CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 119, No. 6 June 2006

(Continued)

CONTENTS—*Continued from preceding page*

NONLINEAR ACOUSTICS [25]

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

UNDERWATER SOUND [30]

ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

 [DOI: 10.1121/1.2198947] Published Online 15 May 2006

Frequency and form as determinants of functor sensitivity in English-acquiring infants

Rushen Shi
Département de psychologie, Université du Québec à Montréal, C.P.8888, Succursale Centre-Ville, Montréal, Québec H3C 3P8 Canada
shi.rushen@uqam.ca

Anne Cutler
Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands
Anne.Cutler@mpi.nl

Janet Werker and Marisa Cruickshank
Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T1Z4, Canada
jwerker@psych.ubc.ca, marisa.cruick@gmail.com

Abstract: High-frequency functors are arguably among the earliest perceived word forms and may assist extraction of initial vocabulary items. Canadian 11- and 8-month-olds were familiarized to pseudo-nouns following either a high-frequency functor *the* or a low-frequency functor *her* versus phonetically similar mispronunciations of each, *kuh* and *ler*, and then tested for recognition of the pseudo-nouns. A preceding *the* (but not *kuh, her, ler*) facilitated extraction of the pseudo-nouns for 11-month-olds; *the* is thus well-specified in form for these infants. However, both *the* and *kuh* (but not *her-ler*) facilitated segmentation for 8-month-olds, suggesting an initial underspecified representation of high-frequency functors.

PACS numbers: 43.71.Ft [JH]
Date Received: January 11, 2006 **Date Accepted:** March 27, 2006

1. INTRODUCTION

The production of an infant's first words is an achievement crowning months of perceptual work. One of the crucial perceptual tasks is the recognition of different acoustic realizations of a word in continuous speech. The speech which infants hear in the first year consists largely of continuous multi-word utterances,[1] in which word boundaries are not reliably marked and in which words may appear in varying phonetic environments. Nonetheless, during the first year of life infants gradually attune perception to the native phonemic repertoire,[2] acquire knowledge of acceptable versus unacceptable sequences of these sounds,[3] and become sensitive to the probable patterns of words of their language.[4] Construction of a vocabulary crucially depends on the ability to extract word forms from a continuous context. With a two-phase familiarization and test procedure, Jusczyk and Aslin[5] demonstrated this extraction ability in prelinguistic infants; with the same type of procedure, the cortical responses to the occurrence in continuous speech of a form previously heard in isolation were observed.[6] Since the first demonstration of noun segmentation,[5] many subsequent studies[7] have examined numerous aspects of content word segmentation.

Function words offer—at least in English and similar languages—a possible aid to infants in this task. More than half of English spontaneous speech consists of functors (e.g., determiners and conjunctions).[8] There are few functor types, so the frequency with which each occurs is high. The high frequency of functors facilitates grammar acquisition in adults' learning of artificial languages.[9] We here inquire whether they may also facilitate, early in language acquisition, the segmentation of potential vocabulary items from speech.

Certainly infants are acquiring sensitivity to functors in the period in which they are beginning to amass a vocabulary. Behavioral[10,11] and ERP[12] evidence indicates that English-learning 10.5- to 11-month-old infants distinguish real from nonsense functors in continuous-speech contexts. Furthermore, infants at 11 and 13 months prefer real functors over minimally modified nonsense functors, suggesting that they recognize real functors with phonetically detailed representation.[11] German-learning 7- to 9 month-old infants[13] and French-learning 6- to 8-month-old infants[14] are able to segment function words from continuous speech. Note that German and French functors are phonologically stronger than English functors, which are predominantly weak.[15] However, across typologically different languages infant-directed speech exhibits contrasting cues, acoustic as well as distributional, which support differentiation of functional from lexical elements.[16]

To examine whether functors facilitate segmentation of potential vocabulary items, we exploited the familiarity effects which appear in the two-phase familiarization and test procedures introduced into infant speech perception research by Jusczyk and Aslin.[5] Infants express familiarity by listening longer in the test phase to known elements, usually items first presented in the familiarization phase. In a variation on this procedure we made familiarity dependent upon infants' ability to segment test items from a preceding functor. If infants are able to perform this segmentation, they will perceive the items presented in isolation in the test phase as familiar. However, if no segmentation occurs (i.e., functor and test item are perceived as an unsegmented whole), the test item when presented alone will be perceived as an unfamiliar form.

As described above, earlier work has shown that 11-month-olds can recognize the correct phonetic form of function words; we therefore began our study with 11-month-olds. Infants were first familiarized with pseudo-nouns preceded by either real or nonsense functors. In the test phase, they heard the same pseudo-nouns without the preceding context. As functor frequency is crucially involved in functors' facilitatory role in language acquisition,[9] and as frequently encountered forms may prove particularly useful in segmentation,[17] we reasoned that it would be important also to examine the frequency factor within the functor class. We therefore contrasted a high-frequency functor (*the*) with a functor with a lower frequency of occurrence (*her*). The difference between these two in the relevant speech type—speech to infants—was established in several corpora. In a corpus of 290 094 words (word types=50 855)[18] of 11 parents speaking to their infants (aged 9 to 13 months; each recorded bi-weekly) the frequency of *the* was 8513, and of *her* was 307. Parental speech to infants up to 24 months of age in three further corpora[19–21] yielded a total of 71 978 word tokens and 16 310 word types; in this total, the token count for *the* is 2484 and for *her* 114. Thus although all functor elements have relatively high frequency compared with content words, *the* is likely to have occurred far more often than *her* in the listening experience of infants.

2. EXPERIMENT 1

Twenty-four 11-month-old monolingual English-learning infants completed the task, 8 in the high frequency and 16 in the low frequency condition. Another 10 infants were tested but were excluded due to fussiness, equipment errors, or insufficient looking time during familiarization. Stimuli were the English functors *the* (high-frequency) and *her* (low-frequency), plus prosodically matched nonsense functors *kuh* and *ler*, which differed from *the* and *her* only in the onset consonants. Stimuli were analyzed for duration, amplitude, and pitch to ensure they were well matched. Both *her* and *ler* were produced with a rhotic /ɚ/ distinct from the schwa in *the* and *kuh*. Both types of functors were paired with the pseudo-nouns *breek* [brik] and *tink* [tiŋk] to form 8 simple noun phrases, *the breek, her tink, ler tink*, etc. A native speaker of American English, a mother of a 6-month-old infant and naive as to the goal of the study, produced the stimuli in an infant-direct speech style. The final set of stimuli included three tokens of each of the eight sequences, and tokens of *breek* (15) and *tink* (15) in isolation.

During the familiarization phase, infants heard six alternating trials (3+3) of "functor+*breek*" and "nonsense functor+*tink*" (or "functor+*tink*" and "nonsense functor +*breek*"). They were then immediately presented with four test trials (two trials with isolated

breek tokens alternating with two trials of isolated *tink* tokens). All trials were 16 s. The interstimulus duration between sequences was 500 ms, with minor adjustments to maintain the same trial length despite slight durational variations among sequences. Orders and functor-pseudoword pairing were counterbalanced across infants, but the test order was constrained: if the last familiarization sequence contained *breek*, the first test trial began with *tink*, and vice versa.

In a sound-attenuated chamber the infant sat on his/her parent's lap facing a central TV monitor and a loudspeaker. During trials, auditory stimuli were presented together with a black-and-white checkerboard on the screen, using HABIT 2000.[22] The parent heard masking music over headphones. The researcher in the next room, blind to the trial orders, observed the infant through a closed-circuit TV, and pressed a computer key whenever an eye fixation occurred. Each trial was initiated upon the infant's eye fixation. Once a trial was initiated, it continued for the entire trial length. All looks to the monitor were recorded online, and were subsequently rescored from videotape records at 30 frames/s; statistical analyses were based on the offline data. Infants who did not reach 15 s of cumulative looking time while receiving the 48-s exposure of each sequence type during familiarization were excluded from the analysis, to ensure that all infants gave sufficient attention to the sequences before the test phase.

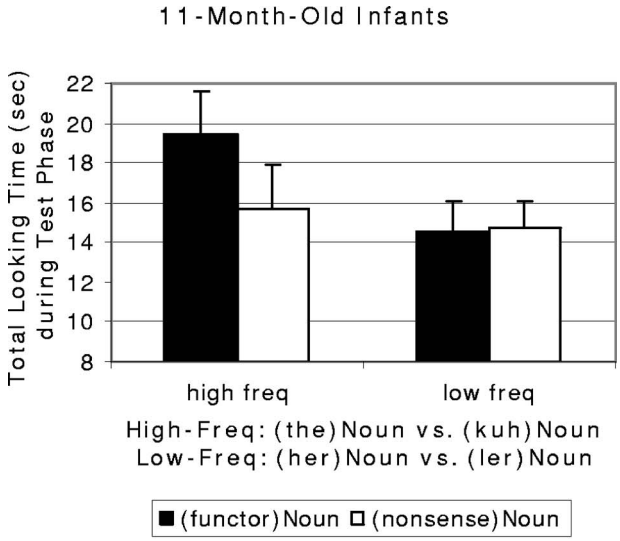
The comparison of interest is relative looking time to the display during presentation of each test pseudo-noun. If the first syllable (functor type) of the familiarization sequences is recognized as familiar, the pseudo-noun with which it is paired will be apprehended as a separate word form and will invoke longer looking time when it appears in isolation in the test phase.

Figure 1(a) shows the very clear pattern which appeared in this experiment. Pseudo-nouns which had been preceded by *the* in the familiarization phase elicited longer looking times than pseudo-nouns preceded by any other of the functor types. For each infant, we calculated the total looking time across *breek* and *tink* trials separately, and conducted a mixed 2×2 ANOVA with familiarization frequency (high, i.e., *the-kuh*, versus low, i.e., *her-ler*) as between-subjects factor, and familiarization functor type (real, i.e., *the-her*, versus nonsense, i.e., *kuh-ler*) as within-subjects factor. This revealed no significant main effect of frequency, but a significant main effect of functor type, $F(1,22)=8.038$; $p=0.01$ (Mean=16.96 s, SE=1.36 s for pseudo-nouns previously paired with real functors; Mean=15.21 s, SE=1.24 s for pseudo-nouns previously paired with nonsense functors); further, the interaction between these factors was significant: $F(1,22)=10.139$; $p=0.004$. Follow-up paired *t*-tests showed that infants listened significantly longer to pseudo-nouns previously familiarized with *the* (Mean=19.41 s, SE=2.18 s) than to those previously familiarized with *kuh* (Mean=15.69 s, SE=2.27 s), $t(7)=3.9$; $p=0.006$, 2-tailed. Every infant in this condition looked longer to *the* than to *kuh* nouns. In contrast, no difference was found in the low-frequency *her-ler* condition.

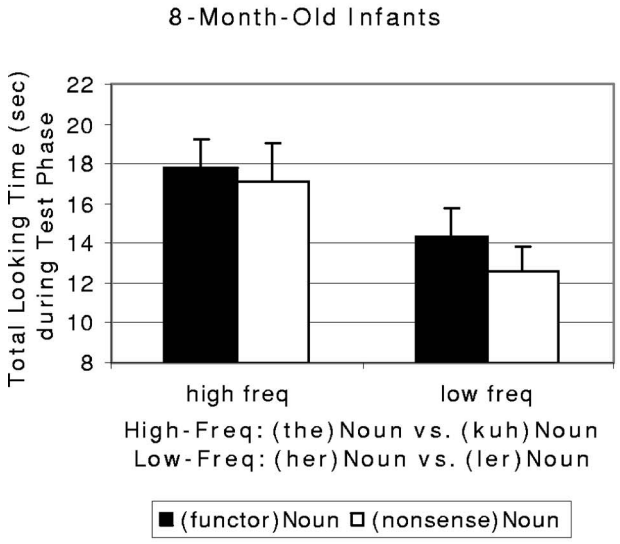
Infants' looking times during familiarization were analyzed in high- and low-frequency conditions; there was no difference between real versus functor sequences, and familiarization exposure was equivalent across the two frequency conditions.

The test phase results clearly indicate that the high-frequency real functor *the* allows 11-month-old English-acquiring infants to segment adjacent word forms from it. Less frequent real functors, and nonsense functors, are significantly less effective. Note that the high-frequency group, in which the pattern is statistically robust, had only half as many participants as the low-frequency group. Thus we can be confident that the failure to find a significant real-functor preference in the low-frequency group is not simply due to lack of power.

In Experiment 2 we repeated the task with younger listeners. In speech to 8-month-olds, *the* is also presumably the most frequent functor type. If frequency is the only determinant of the emerging sensitivity to functor identity, *the* may be equally effective for 8-month-olds.



(a)



(b)

Fig. 1. 11-month-old [(a), Experiment 1] and 8-month-old [(b), Experiment 2] infants' mean total looking time (with SEs) to pseudo-nouns in the test phase, as a function of familiarization functor type. In the high-frequency condition one noun was heard with *the* and another noun with *kuh* during familiarization, i.e., (*the*)Noun, (*kuh*)Noun. In the low-frequency condition the same nouns were heard with *her* and *ler*, respectively, during familiarization, i.e., (*her*)Noun, (*ler*)Noun.

3. EXPERIMENT 2

Twenty-four 8-month-old English-learning infants completed the task. Another five infants were excluded from data analyses due to fussiness, equipment failure, experimenter errors, or insufficient looking time during familiarization. Stimuli and design were as in Experiment 1.

Looking times during familiarization exposure to real versus nonsense functor sequences were compared; there were again no differences across the high- or low-frequency conditions.

The test data were analyzed as a function of familiarization condition, as in Experiment 1. There was no significant main effect of functor type and no interaction of functor type and frequency. However, as shown clearly in Fig. 1(b), there was a significant main effect of frequency: Looking time for pseudo-nouns was longer after familiarization with high-frequency than with low-frequency pairs ($F(1,22)=4.56$, $p=0.044$; Mean=17.43 s, SE =1.52 s for pseudo-nouns previously paired with *the* or *kuh*; Mean=13.46 s, SE=1.07 s for pseudo-nouns previously paired with *her* or *ler*). In other words, when the data were collapsed across *the* and *kuh*, there was evidence of facilitation in the younger listeners' segmentation of the strings whereby the second element produced longer listening times in the test phase. No such effect was exercised by the lower-frequency pair *her* and *ler*, which differed by less than 2 s from each other and by more than 2 s from *the-kuh*.

4. GENERAL DISCUSSION

Our study has revealed that functors indeed assist young infants in segmenting potential word forms from continuous speech. The test words in this experiment were novel forms which the infants could not have already known. The infants were not yet active communicators; few 11-month-olds intentionally produce known words, and no 8-month-olds do. During the second half of their first year, though, infants are actively engaged in learning how to amass a stock of vocabulary items. The occurrence of function words can, we have shown, help them in this task. Constructing a memory representation of words in the mental lexicon requires first segmenting word-size chunks from longer utterances. The presence of a highly frequent, familiar functor assists the infants in treating the adjacent element as a separate word chunk.

Our results suggest it is not the syntactic function of the functors *per se* which exercises the helpful effect. It is, rather, the fact that they occur very frequently in speech (including infant-directed speech). In our study, the longest looking times were recorded to test pseudo-words which had been presented with preceding *the*, the most frequently occurring word in all written and spoken frequency counts of the English language. This functor appears to have occurred often enough in the infants' listening experience such that whenever it is heard, the speech material following it is segmented from *the* and treated as a candidate vocabulary item. Clearly, very high frequency is necessary for this effect, because the real functor *her*, although in absolute terms also a frequently occurring English word, does not facilitate segmentation in the same way either for 11- or for 8-month-olds. As our tallies of word frequency in speech input to infants showed, infants hear *her* very much less often than *the*. The facilitatory effect in segmentation is restricted, at this age, to items which occur very frequently indeed.

Besides the effect of frequency, we discovered a further effect which we would interpret as an effect of acoustic form. For 8-month-olds, not only *the* but also its near match *kuh* facilitated segmentation and thereby produced longer looking times to the pseudo-words presented in isolation. As earlier work had shown,[11] 8-month-olds cannot distinguish real from nonsense functors during word recognition; the earliest age at which English functor representation in phonetic detail is established is 10.5–11 months.[10–12] It is thus no surprise to find that 8-month-olds do not distinguish *the* from *kuh* in this word segmentation task. What is surprising is that both *the* and *kuh* act as significantly more effective facilitators of segmentation for this age group than do *her* and *ler*. Our interpretation of this effect is that for the 8-month-olds, a representation of the highest-frequency form is already in place, but it is as yet acoustically underspecified. Since we only compared *the* with *kuh*, and in particular we did not test the other frequent functor *a*, we cannot make a precise statement about the range of forms which would satisfy this underspecified representation (any syllable with schwa, schwa plus any singleton onset, only acoustically weak onsets, etc.); however, the failure of *her* and *ler* to exercise the facilitatory effect suggests that the form of the underspecified representation has to be reasonably close to the acoustic form of *the* itself.

The difference between full vowels and reduced vowels (mostly schwa) is extremely salient in English; listeners find it acceptable to replace any full vowel by the same vowel with any other level of stress (e.g., the first syllables of *audience, auditorium*, and *audition*, with primary, secondary, and no stress, respectively, may be interchanged), but never tolerate full

vowels exchanging with reduced (e.g., the first syllable of *addition*).[23] Newborn infants are already capable of distinguishing between lists of English content words (which all must have at least one full syllable) and function words (which mostly consist of a monosyllable with a reduced vowel).[24] We suggest that our results indicate that this full-reduced distinction is useful to infants in the crucial work of acquiring a vocabulary. The frequent occurrence of functors with a reduced vowel (not only *the* but also *a*) allows for the early construction of an underspecified "functor" representation with a reduced vowel, and this underspecified representation serves to facilitate segmentation of potential vocabulary items from speech input. At 8 months, this is a useful aid in the beginnings of vocabulary acquisition. Accumulation of further experience makes formation of phonetically detailed functor representations possible; thus at 11 months, *the* facilitates segmentation but *kuh* no longer does. Our previous study[11] showed that a wider range of functors is represented in phonetic detail by age 13 months, so that we would predict that 13-month-old listeners in a two-phase experiment would be able to use *her* also in segmentation.

The task of bootstrapping a vocabulary from continuous speech input is a formidable one, but infants begin to accomplish it by the end of their first year of life. Our results shed further light on how they do this. Frequently occurring forms, perhaps at first represented without full specification, serve as a framework against which potential candidates for vocabulary membership may be identified and extracted.

ACKNOWLEDGMENTS

This research was funded by grants from NSERC and FQRSC to R.S., and NSERC and McDonnell grants to J.W., R.S. and J.W. also acknowledge support from Max Planck Society visiting research stipends. Partial reports of this work were presented at the XIVth and XVth Biennial International Conference on Infant Studies.

[1]J. van de Weijer, "Language input for word discovery," Ph.D. dissertation, University of Nijmegen, Nijmegen 1998.
[2]J. Werker and R. Tees, "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63 (1984).
[3]P. Jusczyk, J. Charles-Luce, and P. Luce, "Infants' sensitivity to phonotactic patterns in the native language," J. Mem. Lang. **33**, 630–645 (1994).
[4]P. Jusczyk, A. Cutler, and N. Redanz, "Infants' preference for the predominant stress patterns of English words," Child Dev. **64**, 675–687 (1993).
[5]P. Jusczyk and R. Aslin, "Infants' detection of the sound patterns of words in fluent speech," Cogn. Psychol. **23**, 1–23 (1995).
[6]V. Kooijman, P. Hagoort, and A. Cutler, "Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech," Cogn. Brain Res. **24**, 109–116 (2005).
[7]J. Saffran, J. Werker, and L. Werner, "The infants's auditory world: Hearing, speech, and the beginnings of language," in *Handbook of Child Psychology*, 6th ed., edited by D. Kuhn and M. Siegler (Wiley, Englewood Cliffs, NJ, 2006).
[8]A. Cutler and D. Carter, "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142 (1987).
[9]V. Valian and S. Coulson, "Anchor points in language learning: The role of marker frequency," J. Mem. Lang. **27**, 71–86 (1988).
[10]M. Shady, "Infants' sensitivity to function morphemes," Ph.D. thesis, SUNY, Buffalo, 1996.
[11]R. Shi, J. Werker, and A. Cutler, "Recognition and representation of function words in English-learning infants," Infancy (in press).
[12]V. Shafer, J. Shucard, D. Shucard, and L. Gerken, "An electrophysiological study of infants' sensitivity to the sound patterns of English speech," J. Speech Hear. Res. **41**, 874–886 (1998).
[13]B. Höhle and J. Weissenborn, "German-learning infants' ability to detect unstressed closed-class elements in continuous speech," Dev. Sci. **6**, 122–127 (2003).
[14]R. Shi, A. Marquis, and B. Gauthier, "Segmentation and representation of function words in preverbal French-learning infants," in *Proceedings of the 30th Boston University Conference on Language Development* (Cascadilla, Somerville, MA, 2006).
[15]A. Cutler, "Phonological cues to open- and closed-class words in the processing of spoken sentences," J. Psycholinguist. Res. **22**, 109–131 (1993).
[16]R. Shi, J. Morgan, and P. Allopenna, "Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective," J. Child Lang. **25**, 169–201 (1998).
[17]H. Bortfeld, J. Morgan, R. Golinkoff, and K. Rathbun, "Mommy and Me: Familiar names help launch babies into speech stream segmentation," Psychol. Sci. **16**, 298–304 (2005).

[18]M. Brent and J. Siskind, "The role of exposure to isolated words in early vocabulary development," Cognition **81**, B33–44 (2001).
[19]E. Bates, I. Bretherton, and L. Snyder, *From First Words to Grammar* (Cambridge University Press, New York, 1988).
[20]N. Bernstein-Ratner and C. Pye, "Higher pitch in babytalk is not universal: Acoustic evidence from Quiche Mayan," J. Child Lang. **11**, 515–522 (1984).
[21]R. Higginson, "Fixing-assimilation in language acquisition," Ph.D. thesis, Washington State University, 1985.
[22]L. Cohen, D. Atkinson, and H. Chaput, "Habit 2000: A new program for testing infant perception and cognition," [Computer software]. Austin, the University of Texas, 2000.
[23]B. Fear, A. Cutler, and S. Butterfield, "The strong/weak syllable distinction in English," J. Acoust. Soc. Am. **97**, 1893–1904 (1995).
[24]R. Shi, J. Werker, and J. Morgan, "Newborn infants' sensitivity to perceptual cues to lexical and grammatical words," Cognition **72**, B11–21 (1999).

Real time decomposition of speech into modulated components

Yadong Wang and Ramdas Kumaresan
Department of Electrical and Computer Engineering, University of Rhode Island, Rhode Island 02881
ydwang@umd.edu

Abstract: Motivated by the active process of the outer hair cell (OHC) in the mammalian inner ear, a real time decomposition of speech into modulated components is presented. A generalized phase lock loop (GPLL) was applied to decompose the speech signal into its envelope and positive instantaneous frequency (PIF) parts, which can be further processed and represented by timing information alone. A log-derivative operator is applied to the bandpass signal. Analytic and antianalytic components occupying non overlapping frequency bands are separated by filtering. The proposed algorithms are used to represent speech signals processed through a bandpass filter bank.

PACS numbers: 43.66.−x, 43.71.−k, 43.60.Mn, 43.25.Ts [DOS]
Date Received: June 7, 2005 **Date Accepted:** March 20, 2006

1. INTRODUCTION

Our signal processing approaches are inspired by the peripheral stages of the auditory system (inner-ear/cochlea). The acoustic signals are separated into frequency components by filtering in the inner ear and represented by streams of spike (nerve impulse) trains in the auditory nerve.[1] Our recent results[2,3] indicate that the information about the phase and envelope modulations of bandpass signals can be represented by certain zero crossings alone (*not those of the signal*), if appropriate adaptive preprocessing (demodulation) is performed on the signals. There is a tantalizing possibility that this sort of adaptive processing may help explain the demodulation performed in the inner ear.

At low sound pressure levels (up to 40–50 dB SPL), cochlear filtering is dominated by an active, sharply tuned nonlinear process that seems to amplify frequency components near the resonant frequency. At higher levels, cochlear filtering is characterized by quasi-linear bandpass filters that are broadly tuned and asymmetric. The human auditory system has incredible sensitivity and enormous dynamic range. Current passive mechanical models of the cochlea have difficulty in reproducing the low-threshold, sharply tuned component of the traveling wave. The only theories to be successful in matching mechanical, hair cell, and auditory nerve tuning suppose that the cochlea contains an active mechanical amplifier. Models that simulate sharp tuning suppose that the outer hair cells feed energy into the traveling wave. Permanent inner ear tones, the so-called spontaneous otoacoustic emissions (SOAEs), are measured in the ear canal of the majority of healthy human ears. They are caused by mechanical oscillations of frequency-tuned outer hair cells. Norena and others[4] found that SOAEs change hearing even in the contralateral ear. The results of Norena and others propose that permanent tones in our ears (SOAEs) are much more than accidental epiphenomena of inner ear mechanics. The $2f_1-f_2$ combination tone can be detected acoustically in the ear canal,[5] again suggesting that an active process is involved. Of them, many details of the active process of one class of cochlear sensory cells, the outer hair cells, remain poorly understood. The main role of the outer hair cells may be to actively influence the mechanics of the cochlea.[6] Hearing sensitivity in mammals is enhanced by more than 40 dB by mechanical amplification thought to be generated by outer hair cells. A recent report by Liberman *et al.*[7] demonstrated successfully that prestin is required for electromotility of the outer hair cell and for the cochlear amplifier. They claimed that "no active process other than that provided by prestin-mediated electromotility need be involved to explain

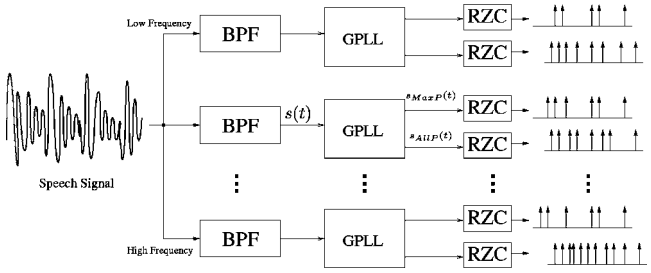


Fig. 1. Tonotopic real zero-crossing converter by generalized phase lock loop in real time. BPF stands for band pass filter and GPLL stands for generalized phase lock loop. The input speech signals are decomposed into overlapping frequency bands. GPLL modules are applied to separate them into $s_{\mathrm{MaxP}}(t)$, which has only envelope information and $s_{\mathrm{AllP}}(t)$, which contains only phase information. Different real-zero conversions (RZs) are used to convert $S_{\mathrm{MaxP}}(t)$ and $S_{\mathrm{AllP}}(t)$ into real-zero signals. The average rate zero crossings of $S_{\mathrm{AllP}}(t)$ are the captured average instantaneous frequency which can be used to recognize speech. The average rate zero crossings of $S_{\mathrm{MaxP}}(t)$ indicates the envelope information in each frequency bands.

exquisite mechanical sensitivity of the mammalian cochlea." Their report[7] seems to conclude a hunt for the motor that amplifies sound vibrations in the inner ear of mammals.

Our question is why there is such an active motor process. Will this process help the cochlea to represent any incoming signals? We suggest that outer hair cells' motors may be the essential parts to code the filtered incoming audio signals in the form of trains of almost identically shaped nerve impulses or spikes, which are transported by the auditory path finally to the auditory cortex. It is likely that this action of the outer hair cells is partly under control of higher centers of the auditory system.[6] The feedback information is believed to be carried by most of the 1800 efferent nerve fibers that make contact with the outer hair cells.[8] We may also propose that distortion product otoacoustic emissions are the by-product of a coding process within inner ear.

Based on these observations, we developed a real time decomposition algorithm which separates the bandpass-filtered speech signal into its modulated components. (See Fig. 1.) We applied a generalized phase lock loop (GPLL), trying to simulate the function of outer hair cells, to decompose the speech signal into its envelope and positive instantaneous frequency (PIF) parts, which can be further processed and represented by timing information alone.[3] A log-derivative operator is applied to a bandpass signal and analytic and antianalytic components occupying nonoverlapping frequency bands are separated by filtering. The proposed algorithms are used to represent speech signals processed through an analysis filter bank. The average of PIF was used as a speech recognition feature. Preliminary results of experiment using HTK with TIDIGIT speech database show its noise robustness. The potential of further improvement is discussed.

2. FIR-LIKE SIGNAL MODELS IN THE ζ PLANE

An analytic signal approach is adopted because it permits an unambiguous characterization of a real signal in terms of its envelope and instantaneous frequency (IF). Recall that if $s(t)$ is a real signal, then the corresponding analytic signal is $s_a(t)=s(t)+j\hat{s}(t)$, where $\hat{s}(t)$ is the Hilbert transform of $s(t)$. The envelope of $s(t)$ is then defined as the magnitude of $s_a(t)(|s_a(t)|)$ and its IF is denoted by the first derivative of $s_a(t)$'s phase function scaled by $1/2\pi$. Characterizing signals by envelopes and IFs is also commonly referred to as AM-FM modeling of signals.

Consider a periodic analytic signal $S_a(t)$, with a period of T s. Let $\Omega=2\pi/T$ denote its fundamental angular frequency. If $s_a(t)$ has finite bandwidth, it may be described by the following model for a sufficiently large M, over an interval of T s:

$$s_a(t) = e^{j\Omega_l t} \sum_{k=0}^{M} a_k e^{jk\Omega t}, \tag{1}$$

where $\Omega_l \geqslant 0$ represents a frequency translation. Ω_l, the low-frequency band edge, is an integer multiple of Ω, say $\Omega_l = K\Omega$. a_k are the complex amplitudes of the sinusoids $e^{jk\Omega t}$; $a_0, a_M \neq 0$. By separating $s_a(t)$ into its the minimum phase (MinP) and maximum phase (MaxP) components, we may express $s_a(t)$ as

$$s_a(t) = A_c e^{j(\Omega_l + Q\Omega)t} \underbrace{e^{\alpha(t)+j\hat{\alpha}(t)}}_{\text{MinP}} \underbrace{e^{\beta(t)-j\hat{\beta}(t)}}_{\text{MaxP}} \tag{2}$$

where Q is the number of zeros (denoted by q_i) outside the unit circle in the complex ζ plane. A_c is $a_0 \Pi_{i=1}^{Q}(-q_i)$. See Ref. 9 for details of the formulas for $\alpha(t)$ and $\beta(t)$.

3. GENERALIZED PHASE LOCK LOOP

We propose to decompose the outputs from bandpass filters or adaptive formant tracking filters into modulations, namely the envelopes and PIFs (positive instantaneous frequencies) by a generalized phase lock loop (GPLL).

The phase lock loop (PLL) has been widely used in communications since the 1930s for FM detection, coherent AM detection, frequency multiplication, frequency synthesis, and carrier phase synchronization. A PLL consists of three basic components: (1) a phase detector, (2) a low-pass filter, and (3) a voltage-controlled oscillator(VCO). The VCO is an oscillator that produces a periodic waveform, according to the value of the applied voltage, which can "lock on" to a sinusoid signal whose frequency is approximately known.

We extend this PLL idea into our speech decomposition model as a GPLL. The input to GPLL is an output from a bandpass filter, which has roughly only one frequency component (around one formant for speech signal). The VCO produces an analytic signal with PIF, whose average IF locks on to the average IF of input signal. The control voltage of VCO comes from a "retain positive frequency" filter, which filters out all the negative frequency in the error signal and keeps only the positive frequency part. Once the VCO has acquired the average of the IF components, the frequency of VCO can track the IF of the input signal.

This algorithm relies on filtering the derivative of the signal's logarithm and has been implemented sample-by-sample. The procedure exploits the algebraic structure of the signal model in Eq. (2). Similar to cepstral processing techniques, we compute the signal's logarithm to convert the product representation into a linear sum. Then we exploit the spectral properties of the MinP and MaxP components to filter them.

Taking the natural logarithm of $s_a(t)$ in Eq. (2) we get

$$\log(s_a(t)) = \log A_c + j\Omega_c t + \alpha(t) + j\hat{\alpha}(t) + \beta(t) - \hat{\beta}(t), \tag{3}$$

where $\Omega_c = \Omega_l + Q\Omega$. Further, taking the time derivative of $\log(s_a(t))$, we have

$$\frac{d}{dt}\log(s_a(t)) = j\Omega_c + \dot{\alpha}(t) + j\hat{\dot{\alpha}}(t) + \dot{\beta}(t) - j\hat{\dot{\beta}}(t). \tag{4}$$

Thus the log-derivative operation on the bandpass signal produces three components. First is the slowly varying component Ω_c that corresponds to the frequency of the most dominant component in $s_a(t)$. Note that $\dot{\alpha}(t) + j\hat{\dot{\alpha}}$ is an analytic signal (spectrum is nonzero for positive frequencies) and $\dot{\beta}(t) - j\hat{\dot{\beta}}$ is also an analytic signal but its spectrum is nonzero only for negative frequencies. We shall call $\dot{\beta}(t) - j\hat{\dot{\beta}}$ the antianalytic (although not a standard mathematical term) component. Thus the analytic and antianalytic components of the signal $d/dt \log(s_a(t))$ occupy nonoverlapping frequency bands and can be separated by filtering. Although these compo-

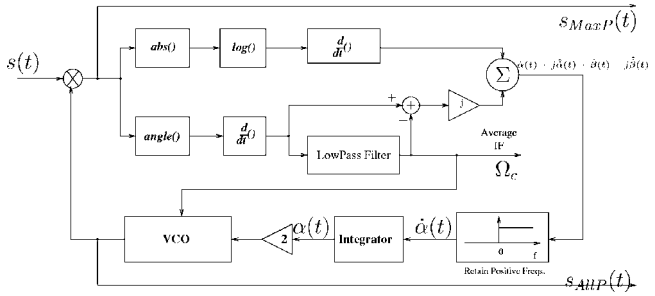


Fig. 2. GPLL module: real time decomposing the speech into modulated components. $\alpha(t)$ is extracted based on log-derivative and fed into the VCO.

nents are band unlimited in general, we need to keep only a finite frequency band and still not lose any information. (This is analogous to the result in cepstral analysis which states that only a finite number of cepstral values are enough to uniquely determine the spectrum of a finite sequence.) The real part of the output of a retain positive frequency filter is $\dot{\alpha}(t)$and its integral yields $\alpha(t)$. The two inputs to the VCO are the average instantaneous frequency and $2\alpha(t)$. The VCO output is simply

$$s_{\mathrm{AllP}}(t) = e^{j(\Omega_l t + Q\Omega t + 2\dot{\alpha}(t))} \tag{5}$$

and the MaxP signal is the real time product of $s_a(t)$ and $s_{\mathrm{AllP}}(t)$:

$$s_{\mathrm{MaxP}}(t) = s_a(t)s_{\mathrm{AllP}}(t) \tag{6}$$

$$= A_c e^{\alpha(t)+\beta(t)-j(\dot{\alpha}(t)+\hat{\beta}(t))}. \tag{7}$$

The procedure is outlined in Fig. 2. There are three outputs from our GPLL-based decomposition: the maximum phase part $s_{\mathrm{MaxP}}(t)$, the all phase part $s_{\mathrm{AllP}}(t)$, and the average instantaneous frequency Ω_c. Both the $s_{\mathrm{MaxP}}(t)$ and $s_{\mathrm{AllP}}(t)$ can be represented by timing information alone. See Ref. 3 for more details.

Figure 3 gives an illustrative example. An analytic signal $s_a(t)$ was synthesized using seven Fourier coefficients: $a_0=1$, $a_1=-0.6024-3.2827i$, $a_2=-5.6441+1.5835i$, $a_3=-0.1454+7.4390i$, $a_4=6.4822-1.1832i$, $a_5=-4.6306-6.7388i$, and $a_6=1.0737+2.7369i$. The real part of $s_a(t)$ is plotted in (b) with the solid line. The IF of $s_a(t)$ and the PIF estimation from the output of VCO are shown in (a). The envelope estimation from MaxP is plotted along the real part of the signal in (b).

4. FORMANT TRACKING

In this section we analyze a speech signal obtained from the TIMIT database. The segment considered is that of a male speaker (waveform timit/test/drl/mrebo/sa2.way). It corresponds to the spoken utterance of "Don't ask me to carry an oily fag like that." Specifically, we consider samples 17 086–36 580, which correspond to the spoken utterance "an oily rag like that" sampled at 16 kHz. It is decimated by a factor of 2 to a sample frequency of 8 kHz. It is then premphasized using a filter with a transfer function $1-0.95z^{-1}$.

For a signal with a stationary spectrum, a fixed bank of bandpass filters (BPFs) centered around regions with significant spectral content may be employed to decompose the signal into components. The drawback of fixed BPFs is significant energy from a nearby filter biases the estimate of energy in a desired filter. For nonstationary signals such a fixed filter bank does not seem reasonable. Hence, we may require that the front filters be slowly varying with time such that they are centered roughly around regions with dominant spectral content (or formant frequencies). To do so, we use an adaptive filter bank. The tracking filters that we employ are

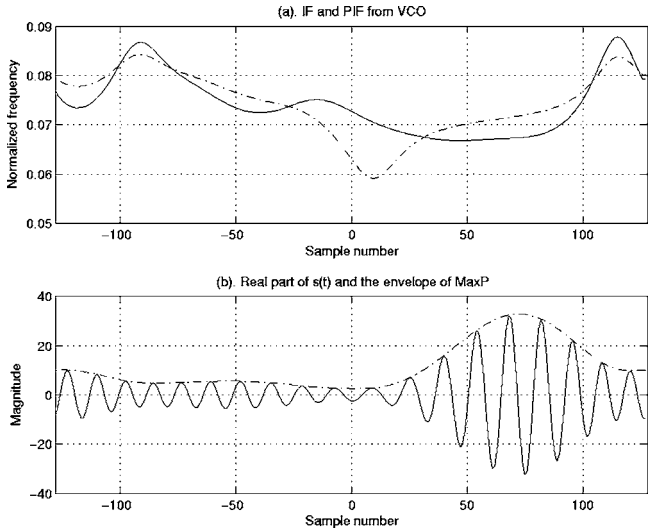


Fig. 3. (a) The IF (dashed line) of $s_a(t)$ and the PIF (solid line) estimation from the output of the VCO. (b) The real part of the original signal is the solid line and the envelope estimation from MaxP is the dashed line.

variations of what are referred to as dynamic tracking filters (DTFs) used for carrier frequency tracking and frequency-feedback demodulation in FM receivers. In addition, we use all-zero filters (AZFs) in a cross-coupled fashion to suppress interferences from strong neighboring components (formant regions) that otherwise may produce large fluctuations in the modulations of $s(t)$ and also significantly alter the estimate of $s(t)$'s average or center frequency. (See Ref. 10 for details about DTFs and AZFs.) After DTFs and AZFs, the band-limited real signals are transformed into analytic signals and the GPLL modules are applied to extract the all phase parts and maximum phase parts. The average PIF from all phase parts (from 4 GPLLs) are plotted in Fig. 4 along with the spectrogram. Observe that they closely follow the formant

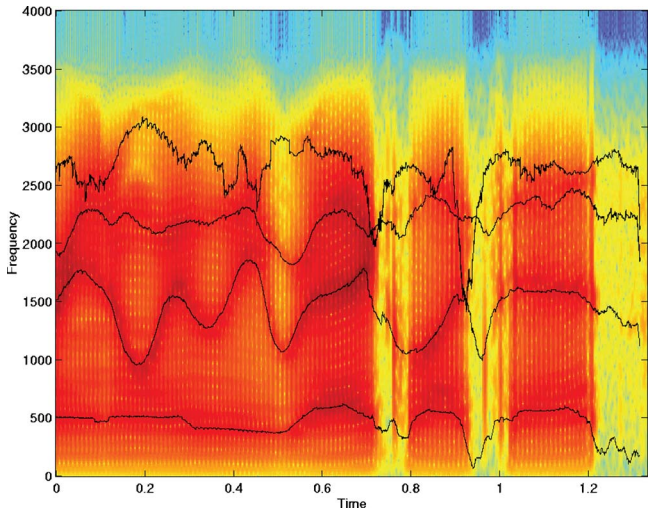


Fig. 4. (Color online) The spectrogram of the input signal and the carrier frequency of the VCO. We use DTF as a formant tracker. It is clear that the average PIFs from the VCOs follow the formants of the input signal.

frequencies. Close to perfect reconstruction of the input speech from the output of the algorithm can be implemented as described by Kumaresan.[3] The results are not reproduced here due to the limited space.

5. DISCUSSION AND CONCLUSIONS

We used average PIFs from 14 VCOs as speech recognition features. Preliminary experiments with HTK for isolated digit from TIDIGITS database show better noise robustness to MFCC.[11]

The GPLL-based algorithm proposed is suitable for implementation in real-time or sample-by-sample in the discrete case with minimal time lag. Additionally, the resulting signal from the decomposition can be represented by timing information alone, while the information about the acoustic signal is carried in the timing information contained in the spike trains emanating from the cochlea in the form of spike rate or inter spike intervals. We emphasize that this model is not intended to be an accurate physiological model of the inner ear but just a means to show that the acoustical signal, like speech, can be filtered, demodulated by active feedback, and decomposed in real time into different parts, which could be represented further by timing information.

ACKNOWLEDGMENTS

The authors would like to thank the two unknown reviewers and especially, the editors for their helpful and useful comments that significantly improved the quality of the paper.

[1]P. Dallos and *et al.* (eds.), *The Cochlea* (Springer-Verlag, New York, 1996).
[2]R. Kumaresan and Y. Wang, "A new real-zero conversion algorithm," in *Proceedings of the ICASSP 2000*, lstanbul, Turkey, June 2000.
[3]R. Kumaresan and Y. Wang, "On representing signals using only timing information," J. Acoust. Soc. Am. **110**, 2421–2439 (2001).
[4]A. Norena, C. Micheyl, J. Durrant, S. Chery-Croze, and L. Collet, "Perceptual correlates of neural plasticity related to spontaneous otoacoustic emissions?" Hear. Res. **171**(1-2), 66–71 (2002).
[5]A. M. Brown and D. T. Kemp, "Suppressibility of the 2fl-f2 stimulated acoustic emissions in gerbil and man," Hear. Res. **13**(1), 29–37 (1984).
[6]B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 4th ed. (Academic, San Diego, 1997).
[7]M. C. Liberman, J. Gao, D. Z. He, X. Wu, S. Jia, and J. Zuo, "Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier," Nature (London) **419**(6904), 300–304 (2002).
[8]J. O. Pickles, *An Introduction to the Physiology of Hearing*, 2nd ed. (Academic, London, 1988).
[9]R. Kumaresan and A. Rao, "Model-based approach to envelope and positive-instantaneous frequency of signals and application to speech," J. Acoust. Soc. Am. **105**(3), 1912–1924 (1999).
[10]A. Rao and R. Kumaresan, "On decomposing speech into modulated components," IEEE Trans. Speech Audio Process. **8**, 240–254 (2000).
[11]Y. Wang, J. Hansen, G. K. Allu, and R. Kumaresan, "Average instantaneous frequencies and average log-envelopes for asr with the aurora 2 database," in *Eurospeech'03, Geneva, Switzerland*, 2003 pp. 21–25.

ACOUSTICAL NEWS—USA

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

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

Preliminary Notice: 4th joint meeting of the Acoustical Society of America and the Acoustical Society of Japan

The 4th joint meeting of the Acoustical Society of America and the Acoustical Society of Japan will be held Tuesday through Saturday, 28 November–2 December 2006 at the Sheraton Waikiki Hotel, Honolulu, Hawaii. A room block has been reserved at the Sheraton Waikiki.

The deadline for submission of abstracts for the meeting is Friday, 30 June 2006. Please see the printed call for papers or the online version for instructions on the preparation and submission of abstracts at ⟨http://asa.aip.org/honolulu/honolulu.html⟩. Other information about the meeting also appears on the ASA Home Page at ⟨http://asa.aip.org/meetings.html⟩.

Greetings from the Presidents of ASA and ASJ

Dear Acoustician,

We enthusiastically invite you to attend the 4th ASA/ASJ joint meeting which will be held from 28 November to 2 December on the beautiful island of Oahu, Hawaii. The meeting will be at the Sheraton Waikiki Hotel, where it has been held for the previous three meetings (1978, 1988, and 1996).

Those of you who have attended these meetings before know that this is a beautiful site on the Pacific Ocean with access to many shops and restaurants. The fact that Hawaii is located between North America and Asia provides a special opportunity to meet with colleagues from these two continents, as well as from around the world. The meeting will start on a Tuesday and finish on Saturday to allow U.S. members to enjoy the full Thanksgiving holiday, and not have to travel on the busy Sunday preceding the meeting. For those who wish to stay longer, Internoise 2006 will meet in the same hotel immediately following the joint ASA/ASJ meeting.

We look forward to receiving your abstract and seeing you in Hawaii.
Sincerely,
Anthony Atchley, President, Acoustical Society of America (2006–2007)
Yôiti Suzuki, President Acoustical Society of Japan (2005–2007)

Technical Program

The Technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Tuesday through Friday, 28 November–2 December. The special sessions described below will be organized by ASA and ASJ Technical Committees.

Special Sessions

Acoustical Oceanography (AO)

Acoustic scattering by aquatic organisms
(Joint with Underwater Acoustics)
Advances in applications, laboratory measurements, and theoretical modeling of acoustic scattering by pelagic and benthic organisms.
Acoustic tomography for coastal and deep water applications
(Joint with Signal Processing in Acoustics)
Recent advances in ocean acoustic tomography in both coastal and deep water, including experimental research, model simulations, and theory.

Animal Bioacoustics (AB)

Acoustic scene analysis
Problems of auditory scene analysis in a broad range of species that use hearing to solve a variety of tasks.
Fish bioacoustics: A tribute to Arthur Myrberg
To honor the memory of Professor Myrberg and his contributions to animal bioacoustics.
Remote monitoring of biological sounds
(Joint with Underwater Acoustics)
Remote acoustical monitoring technologies and methodologies for the study of aerial and aquatic wildlife; results of behavioral studies using remote monitoring.
Vertebrate vocal production: Themes and variations
Review the impressive progress made in the last decade in understanding the physics and physiology underlying vocal production in vertebrates.

Architectural Acoustics (AA)

Acoustic design for Green buildings
(Joint with Noise)
Architectural designs responsive to U.S. Green and LEED certifications and Japan's CASBEE certifications do not specifically address good acoustical design practices.
Acoustics of large indoor and outdoor performance spaces
Issues in acoustic design, measurement, or electroacoustic reinforcement/enhancement relating to large indoor and outdoor performance spaces.
Computer modeling of room acoustics
Research and developments in computer modeling of room acoustics.
Measurement of room acoustics
New technologies to measure impulse response, reverberation time, and problems for measurements; standards and their application; prediction and results of measurements; uncertainty of measurements.
Psychological aspects of speech in rooms
(Joint with Speech Communication and Psychological and Physiological Acoustics)
To encourage collaboration among speech and architectural acoustics researchers.
Recent developments in acoustical materials and structures
(Joint with Structural Acoustics and Vibration, Noise, and Engineering Acoustics)
Numerical study, measurement, and practical installation techniques for sound insulation and absorption.
Session in honor of Manfred R. Schroeder
(Joint with Speech Communication, Musical Acoustics, Psychological and Physiological Acoustics and Signal Processing in Acoustics)
Honoring Manfred R. Schroeder, his contributions and his life in acoustics.

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Acoustic tomography in tissue
(Joint with Signal Processing in Acoustics)
Use of acoustic tomography to image tissue.
Interaction of cavitation bubbles with cells and tissue
Encompasses the formation of cavitation bubbles in human and animal tissue due to ultrasound exposure, the molecular level effects of bubble-cell interactions, and experimental techniques for investigating this phenomena.
Celebration of Floyd Dunn
(Joint with Physical Acoustics)
Celebration of the career of Floyd Dunn.
Elastic imaging
(Joint with Signal Processing in Acoustics)
Use of ultrasound to determine elastic properties of tissue and the enhancement of ultrasound imaging by elastic deformation of tissue.
Topical meeting on shock wave therapy
Latest clinical, regulatory, engineering, and basic scientific advances in shock wave therapy.

 0001-4966/2006/119(6)/3493/10/$22.50

Ultrasound enhancement of drug activation
Use of ultrasound to enhance the activation of drugs in the body.

Education in Acoustics (ED)

Demonstrations and tools in acoustics education
Demonstration of tools for acoustics education, including experimental teaching materials, hands-on tools, simulations, visualization, and auralization. Topics are introduced in brief lectures, details are demonstrated poster style, with table space reserved for each presentation.
Education in acoustics for children
Educational activities and tools to stimulate children's interest in acoustics and science.
Take 5's
Sign up at the beginning of the session. No abstracts required.

Engineering Acoustics (EA)

Acoustic subbottom profiling technology
(Joint with Acoustical Oceanography and Underwater Acoustics)
Recent advances in subbottom profiling techniques.
Developments in microphones: Calibration, standards, and measurements
(Joint with ASA Committee on Standards)
Front end of acoustical measurements including standards.
Hearing aids
(Joint with ASA Committee on Standards and Psychological and Physiological Acoustics)
Latest developments in hearing aid technology.
New electroacoustic transducers utilizing advanced technologies and materials
Latest work on electroacoustic transducers utilizing advanced technologies and materials.
Ultrasonic functional devices
Recent advances in functional ultrasonic devices, including piezoelectric and MEMS-based devices, arrays, sensors, motors, and actuators.

Musical Acoustics (MU)

Acoustic correlates of timbre in music and musical instruments
(Joint with Psychological and Physiological Acoustics)
Evidence for old and new musical timbre correlates using multidimensional scaling of dissimilarity data and other psychometric methods.
Comparing Asian and Western instruments
Comparison of the acoustics of Asian and Western musical instruments.
Music information and communication
Musical aspects relevant to the player and the audience discussed from viewpoints, such as aesthetic evaluation, information retrieval, rhythm recognition, performance skill, and emotional communication.
Simulation and measurement techniques for musical acoustics
Recent developments in modeling and measurement of musical instrument structures including flow visualization, input impedance/admittance measures, digital waveguide methods, and novel analysis/synthesis methods.

Noise (NS)

Acoustics of sports
(Joint with Noise, Physical Acoustics, and Structural Acoustics and Vibration)
Topics of acoustical interest relating to athletic sports and their venues; including sounds of the game, acoustical cues, crowd noise, public address systems, incidental entertainment, broadcast and recording of sporting events.
Measurement, design, and control of consumer product noise
(Joint with ASA Committee on Standards)
Measurement, design, and control of consumer product noise.
Prediction and propagation of outdoor noise
(Joint with Physical Acoustics)
Propagation of noise outdoors in urban environments and its prediction.
Progress in hearing protection and hearing conservation
(Joint with Speech Communication and ASA Committee on Standards)
Developments in design, application, and measurement of hearing protection and promotion of hearing conservation in both occupational and nonoccupational settings.
Soundscapes and cultural perceptio
(Joint with Architectural Acoustics)
Cultural and historical attitudes and influences of soundscape perception.

Physical Acoustics (PA)

Cavitation
General topics in cavitation and cavitation clouds. Topics include, but are not limited to, sonoluminescence, sonochemistry, dynamics of cavitation clouds, and medical applications.
Infrasound
(Joint with Engineering Acoustics)
Low-frequency waves due to natural effects such as volcano activity, weather, storms, and tsunamis.
Sonic boom from supersonic aircraft and projectiles
(Joint with Noise)
Recent research on the generation, propagation, measurement, and human response to sonic boom.
Sound propagation in inhomogeneous media
(Joint with Biomedical Ultrasound/Bioresponse to Vibration)
Focus on mixtures of solid and liquid materials, as well as a mixture of liquid and bubbles.
Sound speeds, phonons, and the thermodynamics of condensed matter
Focus on the connection between sound speeds, elastic moduli, and phonons in condensed matter systems and underlying fundamental thermodynamic processes.

Psychological and Physiological Acoustics (PP)

Auditory grammar
Exploration of how the perceptual allocation of temporal features in forming auditory events may constrain the processing of linguistic and nonlinguistic stimuli. Examples include auditory continuity and phonemic restoration illusions.
New insights on loudness and hearing thresholds
(Joint with ASA Committee on Standards)
Relationship between loudness and hearing thresholds.
Perception of music and speech: Evidence for modularity and for common processing mechanisms
Longstanding debate continues on whether the perception of speech and music involve shared neural substrates or domain-specific processes. This issue will be addressed from multiple perspectives, including data from individuals with unusual abilities or disabilities.

Signal Processing in Acoustics (SP)

Adaptive acoustic signal processing
(Joint with Underwater Acoustics and Acoustical Oceanography)
Recent developments in adaptive signal processing approaches in air and underwater acoustics.
Blind signal processing
Blind source separation and dereverberation of speech and audio signals, blind channel identification of acoustic systems, independent component analysis, sparse component analysis, computational auditory scene analysis.
Sensor array and its applications
DOA estimation, microphone arrays, signal enhancement, and applications based on these technologies, e.g., robotics, hands-free speech communication and recognition, and teleconferencing.
Spatial sound processing: Control and performance evaluation
(Joint with Architectural Acoustics, Engineering Acoustics, and Psychological and Physiological Acoustics)
Focus on signal processing intended to control spatial auditory imagery associated with headphone and loudspeaker reproduction.

Speech Communication (SC)

Communicative speech synthesis and spoken dialog
State-of-the-art expressive speech synthesis and spoken dialogue systems and their potential in future collaborative research topics in speech communication.
Second language acquisition
(Joint with Psychological and Physiological Acoustics)

Explores how researchers in phonetic science and speech technologies might collaborate to create new methodologies for research and new instructional approaches for second-language acquisition.
Speech timing and pronunciation training for the Japanese language
Understanding the rhythm of Japanese timing and methods for training in Japanese by foreign speakers.

Structural Acoustics and Vibration (SA)

Numerical methods in structural acoustics
Application and validation of numerical methods in structural acoustics and vibration.
Vehicle interior noise and vibration
Sources of and propagation paths for vibration in vehicles, such as automobiles, that produce noise inside the vehicle.
Vibration and impact sound in buildings
(Joint with Architectural Acoustics)
Vibration generated and transmitted inside and from outside buildings and impact sound generated by human activity inside buildings.

Underwater Acoustics (UW)

Sediment acoustic processes: Quantifying the effects of sediment properties on dispersion, attenuation, reflection, scattering, and buried target detection
(Joint with Acoustical Oceanography)
Recent results on sound speed dispersion and attenuation in marine sediments, acoustics of the surficial transition layer (e.g., bottom reflection and depth dependence of compressional and shear wave velocities), bottom scattering, and buried target detection.
Session in honor of Frederick H. Fisher
Honors the many contributions of Fred Fisher to underwater acoustics and acoustical oceanography.
Session in honor of Leonid Brekhovskikh
Honors the many contributions of Leonid Brekhovskikh to the understanding of wave propagation in layered media.

Other Technical Events

Opening Plenary Lectures

The 4th joint ASA/ASJ meeting will start with an opening plenary session on Tuesday, 28 November 2006. Two plenary talks will be presented, one by Dr. Lawrence Crum from the ASA and the University of Washington, and another by Professor Masayuki Morimoto from the ASJ and Kobe University. Dr. Crum's talk is entitled "Therapeutic Ultrasound." He will discuss the applications of ultrasound in medicine including: Tissue ablation, acoustocautery, lipoplasty, site-specific and ultrasound mediated drug activity, extracorporeal lithotripsy, and the enhancement of natural physiological functions, such as wound healing and tissue regeneration. Professor Morimoto's talk is entitled "How can auditory presence be generated and controlled?" He will discuss key results of many listening tests on auditory presence, especially auditory localization (AL) and auditory spatial impression (ASI). AL and ASI are the primary characteristics of auditory sensation associated with the acoustics of a space. This paper aims to explain the relationship between those auditory phenomena and the acoustic signals received by a human listener. The author gave the first demonstration that auditory localization in any direction can be simulated through two loudspeakers using head-related transfer functions (HRTFs).

Online Meeting Papers

Papers voluntarily submitted by authors may be obtained at an online site which can be found at ⟨http://scitation.aip.org/asameetingpapers/ ⟩. Authors of papers to be presented at meetings will be able to post their full papers or presentation materials for others who are interested in obtaining detailed information about meeting presentations. The online site will be open for author submissions in September 2006. Submission procedures and password information will be mailed to authors with the acceptance notices.

Those interested in obtaining copies of submitted papers for this meeting may access the service at any time. No password is needed.

Inter-Noise 2006

Inter-Noise 2006 will follow the joint ASA/ASJ meeting on 3–6 December 2006. This conference on noise control engineering will also be held in the Sheraton Waikiki, so it will be convenient to attend after the ASA/ASJ meeting. Details on abstract deadlines, technical sessions, plenary speakers, and other aspects of Inter-Noise 2006 can be found at ⟨www.internoise2006.org⟩.

Special Meeting Features

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 24 October 2006 to: Jolene Ehl, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359, Fax: 516-576-2377, E-mail: jehl@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Honolulu meeting. This award is designed for young professionals who have completed a doctorate in the past five years (not currently enrolled as a student), who plan to present a paper at the Honolulu meeting. Each award will be of the order of $300. It is anticipated that the Committee will grant a maximum of three awards. Applicants should submit a request for support, a copy of the abstract they have submitted for the meeting, and a current resume/vita which provides information on their involvement in the field of acoustics and to the ASA to: Dr. Donna L. Neff, Boys Town National Research Hospital, 555 North 30th Street, Omaha NE 68131; Tel.: 402-452-5069; Fax: 402-452-5027; Email: neff@boystown.org. Deadline for receipt of applications is 10 October 2006.

Students Meet Members for Lunch

The ASA Education Committee provides a way for a student to meet one-on-one with a member of the Acoustical Society over lunch. The purpose is to make it easier for students to meet and interact with members at ASA meetings. Each lunch pairing is arranged separately. Students who wish to participate should contact David Blackstock, University of Texas at Austin, by email ⟨dtb@mail.utexas.edu⟩ or telephone 512-343-8248 (alternative number 512-471-3145). Please give Dr. Blackstock your name, university, department, degree you are seeking (BS, MS, or PhD), research field, acoustical interests, and days you are free for lunch. The sign-up deadline is ten days before the start of the meeting, but an earlier sign-up is strongly encouraged. Each participant pays for his/her own meal.

Banquet, Awards Ceremony, and Social Events

A buffet social with cash bar will be held on Wednesday evening, 29 November 2006, at the Sheraton Waikiki Hotel.

A banquet will be held on Friday evening, 1 December 2006. Banquet festivities will start at 6:30 p.m. in the Sheraton Waikiki second floor foyer with a no-host social hour. The banquet, starting at 7:00 p.m., features a seven-course Chinese dinner. The awards ceremony will be held at the banquet. There will be a special presentation of Hawaiian music as part of the entertainment. Tickets for the banquet will be $40 each (students $25). Banquet tickets are limited, and are not included in the registration fee. Those who wish to attend the banquet must purchase tickets using the registration form in the printed call for papers or online at ⟨http://asa.aip.org⟩. Banquet ticket sales will close at 5:00 p.m., Wednesday, 29 November 2006.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Thursday, 30 November 2006. Those who wish to attend this luncheon must register using the registration form in the printed call for papers or online at ⟨http://asa.aip.org ⟩. The fee is $15 (students $5) for preregistration by 24 October 2006 and $20 (students $5) thereafter including on-site registration at the meeting.

Transportation and Hotel Accommodations

Air Transportation

Honolulu International Airport (HNL) is located on Oahu and is served by the following airlines: Air Canada, Air New Zealand, Air Pacific Airways, All Nippon Airways, Aloha Airlines, American Airlines, American Trans Air, China Airlines, Continental Air Micronesia, Continental Airlines, Delta Airlines, Hawaiian Airlines, Korean Airlines, Japan Airlines, Japan Air Charter, JTB Aloha Service, Northwest Airlines, Philippine Airlines, Rich International, Quantas Airways, and United Airlines. For further information, see ⟨http://www.hawaii.gov/dot/airports/index.htm⟩.

Ground Transportation

Airport Waikiki Express provides transportation from the airport to any hotel in Waikiki. Operates 24 hours: Every 25–30 minutes between the hours of 6:00 a.m. and 10:00 a.m. and every 20–25 minutes after 10:00 a.m. Fees: Adults (one way)—$8.00; Adults (round trip)—$14.00; Children 3 years old or younger ride free.
Baggage—Two (2) bags per person—no fee; Golf bag, car seat, stroller, large boxes, and boogie boards + any bag over the (2) per person limit—$3.00.
Airport Shuttle—Tel. (808) 680-0000
1 person $20.00, ea add'l person $4.00 up to a group of 6 (6:00 a.m.–10:00 p.m.) After 10:00 p.m., $5.00 extra for whole group.
Aloha Airport Shuttle—Tel. (808) 538-3355
$30.00 for 4 people or less (minivans hold 4 people + bags).
Car Rental: Most major car rental companies have rental counters at the Honolulu International Airport located in the baggage claim area. The self-parking rate at the Sheraton Waikiki is $10.00 per day and valet parking is $15.00 per day.
Taxi
AMPCO Express—Tel. (808) 861-8294 is the managing contractor of the Airport's Open Taxi System at Honolulu International Airport. Taxi service is available on the center median fronting the terminal baggage claim areas. See the taxi dispatchers (white shirts and blue vests with a yellow stripe, and the wording "TAXI" in black and white lettering on the front) for service. The fare from the airport to Waikiki during nonrush hour periods is approximately $25.00–$35.00.
Public Transportation—"TheBus" Service
(808) 848-4500 Customer Service; (808) 848-5555 Route Information
Arrival Times: City buses arrive approximately every 30 minutes depending on route number. Bus stops are located in front of the Interisland Terminal and Lobbies 4 and 7 of the main terminal.
Fees: Adults—$2.00; Children and students (up to high school)—$1.00; Visitor pass (Four consecutive day pass. Unlimited use.)—$20.00
Baggage rules: No backpacks with metal frames allowed. All bags must be able to be placed under the feet or on the lap without protruding into the aisle. None larger than 18"×24"×12".
Bus Route No. 19: Eastbound route goes from the airport, then to Kalihi Kai (via Nimitz Highway), downtown Honolulu, the Honolulu Civic Center, Ala Moana Shopping Center, and ends in Waikiki. Westbound route goes from the airport and ends at Hickam Air Force Base.
Bus Route No. 20: Eastbound route goes from the airport, then to Kalihi (via Nimitz Highway), downtown Honolulu, the Honolulu Civic Center, Ala Moana Shopping Center, and ends in Waikiki. Westbound route goes from the airport to the Arizona Memorial, Aloha Stadium, Aiea, and ends at Pearlridge Shopping Center.

Driving Directions from Airport
From Honolulu International Airport, take Nimitz Highway to Ala Moana Boulevard, then turn right onto Kalakaua Avenue. Turn right onto Royal Hawaiian Avenue.

Hotel Accommodations

All meeting functions will be held at Sheraton Waikiki Hotel where a block of sleeping rooms is available. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America (ASA). The reservation cut-off date for the special discounted ASA meeting rates is 24 October 2006; after this date, the meeting rates will no longer be available. Parking costs at the Sheraton are $10 per day for self-parking; $15 per day for valet parking. **Early reservations are strongly recommended**. Note that the special ASA meeting rates are not guaranteed after 24 October 2006. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

The hotel features a fitness center, two swimming pools, twenty gift and specialty shops, and Keiki Aloha Children's Program. See ⟨http://www.sheraton-waikiki.com/⟩ for further details.

In-room amenities include: Iron and ironing board, in-room safe, hair dryer, cable television, shampoo and conditioner, high-speed internet access, coffee maker, and lanai.

Sheraton Waikiki Hotel
2255 Kalakaua Avenue
Honolulu, HI 96815-2571
Tel.: 808-922-4422
Fax: 808-923-8785
⟨http://www.starwoodmeeting.com/StarGroupsWeb/res?id=0603240154&key=C3067⟩

Rates [in US dollars; Single/double occupancy; plus state tax and accommodation tax]:
$159.00—City view
$170.00—Mountain view
$190.00—Partial ocean view
$210.00—Ocean front view
$60.00—Extra person in room (after 2 people)

Rates are available 3 days prior and 3 days after the contracted dates, based upon availability.

Room Sharing

ASA will compile a list of those who wish to share an hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, and smoker or nonsmoker preference, by 16 October to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Acoustical Society of America, Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

Weather on all of the Hawaiian Islands is very consistent, with only minor changes in temperature throughout the year. This is due to year-round warm sea surface temperatures, which keep the overlying atmosphere warm as well. In practical terms, there are only 2 seasons: The summer months (called Kau in Hawaiian) that extend from May to October, and the winter months (Ho'oilo) that run from November to April. The average daytime summer temperature at sea level is 85°F (29.4°C), while the average daytime winter temperature is 78°F (25.6°C). Temperatures at night are approximately 10°F lower. For additional information on Honolulu's weather, visit ⟨www.gohawaii.com⟩.

Security

Attendees should be aware that December is a "big wave" season in Hawaii. While waves of 10 to 30 ft are exhilarating to watch, their reach on the beach can be treacherous. We lose a few visitors each year due to lack of respect for the ocean. Swimmers should be aware that stinging jellyfish may reduce the pleasure of a winter swim. Please observe lifeguard warnings and any cautionary notices posted at beaches.

All attendees are cautioned not to carry significant amounts of cash with them, not to carry handbags on their sides next to a street, and not to leave possessions unattended, particularly at beaches or in vehicles parked at beaches and tourist viewpoints. While people generally are quite safe in their persons, Honolulu is a tourist center, and there is a crime element that targets tourists. It is not impolite to refuse solicitations and police should be called if people become too aggressive. Young children should not be left alone in hotel rooms (see Child Care below).

Child Care

The Sheraton Waikiki provides free child care in its "Children's Center" for guests of all Sheraton Hotels in Waikiki. The Center programs varied activities, from 9:00 a.m. to 5:00 p.m., for children ages 5–12. A "Convention Care" program is available from 6:00 p.m. to 11:00 p.m. by request for an hourly fee.

General Information

Assistive Listening Devices

Anyone planning to attend the meeting who will require the use of an assistive listening device, is requested to advise the Society in advance of the meeting: Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, asa@aip.org.

Accompanying Persons Program

Spouses and other visitors are welcome at the Honolulu meeting. The registration fee for accompanying persons is $50 for preregistration by 24 October and $75 thereafter, including on-site registration at the meeting. A hospitality room for accompanying persons will be open in the Sheraton Waikiki, from 8:00 a.m. to 5:00 p.m. each day throughout the meeting where information about activities in and around Honolulu will be provided. There will be a short program daily lasting for about 1 hour and will include activities such as lei making, hula demonstration, and history, and other presentations of general interest.

Tourism is the major industry of the State of Hawaii and there are many sites to see, activities to participate in, and places to shop. Some examples of interesting sites to visit:

Polynesian Cultural Center on the North Shore of Oahu, about an hour's drive from Waikiki. See recreation of Pacific Island villages and demonstration of cultural activities. There is an admission price. See the hotel concierge.

Iolani Palace in downtown Honolulu is the only Royal Place in the United States and was built by Hawaiian King David Kalakaua. The restored palace and its furnishings are on display. Guided tour must be reserved in advance and an admission fee is required. For more information, call 522-0832 or see the hotel concierge.

USS Arizona and USS Bowfin next to the famous Pearl Harbor. A must see for first-time visitors.

USS Missouri on Ford Island, a part of the Pearl Harbor environment, is a restored battleship on which the surrender documents ending WWII were signed.

Hanauma Bay is the place for you if you ever dreamt of swimming in an aquarium full of fishes. This is the favorite and popular spot for snorkeling in the whole. Closed on Tuesday with an admission fee of $5/person. 396-4229. Go early and take lots of sunscreen.

The Diamond Head hike in the crater of Diamond Head consists of climbing 99 step stairs plus a spiraling staircase and another set of stairs to a WWI vintage outpost with a commanding view of Waikiki.

The Pali lookout is a windy site easily reached by automobile with a commanding view of the windward side of Oahu and the beach town of Kailua and Kaneohe.

Ala Moana Shopping Center is the first shopping center opened in Hawaii and is the most prestigious of all shopping centers in the state. It is located on the town end of Waikiki.

Aloha Stadium swap meet is a place to obtain bargain prices for a wide variety of new items. It is opened on Wednesday, Saturday, and Sunday, 8:00 a.m. until 3:00 p.m. This is the place to get all your souvenirs for your family and friends back home.

Registration Information

The registration desk at the meeting will open on Monday afternoon, 27 November 2006, at the Sheraton Waikiki Hotel. To register, use the registration form in the printed call for papers or register online at ⟨http://asa.aip.org⟩. **If your registration is not received at the ASA headquarters by 31 October 2006, you must register on-site.**

Registration fees are as follows:

Category	Preregistration by 24 October	Registration after 24 October
ASA/ASJ Members	$350	$425
ASA/ASJ Members One-Day	$175	$215
Nonmembers	$400	$475
Nonmembers One-Day	$200	$240
Nonmember Invited Speakers One-Day Only	Fee waived	Fee waived
Nonmember Invited Speakers (Includes one-year ASA membership upon completion of an application)	$110	$110
ASA/ASJ Student Members (with current ID cards)	Fee waived	Fee waived
Nonmember Students (with current ID cards)	$40	$50
Emeritus members of ASA (Emeritus status preapproved by ASA)	$50	$75
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$50	$75

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2007) of membership. Invited speakers who are members of the Acoustical Society of America or the Acoustical Society of Japan are expected to pay the registration fee, but **nonmember invited speakers** may register for one-day only without charge. A nonmember invited speaker who pays the full-week registration fee, will be given one free year of membership upon completion of an ASA application form.

Note: A $25 processing fee will be charged to those who wish to cancel their registration after 24 October 2006.

Online Registration

Online registration is available at ⟨asa.aip.org⟩.

Members of the Joint Meeting Committee

ASA Meeting Committee

General Chair—Whitlow W. L. Au; Technical Program Chair—Anthony A. Atchley; Audio-Visual—Timothy F. Noonan; Accompanying Persons Program—Dorothy E. Au; Signs/Publicity—Marc O. Lammers; Meeting Room Coordinator—John S. Allen; Food—David L. Adams; Posters—Paul E. Nachtigall; Banquet Entertainment—Todd R. Beiler; Special Affairs—Neal Frazer; Public Relations—William Friedl; and Consultant—John C. Burgess.

ASJ Meeting Committee

General Chair—Sadaoki Furui; Technical Program Chair—Yôiti Suzuki; and Technical Program Secretary—Hiroshi Sato.

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

17–21 Sept. INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA; WWW: [www.interspeech2006.org]

28 Nov.–2 Dec. 4th Joint Meeting of the Acoustical Society of America and the Acoustical 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; WWW: http://asa.aip.org]. Deadline for receipt of abstracts: 30 June 2006.

2007

4–8 June 153rd Meeting of the Acoustical Society of America, Salt Lake City, 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; WWW: 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; E-mail: asa@aip.org; WWW: http://asa.aip.org].

2008

28 July–1 Aug. 9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, P.O. Box 1609, Groton, CT 06340-1609, Tel. 860-572-0680; WWW: 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 card. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

Revision List

New Associates

Anderson, Victoria B., Univ. of Hawaii, Linguistics, 569 Moore Hall, 1890 East-West Rd., Honolulu, HI 96822.

Andria, Ken P., Acoustic Dimensions, 145 Huguenot St., Suite 406, New Rochelle, NY 10801.

Bachand, Corey L., BTech Acoustics, LLC, 17 Surry Rd., Barrington, RI 02806.

Bachrach, Benjamin I., Quality Solutions Discovery, 21835 Cherry Hill St., Dearborn, MI 48124-1149.

Booth, Edward M., Northwest Wood Reflections, Inc., 2687 Terrace View Dr., Eugene, OR 97405.

Carim, Hatim, 3M Company, 3M Ctr., Bldg. 270-4N-09, St. Paul, MN 55144.

Cerniway, Matthew, 100 Branchside Ln., Holly Springs, NC 27540.

Clark, Grace A., Electronics Engineering, Lawrence Livermore Natl. Lab., Univ. of California, 7000 East Ave., L-130, Livermore, CA 94550.

Colen, Marc, LOMSC, P.O. Box 2218, Agoura, CA 91376-2218.

Darling, Timothy, Univ. of Nevada, LANL, MS220, Physics, Reno, NV 89557.

Davidson, Keith L., 600 SW Kenyon St., Apt. K202, Seattle, WA 98106.

Dustin, Carr W., 5108 Rose Quartz Ave., NW, Albuquerque, NM 87114.

Fitzgerald, Matthew, Otolaryngology, New York Univ. School of Medicine, 550 First Ave., NBV-5E5, New York, NY 10016.

Gage, Stuart H., Entomology, Michigan State Univ., 243 Natural Science Bldg., East Lansing, MI 48824.

Gervais, Rod A., Sound Solutions, Box 125, 18 Lester Dr., Central Village, CT 06332.

Giguere, Christian, Univ. of Ottawa, Audio. to Speech Lang. Path. Prog., 451 Smyth Rd., Ottawa, ON K1H 8M5, Canada.

Giles, Troy, Zounds, Inc., 1630 S. Stapley Dr., Ste. 100, Mesa, AZ 85204.

Good, James A., Veneklasen Associates, 1711 16th St., Santa Monica, CA 90404.

Green, David N., AWE Blacknest, Seismology, Brimpton, Reading, Berkshire RG7 4RS, U.K.

Habib, Joseph M., Audio Lab., 6th of October City, 2nd District, 5th Division, 1172 Michail El Masry Villa, Giza, Cairo 12451 Egypt.

Haiat, Guillaume, Lab. de Mecanique Physique, Univ. of Paris 12, B20A, 61 Ave. du gen de Gaulle, Creteil 94010 Cedex, France.

Headrick, Robert H., Office of Naval Research, Code 32, 875 North Randolph St., Suite 1425, Arlington, VA 22203-1995.

Hertz, Susan R., 24 Highgate Cir., Ithaca, NY 14850.

Hesedahl, Jeanette, 783 Hillrise Pl., NE, Bremerton, WA 98311.

Hosokawa, Atsushi, Dept. of Electrical and Computer Eng., Akashi National College of Technology, 679-3 Nishioka, Uozumi, Akashi, Hyogo, 674-8501, Japan.

Hsu, Hanson K., Delta H. Design, 810 South Flower St., Suite 901, Los Angeles, CA 90017.

Hughes, Michelle L., Ctr. for Childhood Deafness, Boys Town National Res. Hospital, 425 North 30th St., Omaha, NE 68131.

Hurlebaus, Stefan, Zeachry Dept. of Civil Eng., Texas A&M Univ., 3136 TAMU, College Station, TX 77843-3136.

Jacobs, Gordon E., Polysonics Corp., 217 East Main St., Remington, VA 22734.

Kamata, Hirofumi, Dept. of Psychology, Sophia Univ., Kio-cho 7-1, Chiyoda, Tokyo 102-8554, Japan.

Kim, Cheol-An, Medison, R&D, Medison Blvd., 1003, Deachi-dong, Gangnam-gu, Seoul 135-280, South Korea.
Kupersztoch, Daniel J., JEAcoustics, 1705 West Koenig Ln., Austin, TX 78756.
Laible, Henry A., 2607 31st Ave., West, Seattle, WA 98199.
Lally, Emily S., 197 East 3rd St., Apt. 16, New York, NY 10009.
Leasure, Joshua D., Veneklasen Associates, 1711 16th St., Santa Monica, CA 90404.
Leembruggen, Glenn A., Acoustic Directions, P.O. Box 205, Summer Hill, NSW 2130, Australia.
Lin, Ying-Tsong, Applied Ocean Physics and Eng., Woods Hole Oceanographic Inst., 210 Bigelow Bldg., MS #11, Woods Hole, MA 02543.
Lissek, Herve, Ecole Polytechnique Federale de Lausanne, Lab. D'Electromagnetisme Et D'Acoustique, EPFL Sti Itop Lema, Station 11, Lausanne, CH-1015, Switzerland.
Liu, Yi, 3531 Sherburne Ln., Apt. C, Indianapolis, IN 46222.
McMurray, Bob, Dept. of Psychology, Univ. of Iowa, E11 SSH, Iowa City, IA 52242.
Miller, Gordon M., 1651 Castle Rd., North Branch, MI 48461.
Moretti, David J., Naval Undersea Warfare Ctr., Ranges, Engineering and Analysis, Bldg. 1351, Newport, RI 02841.
Nelson, Misty D., 9 Blair Ln., Falmouth, MA 02540.
Park, Joo-Bae, Hankook Tire Co., Ltd., NVH Research Team, 23-1 Jang-Dong, Yuseong-gu, Daejeon, Choong Nam, 305-725, South Korea.
Petnikov, Valery G., A.M. Prokhorov General Physics Inst., Russian Academy of Sciences, Wave Research Ctr., Vavilov St. 38, Moscow 119991, Russia.
Raven, Paul A., 3445 Uplands Dr., Unit 137, Ottawa, ON K1V 9N5, Canada.
Salter, Ethan C., Charles M. Salter Associates, Inc., 130 Sutter St., San Francisco, CA 94104.
Saunders, Melvin, Saunders and Associates, 3467 Forest Ln., Suite 124, No. 1, 203, Dallas, TX 75244.
Shah, Amee P., Speech and Hearing, Cleveland State Univ., 2121 Euclid Ave., MC 431-B, Cleveland, OH 44115.
Shim, Taebo, Soongsil Univ., Information, Communication, and Electronics, 1-1 Sang-do-Dong, Dongjak-gu, Seoul 156-743, South Korea.
Singh, Balbir, Flat No. 7D, Block F, Saket, New Delhi, 110017, India.
Sisneros, Joseph A., Univ. of Washington, Psychology, Guthrie Hall Box 351525, Seattle, WA 98195.
Sporer, Thomas R., IDMT, Fraunhofer Society, Ehrenbergstr. 29, Ilmenau 98693, Germany.
Steiner, Erik A., Steiner Hearing Instruments Ltd., 10 Abbas St., P.O. Box 9400, Haifa 31094, Israel.
Van de Ven, Thomas E., 1835 Skidaway Rd., No. 202, Savannah, GA 31404.
Warusfel, Olivier, IRCAM, Room Acoustics Team, 1 Place Stravinsky, Paris 75011, France.
Wilkinson, Matthew J., 115 Sedgwick St., Cambridge, CB1 3AL, U.K.
Wilson, T. Scott B., Accent Acoustics and Soundproofing, P.O. Box 50935, Eugene, OR 97405.
Wolfe, Joe, School of Physics, Univ. of New South Wales, Sydney NSW 2052, Australia.
Yalaju, Alfred K., CI Assoc. Architects & Consultants, PC, Four Atrium Dr., Suite 280, Albany, NY 12205.
Yargus, Michael W., 8035 Banner Rd., SE, Port Orchard, WA 98367.
Zeddies, David G., Univ. of Maryland, C-CEBH, Biology, Biology-Psychology Bldg. 144, College Park, MD 20742.
Zhang, Xian Xian, Texas Instruments, 12500 TI Blvd., P. O. Box 660199, MS 8649, Dallas, TX 75266-0199.
Zhou, Yufeng, ImaRx Therapeutics Inc., 1635 East 18th St., Tucson, AZ 85719.

New Students

Akar, Ali O., Undersea Warfare, Naval Postgraduate School, 1 University Circle, Monterey, CA 93943.
Andrews, Mark R., 43 Saint Paul St., Apt. 1, Brookline, MA 02446.
Aroutiounian, Lilia, Teachers College, Columbia Univ., Biobehavioral Sciences, 520 West 120 St., New York, NY 10025.
Berger, Michael A., 191 Raleigh St., Rochester, NY 14620.
Bergevin, Christopher, 16 Greenough Ln., No. 103, Boston, MA 02113.
Bishop, Jason B., English Language and Literature/Linguistics, Eastern Michigan Univ., 202 Welch Hall, Ypsilanti, MI 48170.
Blaszak, Magdalena, Adam Mickiewicz Univ., Acoustics, Umultowska 85, Poznan, Wielkopolska 60-414, Poland.
Bosseler, Alexis N., Inst. of Learning and Brain Sciences, Speech & Hearing Sciences, Univ. of Washington, Seattle, WA 98195.
Bradley, Dunkin, Graduate Program in Acoustics, Pennsylvania State Univ., P.O. Box 30, State College, PA 16802.
Bresch, Erik, Electrical Engineering, USC, 3740 McClintock, Room 400, Los Angeles, CA 90089-2564.
Bridges, Brian D., National Univ. of Ireland, Maynooth, Music Dept., Maynooth, Ireland.
Carbo, Randy M., 331 7th Ave., 2nd Floor, Troy, NY 12182.
Case, Julie E., 365 W. 20th St., Apt. 7C, New York, NY 10011.
Cerka, Amanda J., Northwestern Univ., Communication Sciences and Disorders, 2240 Campus Dr., Evanston, IL 60208.
Chi, Zuemin, Massachusetts Inst. of Technology, 77 Massachusetts Ave., Rm. 36-549, Cambridge, MA 02139.
Chiu, Yung-Sheng, National Taiwan Univ., Engineering Science and Ocean Eng., No.1, Sec. 4, Roosevelt Rd., Taipei 106, R.O.C. Taiwan.
Cimmino, Gary A., 18 Ashley Park Dr., North Branford, CT 06471.
Covato, Andrew R., 14 Hamptonbrook Dr., Toronto ON M9P 1A1, Canada.
Crone, Timothy J., School of Oceanography, Univ. of Washington, Box 357940, Seattle, WA 98195-7940.
Dalberg, James L., 329046 Georgia Tech Station, Atlanta, GA 30332.
DeRuiter, Stacy L., Biological Oceanography, MIT/Woods Hole Oceanographic Inst., MS 50, Woods Hole, MA 02543.
Draudt, Andrew B., 35 Orvis Rd., Arlington, MA 02474.
Dutta, Indranil, Univ. of Illinois, Linguistics, 40480 FLB, 707 S. Matthews, Urbana, IL 61801.
Flint, Jesse D., Binghamton Univ., Psychology, P.O. Box 6000, Binghamton, NY 3902-6000.
Gaudrain, Etienne, 8 rue Sainte Catherine, Lyon 69001 France.
Gauthier, Jean-Philippe, Universite de Sherbrooke, Genie Mecanique, 2500 boul. Universite, Sherbrooke, QC J1K 2R1, Canada.
Glesser, Martin, Universite de La Rochelle, LEPTAB, Ave. Michel Crepeau, La Rochelle17000, France.
Goto, Tomoya, Innai 1-9-39, Funabashi-City, Chiba 273-0039, Japan.
Graves, Amanda G., School of Behavioral and Brain Science, Univ. of Texas at Dallas, 811 Synergy Park Blvd., Richardson, TX 75080.
Gubbala, Rabindranath, 457 Park Dr., Apt. 4, Boston, MA 02215.
Gunduz, Mustafa E., 3197 Post Woods Dr., Apt. G, Atlanta, GA 30339.
Gurski, Claire M., French Dept., Univ. of Western Ontario, 1151 Richmond St., London ON N6A 5B8, Canada.
Hanke, Ryan M., 514 West Doty St., No. 208, Madison, WI 53703.
Haywood, Nicholas R., 204 Tullibardine Rd., Sheffield, South Yorkshire, S11 7GQ, U.K.
Herrmann, Frank, Damaschkestr. 21, Bad Duerrenberg, Sachsen Anhalt 06231, Germany.
Hickox, Ann, 8020 E. Garland Rd., Tucson, AZ 85750.
Ho, Connie, Vtech Telecommunications Ltd., 23/F Tai Ping Industrial Centre, Block 1, 57 Ting Kok Rd., Tai Po, Hong Kong, P.R. China.
Holt, Yolanda F., The Ohio State Univ., Speech and Hearing Science, Pressey Hall, 1070 Carmack Rd., Columbus, OH 43210.
Hooper, Stacie L., Univ. of California, Davis, VM: Population Health and Reproduction, One Shields Ave., Davis, CA 95616.
Howell, Eric W., 267 Big Terra Ln., Gurnee, IL 60031.
Huyck, Julia, Northwestern Univ., Communication Sciences and Disorders, 2240 Campus Dr., Evanston, IL 60208.
Ihlefeld, Antje, Boston Univ., Cognitive and Neural Systems, 677 Beacon St., Boston, MA 02215.
Jackson, Scott R., Linguistics, Univ. of Arizona, Douglass Bldg., Tucson, AZ 85721.
Jados, Benjamin J., 2235 Keim Dr., Aurora, IL 60503.
Jain, Rashi, New Jersey Inst. of Technology, Mathematical Sciences, University Heights, Newark, NJ 07102-1982.
Jassar, Sandip S., 172 Canon St., Leicester LE4 6GH, U.K.
Kennedy, Elizabeth A., Audiology, Univ. of Nebraska-Lincoln, 14th and 'R' St., Lincoln, NE 68588.
Koblitz, Jens C., Eichenallee 32a, Neuss NRW 41469, Germany.
Kosko, Andrea J., 1351 West Lobo Ave., Mesa, AZ 85202.

Kreidler, Sandra M., 2066A Park Run Dr., Columbus, OH 43220.
Lambert, Jeremy A., 205 Princeton Ave., No. 4, Amherst, NY 14226.
Landera, Mario A., 3600 Windmeadows Blvd., Apt. 24, Gainesville, FL 32607.
Lau, Kent C., 1125 Bloomdale Rd., Philadelphia, PA 19115.
Lee, Andrian K. C., Auditory Neuroscience Lab., Cognitive and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215.
Lenhart, Erica M., 734 Bear Claw Way, No. 308, Madison, WI 53717.
Lesley-Neuman, Diane F., Linguistics, Univ. of Colorado, Hellums 295–295UCB, Boulder, CO 80309.
Lorenz, Elbereth G., 2301 East University Dr., #108, Mesa, AZ 85213.
Lu, Huancai, 3649 Glenwood Ave., Windsor ON N9E 2Y6, Canada.
Mahaffey, Christie, Mechanical Engineering, Univ. of Maine, 5711 Boardman Hall, Orono, ME 04469-5711.
Marozeauu, Jeremy P., Northeastern Univ., Speech-Language Pathology and Audiology, 360 Huntington Ave., No. 106 Fr., Boston, MA 02115.
Martin, Vanessa A., GeM, Inst. de Recherche en Genie Civil et Mecanique, IUT de St. Nazaire, 58 rue Miche Ange, Saint Nazaire 44600, France.
McPherson, Craig R., James Cook Univ., Electrical Engineering, Townsville, QLD 4811, Australia.
Messina, Andrew, Blessed XXIII National Seminary, 558 South Ave., Weston, MA 02493.
Meyer, Matthias, Gartenstrasse 13, Wiefelstede 26215, Germany.
Migliazzo, Jacob J., 3311 Red River, No. 311, Austin, TX 78705.
Moore, Kristy A., Ocean Engineering, Univ. of Rhode Island, 217 Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882.
Moraveji, S. Javid, Univ. of Michigan, Mechanical Engineering, 2600 Draper Rd., Ann Arbor, MI 48109.
Moreau, Sebastien, 23/25 Boulevard Edgard Quinet, Colombes, 92700, France.
Muller, Marie, Laboratoire d'Imagerie Parametrique, 15 rue de lecole de medecine, Paris 75006, France.
Murray, Stephen R., Physics, UMass Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747.
Naluai, Nathan K., Graduate Program in Acoustics, Pennsylvania State Univ., 217 Applied Science Bldg., University Park, PA 16802.
Nardone, Marco, 128 Biddle St., Bowling Green, OH 43402.
Pape, Nathan A., Univ. of Hartford, Engineering, 200 Bloomfield Ave., Box 1608, West Hartford, CT 06117.
Patel, Akshay M., 2712 SW 34th St., Apt. 164, Gainesville, FL 32608.
Payan, Cedric, LCND, IUT GMP; Ave. Gaston Berger, Ain en Provence, 13625 cedex 1, France.
Penny, Isaac D., Mechanical Engineering, Georgia Inst. of Technology, 801 Ferst Dr., NW, Atlanta, GA 30332.
Petrites, Anthony E., Brown Univ., Ecology and Evolutionary Biology, 80 Waterman St., Box G-W, Providence, RI 02912.
Pinto, Michelle G., 52 Eastern Ave., East Providence, RI 02914.
Proctor, Michael, P.O. Box 207053, New Haven, CT 06520.
Radhakrishnan, Sreeram, Stevens Inst. of Technology, Ctr. for Maritime Systems, 711 Hudson St., Hoboken, NJ 07030.
Romero, Julian P., Flat 22 Matthias Ct., Silk St., Salford, Greater Manchester, M3-6JD, U.K.
Samaran, Flore, CEBC-CNRS, Villiers En Bois, Franca 79360, France.
SampathKumar, AshwinKumar, Boston Univ., Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215.
Sarkar, Bikramjit, Marine Physical Lab., Scripps Inst. of Oceanography, 9500 Gilman Dr., MC 0238, San Diego, CA 92093-0238.
Sastre, Marcos M., Systems, Architecture, Design and Integration Directorate, Raytheon Integrated Defense Systems, 1847 West Main Rd., Portsmouth, RI 02871.
Scarlotta, Theresa M., 9261 West 100th Way, Westminster, CO 80021.
Schwartz, Kyle W., 500 Broce Dr., Apt. 60, Blacksburg, VA 24060.
Sefton, Samantha E., 2236 Victoria Park Dr., Columbus, OH 43235.
Seo, Hojune, Jugong Apt 111-410, Beon 3dong Gangbuk-gu, Seoul 142-063, South Korea.
Skubal, Jacob J., 773 Old Treeline Trail, DeLand, FL 32724.
Smith, Adam R., Biological Sciences, Bowling Green State Univ., 304 Life Science Bldg., Bowling Green, OH 43403.
Smith, Moira J., Grad. Sch. of Arts and Sciences, Harvard Univ., University Hall, Cambridge, MA 02138.
Snow, Bill, 117 Harrison St., No. 1, Oak Park, IL 60304.
Sonstrom, Kristine E., 150 Swamp Rd., Coventry, CT 06238.
Speciale, Valerie N., Bethel College, Biology, 1001 West McKinley, Mishawaka, IN 46544.
Stamenovic, Marko S., 660 Massachusetts Ave., No. 1, Boston, MA 02118.
Sumoro, Hadi, 716 Washington Place, Apt. 507, Baltimore, MD 21201.
Sung, Woo, Loyola Law School, 13334 Mapledale St., Norwalk, CA 90650.
Sylvestre-Williams, Nicholas, Air and Noise, Pinchin Environmental, 5749 Coopers Ave., Mississauga, ON L4Z 1R9, Canada.
Szczepaniak, Olimpia M., Adam Mickiewicz Univ., Physics, Umultowska 85, Poznan, Wielkopolska 60-414, Poland.
Taghvaei, Navid, Psychological and Brain Sciences, Univ. of Louisville, 317 Life Sciences Bldg., Louisville, KY 40292.
Tseng, Yao-Ting, Applied Physics, Curtin Univ. of Technology, GPO Box U1987, Perth, WA 6845 Australia.
Vargas, Hillary S., 33 Azalea Court, North Kingstown, RI 02852.
Vasil, Kristin A., 33 Baxter Rd., Unit 2E, Willington, CT 06279.
Wallace, Andrew B., Brown Univ., Cognitive & Linguistic Sciences, Box 1979, Providence, RI 02912.
Wells, Blake M., 5977 Feather Ln., Sanford, FL 32771.
Wu, Tianming, Boston Univ., Aerospace & Mechanical Engineering, 110 Cummington St., Boston, MA 02215.
Zhao, Sherry Y., Research Lab. of Electronics, Massachusetts Inst. of Technology, 77 Massachusetts Ave., Rm. 36-545, Cambridge, MA 02139.

New Electronic Associates

Arnon, Osnat, 15 Tschernichovsky St., Jerusalem, 92581, Israel.
Bento, Pedro S., R. Hardim 678-Dir. Traseiras, Vilas do Paraiso VNG, 4405-824 Portugal.
Birkett, Stephen H., Systems Design Engineering, Univ. of Waterloo, 200 University Ave., West, Waterloo, ON N2L 3G1, Canada.
Blair, Jerome J., 3495 East Tompkins Ave., Las Vegas, NV 89121-5912.
Boehmer, Gerd G., Dept. Physiology and Pathophysiology, Johannes Gutenberg Univ., Duesbergweg 6, Mainz RPF D-55099, Germany.
Booth, Cormac, Univ. of St. Andrews, The Observatory, Buchanan Gardens, St. Andrews KY1 69LZ, U.K.
Bruno, Giordano L., CIRMMT-Schulich Sch. of Music-McGill Univ., 555 Sherbrooke St. West, Montreal, QC H3A 1E3, Canada.
Bullock, Bruce, Bruce Bullock and Associates, 79 Atwood Ln., Chatham, MA 02633.
Burkholder, Rose A., 1136 Surrey Ln., Lyon, MI 48178.
Canestaro, Christopher L., Penguin Jungle, 2706 Oak Crest Ave., Austin, TX 78704.
Castellote, Manuel, L'Oceanografic of Valencia, Research, Junta de murs i valls s/n, Valencia 46013, Spain.
Cervera-Crespo, Teresa, Literato Gabriel Miro 49,4, Valencia 46008, Spain.
Cheolung, Cheong, Korea Res. Inst. of Standards and Science, Div. of Physical Metrology, 1 Doryong-Dong, Yuseong-Gu, Daejeon, 305-340, South Korea.
Cox, Rupert A., Manchester Univ., Social Anthropology, Roscoe Bldg., Brunswick St., Manchester, M139PL, U.K.
Davenny, Ben C., Acentech, 33 Moulton St., Cambridge, MA 02138.
Davis, Penny R., Biology Dept., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431.
Elizalde-Dounce, Roberto, Ford Motor Company, CAD/CAE, G. Gozalez Bocanegra 1500, Mexico 01210, Mexico.
Gausland, Ingebret, Kveinveien 13, Stavanger, Rogaland 4017, Norway.
Geerits, Tim W., Baker Hughes, 2001 Rankin Rd., Houston, TX 77073.
Gruhler, Gerhard, Heilbronn Univ., Max-Planck-Str. 39, Heilbronn, Baden-Wuerttemberg D-74081, Germany.
Guan, Shane, NOAA/NMFS, Office of Protected Resources, 1315 East-West Highway, SSMC-3, Suite 13756, Silver Spring, MD 20910.
Gupta, Amitava, Johnson & Johnson, 2797 Frontage Rd., Roanoke, VA 24017.
Haddock, Steven H., MBARI, 7700 Sandholdt Rd., Moss Landing, CA 95039.
Holcomb, Sara S., 16840 Garfield Rd., Olivet, MI 49076.
Jones, Ossie, Jr., 196-15 120th Ave., St. Albans, NY 11412.
Kato, Kosuke, Matsuzono 10-43, Nisinomiya, Hyogo 662-0053, Japan.
Kirchner, Tobias, Fjardholmsgrand 11B 6Tr, Stockholm, 12744, Sweden.

Kourakata, Itaru, Niigata Univ., Ctr. for Cooperative Research, 8050 Ikarashi 2-nocho, Niigata City, 950-2181, Japan.
Lee, Seung-Jae, Samsung Heavy Industries, Munji-dong, Yuseong-gy 103-28, Daejeon 305-380, South Korea.
Lekner, John, SCPS, Univ. of Wellington, P.O. Box 600, Wellington 6015, New Zealand.
Leroux, Thierry R., ELDIM, R&D, 1185, Rue d'Epron, Herouville Saint Clair, Calvados 14200, France.
Levitsky, Semyon P., Sami Shamoon College of Engineering, Mathematics, Bialik/Basel Sts., Beer-Sheva 84100, Israel.
Maruyama, Xavier K., Naval Postgraduate School, Physics, 1 University Circle, Monterey, CA 93943.
Mellow, Tim J., Nokia UK Ltd., Product Technologies, Summit Ave., Farnborough, Hampshire, GU114 0NG, U.K.
Petersen, Claus M., Carl Bro a/s, Granskoven 8, Glostrup DK-2600, Denmark.
Pompei, Joseph, Holosonics, 400 Pleasant St., Watertown, MA 02472.
Riddle, Davilla E., 20 Manino Cr. No. 103, Kihei, HI 96753.
Rumberg, Martin, Rauhweid 3, Kaiserslautern, Rhineland-Palatinate 67663, Germany.
Samarra, Filipa, The Cottage, 31 South St., St. Andrews, KY16 9QR, U.K.
Sarradj, Ennes, Brandenburgische Technische Universitat, Fak. 3, JP Aeroakustik, Siemens-Halske-Ring 14, Cottbus 03046, Germany.
Saudrais, Eddie, 12 rue du Gabon, Auxerre, 89000, France.
Shabad, Sergey A., 3-iy Baltiyskiy per., 6-3-76, Moscow 125315, Russia.
Shlivinski, Amir, Ben-Gurion Univ. of the Negev, Electrical and Computer Engineering, Beer Sheva 84105, 82105, Israel.
Simpson, Carl G., 16215 Santa Cathrena, San Antonio, TX 78232.
Spitzer, Jaclyn B., College of Physicians and Surgeons, Columbia Univ., Otolaryngology/HNS, 180 Ft. Washington Ave., HP811, New York, NY 10032.
Thomas, Connor L., 12711 Magna Carta Rd., Herndon, VA 20171-2706.
Velazquez, Marvin, SARA, Inc., 6300 Gateway Dr., Cypress, CA 90630.
Wijnant, Ysbrand H., Univ. of Twente, Faculty of Engineering Technology, P.O. Box 217, Enschede, Overijssel 7500AE, The Netherlands.
Williams, Michael T., National Marine Fisheries Service, Protected Resources Division, 222 West 7th Ave., Box 43, Anchorage, AK 99513.
Wright, David, 8010 Englewood, Indianapolis, IN 46240.
Zhou, Ruohua, EPFL, Signal Processing Inst., EPFL-STI-ITS-LTS3 Station 11, Lausanne, CH-1015, Switzerland.
Zoidis, Ann M., Cetos Research Organization, 11 Des Isle Ave., Bar Harbor, ME 04609.

New Corresponding Electronic Associates

Bi, Chuanxing, Inst. of Sound and Vibration Research, Hefei Univ. of Technology, Tunxi Rd. 193, Hefei, Anhui, 230009, P.R. China.
Fernandes, Alberto P., Rua Dona Zulmira 88/204, Rio de Janeiro 20550160, Brazil.
Luo, Ercang, Technical Inst. of Physics and Chemistry, Research Ctr. of Thermoacoustic Technology, Zhongguancun Beiyitao, Beijing 100080, P.R. China.
MyoungWoo, Jan, 112-1504 Hyundai Shinsa-Dong, Eunpyong-gu, Seoul 122-763, South Korea.
Naresh, Mandayam D., Central Leather Research Inst., Biophysics Dept., Sardar Patel Rd., Chennai Tamilnadu 600 020, India.
Zenker, Miguel, Esuela Nacional de Musica/UNAM, Workshop for Musical Instrument Making, Xicotencatl 126, Del Carmen, Coyoacan, Mexico, D.F. 04100, Mexico.
Zhang, Zhiewei, Shandong Univ., School of Physics and Microelectronics, Hongja Lou 5, Kinan, Shandong, 250100, P.R. China.

Reinstated

Shawn L. Makepeace, *Student*

Members Elected Fellows

A. G. Askenfelt, A. U. Case, T. Dau, C. Y. Espy-Wilson, J. G. McDaniel, L. A. Miller, B. Mohl, S. S. Narayanan, S. D. Richards, R. A. Stephen, D. Tang, and G. R. ter Haar

Associates Elected Members

T. R. Beiler, J. S. Bibb, C. Buma, K. J. Bastyr, K. J. Benoit Bird, A. L. Carey, D. Chhetri, D. Clayton, J. L. Cipolla, D. B. Copeland, M. J. Daley, L. Davidson, N. F. Declercq, S. H. Ferguson, F. Farassat, A. M. Farbo, B. W. Gick, J. Guan, R. Hagiwara, J. M. Harte, L. L. Holt, C. Q. Howard, L. Huang, D. Hurley, Y. Jiang, S. Keith, B. J. Kim, M. Kob, R. E. Johnson, J. A. Keefe, G. D. Larson, P. H. Mathuria, R. A. McArdle, J. A. McAteer, A. W. Meyer, D. Moenssen, P. J. Murphy, T. B. Neilsen, B. K. Newhall, A. A. Parthasarathi, N. H. Philipp, M. Pouplier, C. J. Richie, C. B. Roads, T. L. Rogers, D. Rowan, L. S. Smith, J. W. Solway, A. Souslov, G. C. Stecker, R. M. Stern, B. H. Story, O. G. Symko, S. L. Thomson, W. R. Worley, K. D. Van Wyk, R. A. Wright, Y. Xu, and M. K. L. Yeung

Members to Electronic Associates

S. D. Arlinger, D. T. DiPerna, T. O. Lahti, C. Bjerrum-Niese, Y. A. Chen, T. Holman, and C. A. Sammeth

Associates to Electronic Associates

S. P. Bant, M. Bertsch, D. A. Cabrera, A. C. Carballeira, Y. C. Chiang, P. J. Clemins, C. L. Davies, L. Desutter-Grandcolas, R. Drullman, R. M. Glaese, M. J. Greaves, S. Z. A. Hamid, P. D. Herstein, J. E. Joseph, R. H. Keith, J. Kim, S. A. Kleinknecht, B. R. LaCour, B. W. Lewis, G. G. Martin, L. Menard, A. K. Morozov, S. F. Morse, R. A. Moscos, I. Nakano, T. F. Noonan, E. T. Nykaza, R. Patel, C. J. Pollock, M. K. Rami, P. Sabransky, M. Sakurai, M. R. Schuepbach, A. Song, S. P. Y. Stenfelt, R. M. Su, J. M. Suarez, R. Wakeland, R. Wayland, and U. T. Zwicker

Associates to Corresponding Electronic Associates

O. D. Kripfgans, J. Park, and N. P. Zagrai

Associates to Students

M. A. Andrew, T. Beiler, A. G. Deloach, and K. M. Tsiappoutas

Electronic Associates to Associates

J. A. Corey, F. Desharnais, G. M. Hughes, X. Jiang, E. Rambod, S. M. Unger, and E. L. Zechmann

Students to Associates

A. Agwuele, R. Aravamudhan, T. C. Bent, P. T. Calamia, N. F. Declercq, D. R. Feinberg, A. J. S. Ferreira, G. L. Gattemy, K. L. Gee, J. W. Gorr, J. W. Gregory, C. A. Hartnett, N. Kubo, R. H. Lande, P. W. Lentz, D. E. Lewis, W. Li, C. C. Lin, J. Mamou, C. McGinnes, D. J. Meinke, A. C. H. Morrison, E. M. Oleson, S. C. Peng, L. Sui, A. Stimac, M. Sundara, C. Tollefsen, and T. C. Weber

Students to Electronic Associates

P. B. Allen, A. Bell, F. Bunta, G. A. Coudriet, J. Liu, J. L. Loebach, J. R. Motta da Silva, A. Jesse, J. K. McElveen, M. S. Palahanska-Mavrov, A. Salomon, R. G. Sieler, M. Xu, and B. Zeitler

Students to Corresponding Electronic Associates

S. H. Jang, S. Paik, and Y. Zazicioglu

Resigned

R. J. Bobber, and P. Weissler—*Fellows*
A. Beck, R. Carbo-Fite, K. Dibble, T. H. Fay, W. A. Friedl, D. Gardner, G. M. Gerken, R. Heng, E. Kittinger, T. Machac, D. L. Nichols, V. Soudek, H. G. Tillmann, A. J. Tucker, and E. Yogeswaren—*Members*
I. Akio, M. Anderson, L. J. M. Boe, T. E. Burton, R. A. Corrigan, M. Cooper, S. Fischer, S. Gorbatov, J. Hogan, W. J. Jung, J. E. Newton, T. Rivard, and W. F. Wenneman—*Associates*

G. Kennedy, and S. Shakespeare—*Electronic Associates*
A. Almansouri, L. Botto, B. Byron, M. Fung, J. G. Sibley-Schwartz, Y. Ito, T. Lechler, J. A. Monet-Descombey, H. Parrish, and E. Mace–*Students*

Deceased

F. J. Fry, J. Hind. Jr., J. W. Kopec, P. Ladefoged, H. Medwin, J. R. Mundie, and A. Nelkin—*Fellows*
J. R. Arrington, W. C. Bohn, M. E. Dempsey, P. Ebaugh, F. A. Everest, V. P. Garwood, R. H. Gonter, C. G. Gordon, C. W. Koutstaal, M. Lepor, H. Schecter, W. R. Turner, and J. H. Smith—*Members*

Fellows	885
Members	2249
Associates	2535
Students	886
Electronic Associates	583
	7138

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.

June 2006

4–6	**8th International Symposium of 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 (WESPAC 9)**, 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.tuwien.ac.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

6–8	*__2nd International Symposium "Material-Acoustics-Place 2006,"__ Zvolen, Slovakia (Web: www.acoustics.sk/map/)
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)
4–6	*__33rd International Acoustical Conference "Acoustics High Tatras 06"—EAA Symposium__, Štrbské Pleso, Slovakia (e-mail: 33iac@skas.sk; Web: www.skas.sk/acoustics/2006)
11–13	**Annual Conference of the Canadian Acoustical Association**, Halifax, Nova Scotia, Canada (Web: www.caa-aca.ca/halifax-2006.html)
16–17	*__Institute of Acoustics Autumn Conference__, Oxford, UK (Web: www.ioa.org.uk/viewupcoming.asp)
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**, İstanbul, Turkey (Web: www.internoise2007.org.tr)
27–31	**Interspeech 2007**, Antwerp, Belgium (Web: www.interspeech2007.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, (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

30–4	*__Acoustics '08 Paris: 155th ASA Meeting+5th Forum Acusticum (EAA)+9th Congrès Français d'Acoustique (SFA)__, Paris, France (Web: www.acoustics08-Paris.org)

July 2008

7–10	*__18th International Symposium on Nonlinear Acoustics (ISNA18)__, Stockholm, Sweden (temporary e-mail: ⟨BengtEnflo⟩ benflo@mech.kth.se)
28–1	**9th International Congress on Noise as a Public Health Problem**, Mashantuket, 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)

 0001-4966/2006/119(6)/3503/1/$22.50

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 Source Localization

Arthur N. Popper and Richard R. Fay

Springer, New York, 2005. 330 pp. Price $129.00 (hardcover). ISBN 038724185X.

This latest addition to the Springer Handbook of Auditory Research focuses on the issue of sound source localization from an evolutionary and comparative perspective. According to the editors, "sound source localization is arguably one of the most important functions of the auditory system" (p ix). The importance of this process, both in terms of localizing sounds themselves as well as acoustic guiding of the eyes towards novel or important environmental events, suggests that mechanisms of source localization can profitably be examined in a wide range of animal species to search for common mechanisms. The complexity of such a search derives from the fact that ears of different animal groups differ greatly in structure and operation. The diversity of sound receiving structures is all the more daunting for studying localization because only a few potential cues for source localization are available (differences in interaural time and intensity, spectral cues, and the axis of acoustic particle motion). How, then, do different types of ears process these cues, and what is the comparative efficiency of different kinds of peripheral processing? Given differences in peripheral processing, what, if any, are the underlying similarities in subsequent central processing? What are the commonalities in localization performance across invertebrate and vertebrate animals, and can similarities in performance help us devise realistic computational models for source localization? The eight chapters in this volume (including the editors' introduction) address these questions by critically examining behavioral and physiological evidence for sound source localization in a variety of invertebrate and vertebrate species. The compilation of chapters from different animal groups with different kinds of ears into one volume not only aids interspecies comparisons but also facilitates development and evaluation of general computational models of localization. The perspective of these chapters is wide, with most including data from acoustic, behavioral, physiological, and modeling approaches.

The structure and operation of different kinds of peripheral receivers are detailed in the chapters on insects (Robert), fishes (Fay), and nonmammalian tetrapods (Christensen-Dalsgaard). Two general types of receivers—pressure receivers and pressure gradient receivers—are found in different animal groups. Robert shows that in insects these two basic types exist in many variations. Tympanal ears that function as pressure receivers are typically found on insects that are relatively large in body size (noctuid moths, for example) compared to the sound frequencies they must localize. The nontympanal, flagellar ears in such insects as mosquitoes, caterpillars, and crickets may act as particle velocity receivers, similar in operation to those found in fishes. As Fay discusses, the ears of all species of fish function as inertial accelerometers, responding directly to particle motion in both the near and far fields. In some species, called hearing specialists, pressure detection via a swimbladder also occurs. Christensen-Dalsgaard argues that a pressure gradient ear is the primitive condition in all nonmammalian tetrapods, while a pressure sensitive ear is a later specialization. A pressure gradient ear can produce a directional response at low frequencies. But because the two ears are coupled through the respiratory pathway, distinct monaural cues for sensing direction are not present. The evolution of a pressure sensitive ear, in which the two ears are not acoustically coupled, facilitates binaural differentiation of high frequencies. As Fay and Popper point out in their introductory chapter, mammals are the only vertebrate group without the ability to sense direction by pressure gradients. Apparently this ability was lost over evolutionary time in preference to the advantages gained from high-frequency hearing.

The diversity of peripheral acoustic receivers poses the greatest challenge for identifying general central processing mechanisms for computation of interaural difference cues. The most influential model of source localization has been the cross-correlation model of Jeffress, which computes interaural time differences on the basis of a series of anatomical delay lines converging on central coincidence detectors. Trahiotis, Bernstein, Stern, and Buell in their chapter show that certain phenomena in binaural hearing, based primarily on data from human listeners, are consistent with the operation of an interaural correlation mechanism. Colburn and Kulkani discuss recent variations of the original Jeffress cross-correlation model for processing of sound azimuth, including additions to the model that could handle interaural intensity differences. These authors briefly reference physiological and anatomical evidence for cross-correlation processing, topics that are more fully addressed by Christensen-Dalsgaard in his discussion of central mechanisms of binaural processing in amphibians, reptiles, and birds, and by Kubke and Carr in their analysis of the anatomical and biochemical development of binaural pathways. Both chapters argue that the functional organization of brainstem auditory pathways in avians (barn owls and chickens) is highly consistent with the operation of a Jeffress cross-correlation mechanism, while binaural processing in the mammalian auditory system may operate by a somewhat different mechanism. As Christensen-Dalsgaard points out, computation of interaural cues in animals using a pure or a mixed pressure gradient ear probably occurs by a different mechanism, as yet unspecified.

Further complexities for our understanding of binaural processing are addressed in the chapters by Brown and May on behavioral aspects of sound localization in mammals, and again by Colburn and Kulkani in their presentation of models of sound localization, again based primarily on data from mammalian species. Both chapters explicitly discuss not only the importance of interaural cues, but also the importance of monaural spectral cues for localizing sound sources. Brown and May point out that the relative importance of different cues for source localization is a function of the different ecologies and life histories of different mammalian species. Animals may perceive acoustic distance or depth, but little research has focused on how animals integrate the various cues available to them to localize sources in complex, natural acoustic environments. Colburn and Kulkani point to the absence of well-developed models for localization in both distance and direction as a major gap in the existing literature.

All chapters in this volume give a balanced yet critical overview of the literature, with references current up to 2003. The authors point out areas where further empirical evidence is needed. This latter feature makes the volume a valuable resource for graduate and postdoctoral students devising independent research projects, and for use in graduate seminars. The production quality is high, with even anatomical figures being reproduced clearly.

ANDREA MEGELA SIMMONS
Departments of Psychology and Neuroscience,
Brown University,
Providence, Rhode Island 02912

 0001-4966/2006/119(6)/3505/1/$22.50

Acoustics of Wood (2nd edition)

Voichita Bucur

Springer, 2006. 393 pp. Price: $279 (hardcover). ISBN: 3540261230

The first edition of this book was published by CRC Press in 1995. For a review, see J. Acoust. Soc. Am. **109**, 1274 (2001). Let us denote the first edition by AW1 and the second edition by AW2.

The scope and format of AW2 are the same as AW1. Thus, AW2 covers many aspects of wood and its acoustical properties. Also, AW2 is neither a textbook nor a research monograph. It is written as an extended survey, giving brief descriptions of the cited literature, together with many tables and figures.

We should expect that a second edition is better than a first edition, so let us compare AW2 with AW1. To begin, we notice that AW1 cost $265 for 284 pages, so that AW2 seems to be better value for money. (This comparison is not quite so clear because each page in AW2 contains less text.)

I noticed the following improvements. AW2 contains an excellent bibliography, with about 1000 items arranged alphabetically at the end of the book. AW1 had about 850 items, with a separate bibliography for each chapter. The figures and tables are generally clearer and larger in AW2. Most chapters have been updated, some slightly, some substantially: Notable additions include a section (12 pages) on nonlinear effects, much new material in the chapter on "Acoustic Emission," and a new chapter (18 pages) on "Acousto-Ultrasonics." Also, each chapter in AW2 ends with a short summary.

The typography in AW2 is similar to that in AW1. This can be counted as a lost opportunity, because many formulas look ugly, and equations could have been renumbered [there is no good reason to have Eq. (4.18) followed by Eq. (4.18′), for example]. However, I found instances where a correct formula from AW1 has become an incorrect formula in AW2. Some examples of this follow.

Let me give a sample of errors noted, some of which are mere irritations. On p. 24, "Er" in the text becomes E_r in a displayed equation: There are other similar examples. In AW2, the product of two scalar quantities is often written as $a \cdot b$ or $a \times b$ when ab is both clear and simpler. In Eq. (4.9), replace E_{11} by C_{11} (new error). The bottom half of p. 44 is garbled. Below Eq. (4.23), it is not necessary for all the stiffnesses and compliances to be positive. The discussion around Eq. (4.28) is garbled (new error). "Hook's law" above Eq. (4.32) and above Eq. (6.25). There are subscript errors in Eq. (4.37). Equation (4.62) is for V, not v (new error). Amusingly, "Waltham, MA" in the caption to Fig. 4.15 in AW1 has become "Waltham, Maryland" in AW2. In Eq. (4.67), "a" should be "α" (new error). In Table 5.8, D, E, and F should be d, e, and f, respectively (new error). On p. 157, "Hutchin's instruments" (new error). On p. 318, why do the integral signs have a line through them (as commonly used to denote principal values)? Eq. (5.20) in AW1 is correctly given as the quartic

$$MC_{ij}^4 + NC_{ij}^2 + P = 0;$$

in AW2, it has become a quadratic,

$$\mathrm{M} \cdot \mathrm{C_{ij}} + \mathrm{N} \cdot \mathrm{C_{ij}}^2 + \mathrm{P} = 0.$$

Note also that, throughout AW2, quantities such as n_1^2 appear as both $n_1{}^2$ and $n^2{}_1$, whereas V_{QL}^2 appears as $V^2{}_{\mathrm{QL}}$ and C_{ij}^2 appears as $C_{ij}{}^2$: to my eyes, such things are ugly!

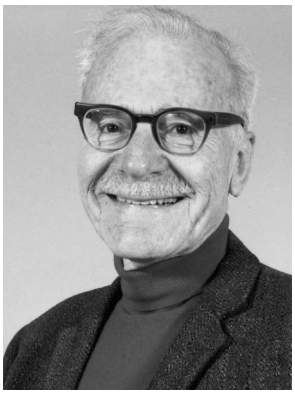
To summarize, I found the textual changes to be poor, with more clutter and less clarity, and the introduction of (mainly minor) new errors. There are also substantial positive changes, as mentioned above. Despite my complaints, this remains a uniquely valuable book because of the way that the author has collected, organized, and reviewed the scattered literature on the *Acoustics of Wood*. It is likely that its value will increase because of the growing interest in all kinds of anisotropic and inhomogeneous materials, including wood.

PAUL A. MARTIN
Department of Mathematical and Computer Sciences,
Colorado School of Mines,
Golden, CO 80401–1887

 0001-4966/2006/119(6)/3506/1/$22.50

OBITUARY

Herman Medwin ● 1920–2006



Dr. Herman Medwin, Fellow, Silver Medalist in Acoustical Oceanography, Gold Medalist, and past President of the Acoustical Society of America, Emeritus Professor at the Naval Postgraduate School, and Founder and CEO of Ocean Acoustics Associates in Pebble Beach, passed away peacefully in his sleep on the afternoon of Monday, January 9, 2006 at the Carmel Hills Care Center, Monterey, California. He had suffered a stroke three months earlier. He was 85.

Herman (Hank) Medwin was born in Springfield, Massachusetts in April 1920. He attended Worcester Polytechnic Institute where he first discovered that he had a special interest in the subject of sound and vibration. Following graduation with a B.S. in Physics in 1941, he worked at Pratt and Whitney's engine test and inspection group and enrolled, part time, in Smith College to further his education. He was the first male student of Smith College. He joined the United States Army Air Corps in 1943 and was assigned to the European Theater of Operation in United Kingdom as a Weather Observer. There he met and married Eileen Huber, from Vienna, Austria, who was serving with the First Allied Platoon of the British Army at the time.

After their return to the United States, Medwin, in the Fall of 1946, enrolled as a graduate student in the Department of Physics at the University of California at Los Angeles (UCLA) where he pursued graduate study in acoustics under an elite academic group of acousticians including Isadore Rudnick, Robert W. Leonard, Vern O. Knudsen, Carl Eckart, and Leo P. Delsasso. While attending graduate school, he was also working as a full-time instructor in Physics at Los Angeles City College. His doctoral research encompassed both theoretical development and experimental verification, extending Eckart's theory for acoustic streaming in fluids. UCLA awarded him a M.S. in Applied Physics in 1948 and a Ph.D. in Physics in 1954. Between 1954 and 1955, he worked at Bolt, Beranek, and Newman, Inc. as a Consultant in Acoustics.

Medwin joined the Naval Postgraduate School (NPS) in 1955 where he began a highly distinguished, productive, and visionary career as an educator and scientist, teaching ocean acoustics to naval officers of the United States and from allied countries, and conducting research in the use of sound waves to measure and learn about ocean physical and biological processes. His research, always involving student participation, included carrying out the very first acoustical measurements of microbubbles in the laboratory and at sea using *in situ* backscatter, attenuation, and dispersion, developing inexpensive scaled laboratory models to understand the physics of sound scattering by a seamount and by a rough sea surface, and providing the first experimental attribution of the Knudsen sea noise spectrum to bubbles due to breaking waves. These are just a few examples of his ample and extensive research contributions.

During this period, he did acoustical consulting for several organizations, including Stanford Research Institute, the Human Resource Research Organization of George Washington University, and the Hudson Laboratories of Columbia University, Dobbs Ferry, NY. In 1961–1962, he served as Liaison Scientist at the U.S. Office of Naval Research, London, England. In 1965–1966, Medwin spent a sabbatical year as a Visiting Scientist at Imperial College, London, England. A six-month sabbatical was spent as a Visiting Scientist at the Royal Australian Naval Research Laboratory, Sydney, Australia in 1972–1973.

Although Medwin retired from NPS in 1980, he had not stopped serving and contributing to NPS and the scientific community in general. Since his so-called retirement, he generously devoted much of his time and energy to continue to educate and serve the scientific community. At NPS, he continued to supervise thesis students, mentor and collaborate with young professors, and most notably, published two textbooks as lead author, one in 1998 and the latest, Sounds in the Sea: From Ocean Acoustics to Acoustical Oceanography, in 2005. These two textbooks, together with an earlier one coauthored with C. S. Clay in 1977, truly defined the field of Acoustical Oceanography.

His services to the Acoustical Society of America were equally abundant and important. Between 1974 and 1988, he served as Member or Chairman of numerous administrative committees of the Society. He chaired the Underwater Acoustics Technical Committee from 1975 to 1977. Between 1989 and 1991, he served as Vice President of the Society and Chairman of the Technical Council. In 1989, he founded the Technical Specialty Group for Acoustical Oceanography that later became a Technical Committee, and which he served as Chairman until 1991. In 1992, he served as President of the Society.

Medwin's other accomplishments include support of the Medwin String Quartet at Worcester Polytechnic Institute, and founder of the Medwin Prize in Acoustical Oceanography. In his hometown, he served on the board and was President of the Monterey Chamber Music Society and was known in local music circles for his string violin performances.

In recognition of his many exceptional achievements, in 2001, Medwin was awarded the highest recognition by the Acoustical Society of America, the Gold Medal, for innovative research in ocean acoustics and leadership and service to the Society. Previously in 1977, he had received the Silver Medal in Acoustical Oceanography for his contributions to the understanding of acoustical scattering, absorption, and ambient noise, particularly in relation to the acoustics bubbles in sea. Additionally, Worcester Polytechnic Institute awarded him an Honorary Doctor of Science Degree in 1992.

A "giant" in Acoustical Oceanography, Medwin pioneered methodologies in and strongly advocated the use of sound waves to sense and study the ocean's physical and biological processes. Both the ocean acoustics and oceanography scientific communities have benefited greatly from his vision and dedication in education and research. Perhaps the most cogent summary of his acoustical career is given by the words of Dr. Larry Crum in introducing Medwin at the ceremony for the award of the Gold Medal of the Acoustical Society of America in 2001: "Hank, you are a man of many adjectives. You are creative and talented, persistent and resourceful, unselfish and genial, and, yes, at times, even stubborn and irascible, but most of all, you are generous, generous with your time, your talent and your treasure. You have been our teacher, our leader, our patron. We are honored to have you in our company of acousticians. Indeed, you are "the" quintessential acoustician, and we are proud to honor you with our Gold Medal."

Professor Medwin is survived by his wife, Eileen, and his older brother, Leo, of Los Angeles.

CHING-SANG CHIU
OSCAR B. WILSON
XAVIER MARUYAMA
VAN HOLLIDAY
STEVEN HADDOCK

 0001-4966/2006/119(6)/3507/1/$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:

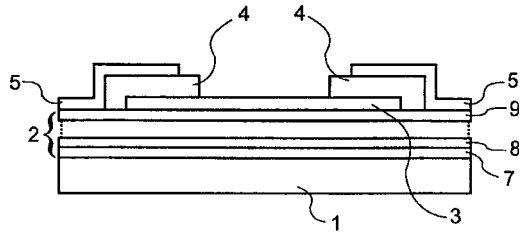
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
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*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
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*

6,975,182

43.20.El FILTER SYSTEM WITH BULK WAVE RESONATOR HAVING A REFLECTION LAYER COMPOSITION THAT VARIES WITH THE LAYER THICKNESS

Hans Peter Loebl *et al.*, assignors to Koninklijke Philips Electronics N.V.
13 December 2005 (Class 333/187); filed in the European Patent Office 14 August 2001

This patent describes a film bulk-acoustic-resonator device that uses quarter-wave stacks beneath itself to achieve decoupling from the substrate within a specified passband. According to the inventors, the resonant materials **4** are best made of oriented piezoelectric materials for increased elec-



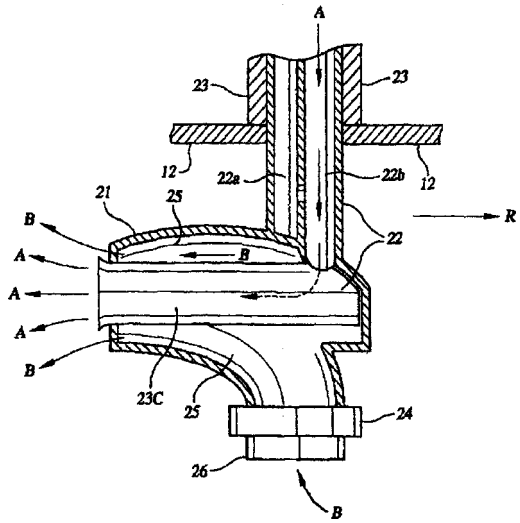
tromechanical coupling. A variety of materials are suggested for the piezoelectric and reflective layers, but no design formulas are given, nor are any measurements of the performance of these devices given. There is little here that appears novel to this reviewer.—JAH

6,976,891

43.30.Nb UNDERWATER SOUND AND BUBBLES FOR POWERED WATERCRAFT

Peter Mele, Crown Point, New York
20 December 2005 (Class 440/38); filed 7 June 2004

A device is described for producing underwater sound effects to assist in alerting swimmers or aquatic animals to the presence of a nonmotorized



watercraft such as a sailboat. The device comprises a housing **21** mounted to the bottom of the sailboat **12**. Fluid flow, indicated by the arrows **B** out of the shaped nozzle end **21**, creates a partial vacuum which draws atmospheric air **A** out of the termination of outlet channel **22**. The air flow produces a jet sound and a visual display of exhaust bubbles. If the watercraft is motionless, fluid flow **B** can be achieved by means of attached hose **26** and a source not shown.—WT

6,980,484

43.30.Sf FISH FINDING DEVICE AND METHOD FOR DETECTING AND DISTINGUISHING MARINE LIFE FROM A SONAR REFLECTED MARKER

Hwang-Hsing Chen, Fort Worth, Texas
27 December 2005 (Class 367/99); filed 14 October 2004

A modified fish finder system consists of a sonar transmitter and receiver, a microprocessor to control the transmitted and received signals, and a sonar reflector/marker. The microprocessor is capable of distinguishing

 0001-4966/2006/119(6)/3509/23/$22.50

echoes from any marine life as well as those from the sonar reflector and converting these two sets of echoes into different signals that can be displayed with different icons so that the operator can move the bait, attached near the sonar reflector, to a more advantageous position near the fish.—WT

6,985,406

43.30.Yj INTERFEROMETRIC SYNTHETIC SONAR ANTENNA

Didier Billon, assignor to Thales
10 January 2006 (Class 367/88); filed in France 19 December 2000

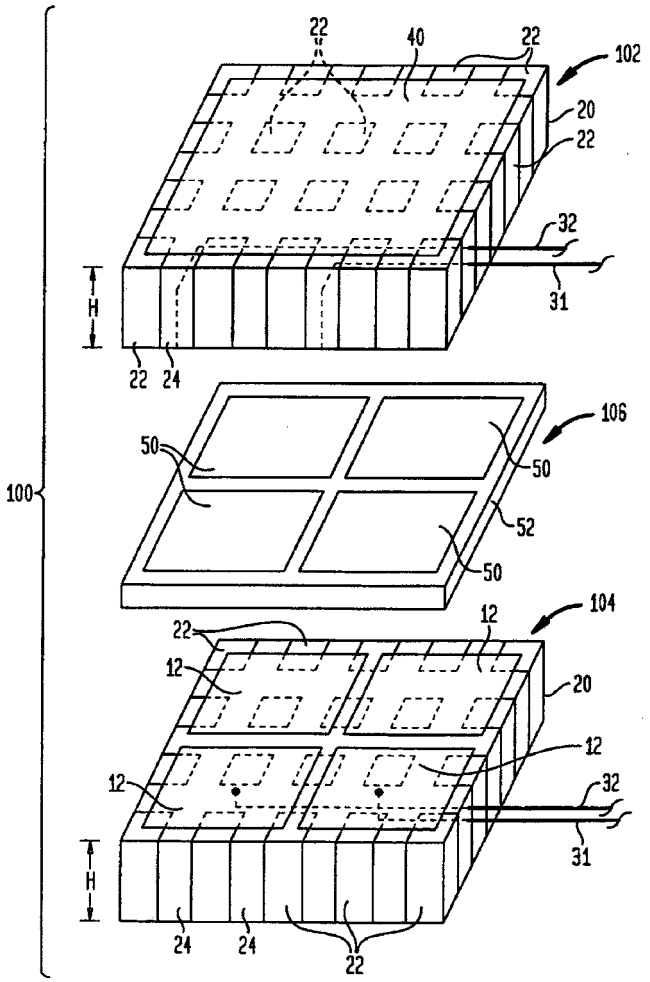
A synthetic sonar receiving array is realized as two parallel line arrays displaced in their length direction by one-half of an individual element spacing. In one mode of operation, the output signals of the two line arrays are processed to form a single beam with higher spatial resolution than that of the two individual arrays. In a second mode of operation, the output signals of the two line arrays are processed to enable the two arrays to be self-calibrated with improved precision.—WT

6,985,407

43.30.Yj MULTI-LAYER COMPOSITE TRANSDUCER ARRAY

Kim C. Benjamin, assignor to The United States of America as represented by the Secretary of the Navy
10 January 2006 (Class 367/153); filed 2 February 2004

The patent describes the construction of a compound wideband transducer intended for sonar applications. The individual sections are tuned to separate frequencies, each operating at its primary resonance.—JME

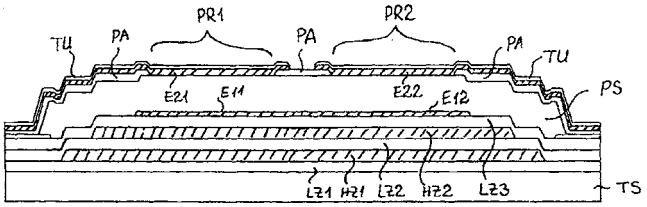


6,985,052

43.35.Ns COMPONENT OPERATING WITH BULK ACOUSTIC WAVES, AND HAVING COUPLED RESONATORS

Pasi Tikka, assignor to EPCOS AG
10 January 2006 (Class 333/189); filed in Germany 13 December 2002

The inventors disclose a film bulk-acoustic-wave resonator having a metal electrode inserted beneath the resonant piezoelectric layer intended to couple input and output electrodes electrically. The nonmechanical coupling path creates additional notches in the transfer function of the filter structure that the inventors maintain is advantageous. Conductive layers **E11** and **E12**



are electrically connected to achieve this, even though the acoustic waves penetrate them and are reflected from the lower acoustic mirrors formed by **LZ1**, **HZ1**, **LZ2**, and **HZ2**. The inventors argue that the effect can usefully be extended to more than two resonators in the same substrate. The patent is brief and contains no data to support the efficacy of this implementation of a ladder network.—JAH

6,975,184

43.38.Ar ADJUSTING THE FREQUENCY OF FILM BULK ACOUSTIC RESONATORS

Li-Peng Wang *et al.*, assignors to Intel Corporation
13 December 2005 (Class 333/188); filed 30 May 2003

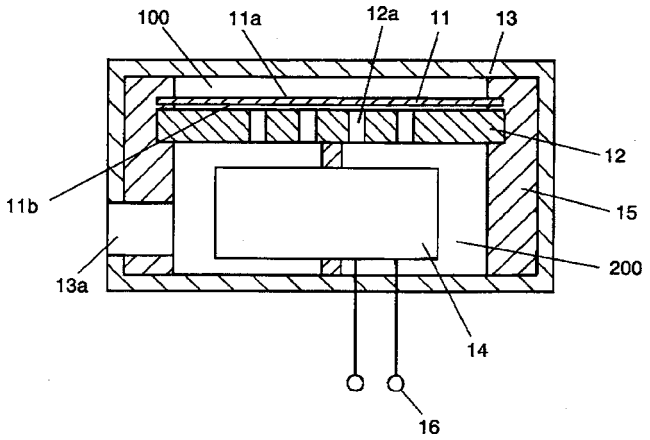
This patent applies to the process of adjusting the resonant frequency of film bulk-acoustic-wave resonators (FBARs) using focused ion-beam milling. It would not surprise most readers to learn that removing mass from an electrode on a piezoelectric resonator raises its resonant frequency and adding mass reduces it. The inventors have discovered that this principle applies even at the nanometer scale of electrodes on an FBAR. Using focused ion beam technology, one can even trim the microscopic resonators used in wireless communications individually while they are still on a common wafer. There is not much of interest here for the acoustically inclined reader.—JAH

6,975,736

43.38.Ar MICROPHONE

Akinori Hasegawa, assignor to Matsushita Electric Industrial Company, Limited
13 December 2005 (Class 381/174); filed in Japan 19 July 2002

This patent describes what is in effect the typical capillary tube in a pressure microphone, which equalizes the pressure difference on the diaphragm at very low frequencies. In this case, the size of the hole has been



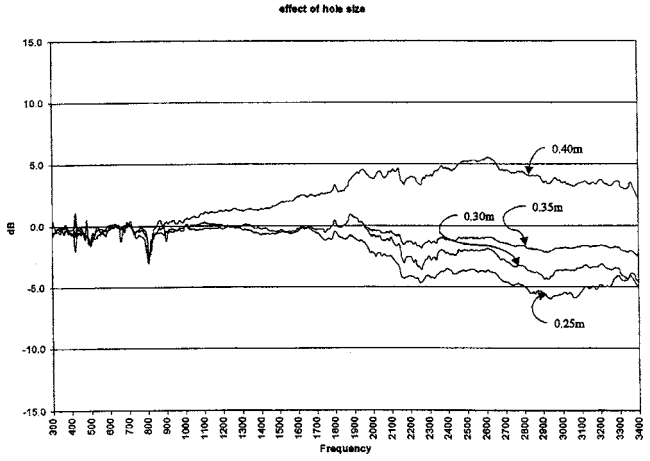
adjusted to minimize the effects of blasts when the microphone is used at close quarters. This treatment will of course change both the sensitivity of the microphone and its directivity pattern in the lower-frequency range.—JME

6,978,033

43.38.Ar MICROPHONE GASKET WITH INTEGRATED ACOUSTIC RESISTANCE

Philippe Moquin and Stephane Dedieu, assignors to Mitel Networks Corporation
20 December 2005 (Class 381/355); filed in the United Kingdom 20 February 2001

In the integration of electret microphones into handsets, cell phones, and other such devices, the acoustical path between the talker and the receiver may go through a number of small bends and turns that may ad-



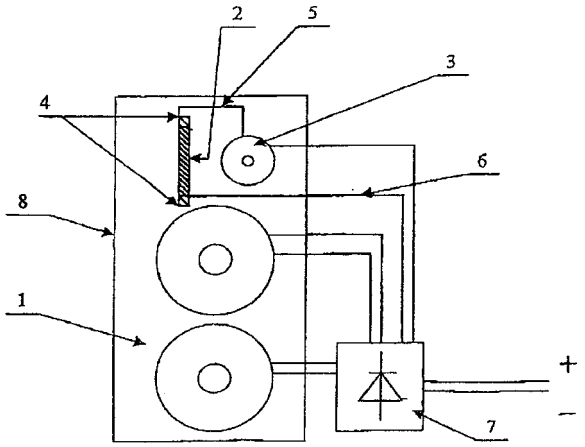
versely affect frequency response. The patent describes the use of a resistive acoustical mesh integral to one of the gaskets or fittings involved to damp the major response peak in the 1500–3000 Hz octave.—JME

6,944,306

43.38.Dv HARMONIC AMPLIFIER AND CORRESPONDING ELECTRO-ACOUSTIC TRANSDUCER

Marc Charbonneaux *et al.*, assignors to Pica-Sound International
13 September 2005 (Class 381/117); filed in France 18 August 1998

This very short patent comprising two figures and less than two pages of text states that the addition of what is, in essence, an air core or polymer



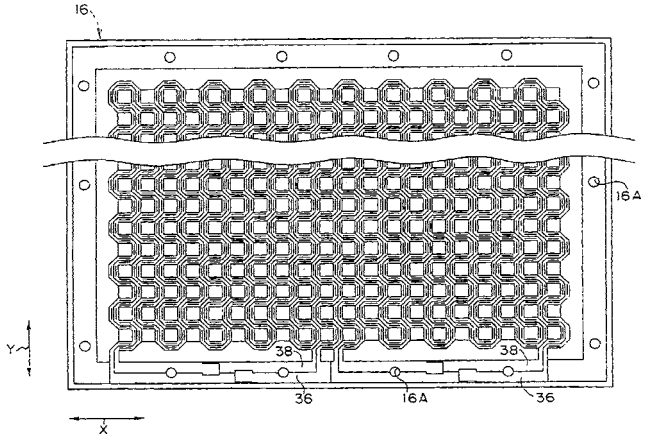
core inductor **2** of practically any shape, and which in the patent is called a harmonic amplifier, is said to "superimpose on the audio sound refined sound information contained in a source signal of the audio sound."—NAS

6,963,654

43.38.Dv DIAPHRAGM, FLAT-TYPE ACOUSTIC TRANSDUCER, AND FLAT-TYPE DIAPHRAGM

Hiromi Sotme and Toshiiku Miyazaki, assignors to FPS Incorporated
8 November 2005 (Class 381/431); filed in Japan 4 October 2001

The patent describes a means for building a flat-panel transducer that is said to reduce the number of through holes needed to implement the current path and also to improve the manufacturability, among other claims.



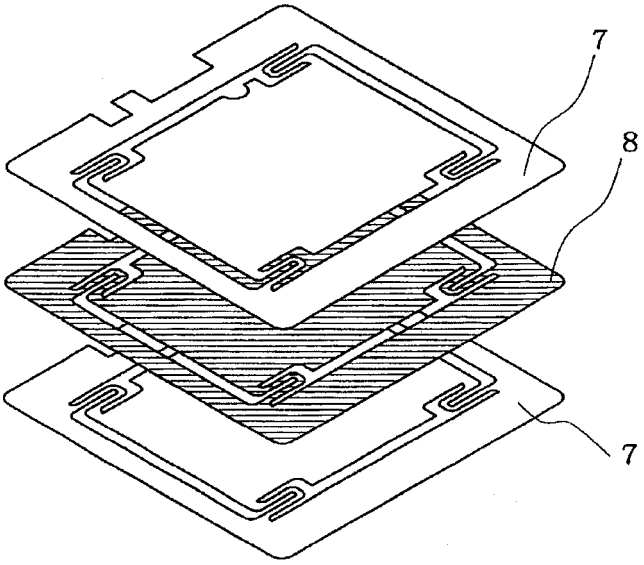
This is done by routing etched conductors **36** and **38** in a serpentine manner so that the path almost surrounds a number of magnets which may be placed at some, but not necessarily all, of the square areas in the figure.—NAS

6,978,032

43.38.Fx PIEZOELECTRIC SPEAKER

Takashi Ogura and Kousaku Murata, assignors to Matsushita Electric Industrial Company, Limited
20 December 2005 (Class 381/190); filed in Japan 29 November 2001

Small piezoelectric loudspeakers are commonly used in portable communications devices. Such a loudspeaker can deliver usable response from about 300 Hz to more than 10 kHz. One design employs a disk-shaped piezoelectric element bonded to a rectangular metal diaphragm. The diaphragm must have a coefficient of thermal expansion roughly equal to that



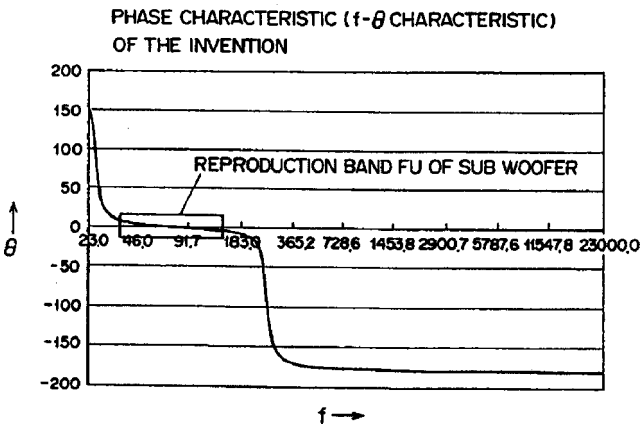
of the piezo element and, for this reason, stainless steel is a desirable choice. Since the stiffness of the diaphragm must be held within fairly close limits, it follows that diaphragm thickness and weight are largely predetermined. This patent discloses an improved sandwich diaphragm structure in which two very thin layers of stainless steel are laminated to a thicker core material. A weight reduction of about 40% is said to be achievable.—GLA

6,955,241

43.38.Ja SPEAKER UNIT FOR LOW FREQUENCY REPRODUCTION

Erik Thorsell *et al.*, assignors to Alpine Electronics, Incorporated
18 October 2005 (Class 181/156); filed 24 January 2005

Adjusting the port and transducer resonances in bandpass-type subwoofer loudspeaker designs is all the rage these days, especially for automobile sound systems. See, for example, Unites States Patent 6,912,290 [reviewed in J. Acoust. Soc. Am. **119**(2), 684 (2006)]. In this patent, which



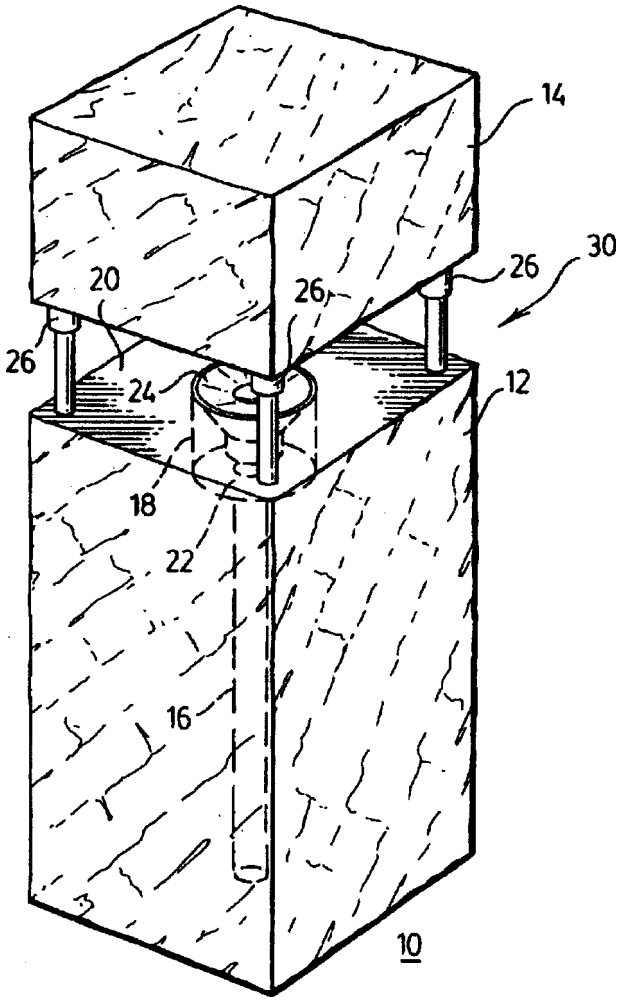
is similar to that referenced, the port resonance is set below the pass band and the transducer resonance is set above the pass band, with the phase characteristic of the assembly through the pass band to be essentially flat. The patent is fairly easy to read.—NAS

6,957,715

43.38.Ja GARDEN SPEAKER

Carl Christiansen, Burlington, Ontario, Canada
25 October 2005 (Class 181/199); filed 21 January 2003

Instead of using concrete or other durable material to manufacture the enclosure for a garden speaker, the present patent describes the manufacture



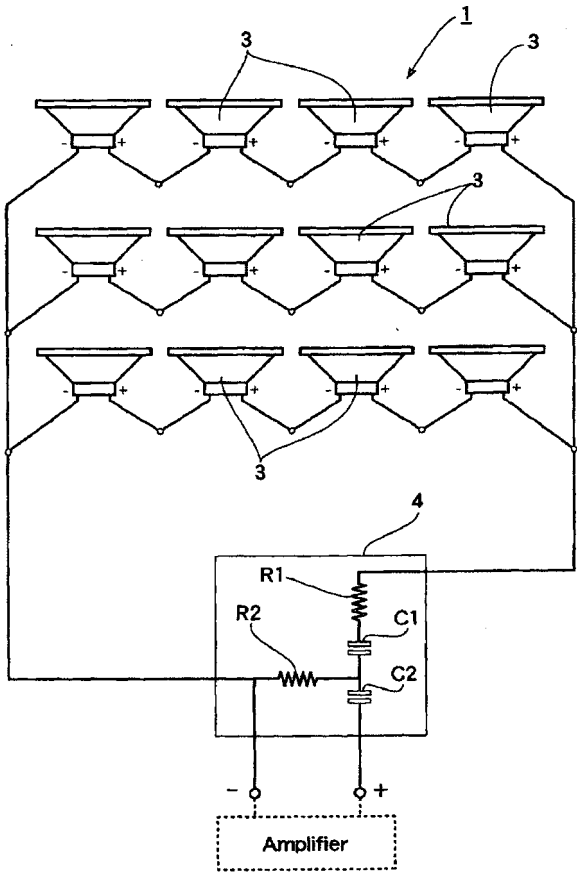
of the enclosure from a solid piece of stone, preferably limestone, that is sawed and cored. Various implementations using different types and numbers of transducers are described in this very short patent.—NAS

6,961,438

43.38.Ja LOUDSPEAKER SYSTEM HAVING WIDE-DIRECTIONAL CHARACTERISTICS

Takeshi Fujita, assignor to Globo Technology, Incorporated
1 November 2005 (Class 381/182); filed 20 December 1999

This is a new addition to the long line of patents that subscribe to the proposition that individual drivers regularly arrayed on the faces of a polyhedron will provide for wide directional dispersion of sound. The patent



notes that along certain directions, the drivers may interact so that "it is difficult to listen to the sounds... ." This difficulty is said to be corrected by the use of RC or LCR networks of the type described in the patent that connect the various loudspeakers together.—NAS

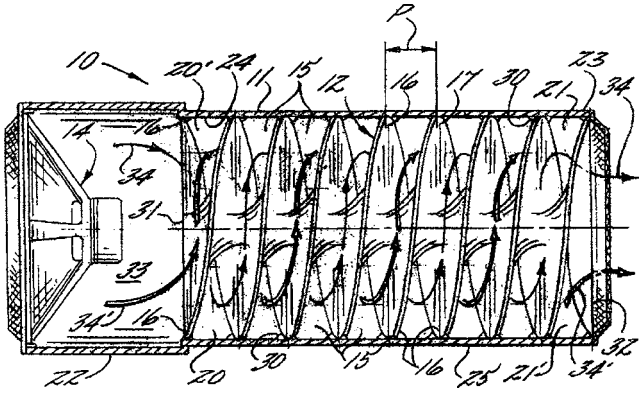
6,973,994

43.38.Ja APPARATUS FOR INCREASING THE QUALITY OF SOUND FROM AN ACOUSTIC SOURCE

Ian J. Mackin, Belmont and William L. Weir, Huntersville, both of North Carolina

13 December 2005 (Class 181/156); filed 4 November 2002

A loudspeaker on a pipe behaves pretty much like a loudspeaker on a pipe, no matter how the pipe is twisted or bent. At least two earlier patents describe a single helical air path, and one earlier patent describes a double helix. In the case at hand, the inventors have chosen to work with a double helix. Nothing about the concept shown appears to be unusual, yet 46 patent claims were granted.—GLA



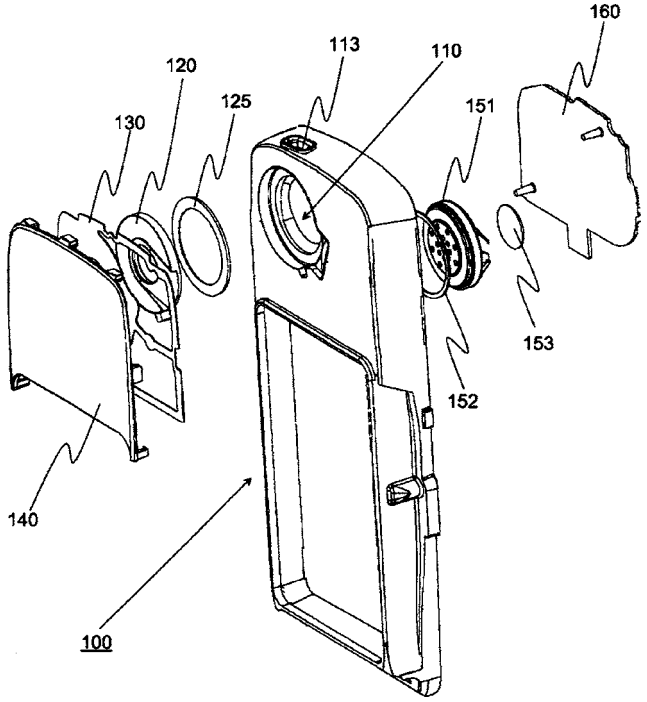
6,975,740

43.38.Ja WATERPROOF ACOUSTIC STRUCTURE APPLICABLE IN CONJUNCTION WITH SPEAKER

Janne Rautio and Juha Petteri Jylänki, assignors to Nokia Corporation

13 December 2005 (Class 381/386); filed in Finland 24 October 2002

Cellular phones may be used in the midst of smoke, dust, rain, or other atmospheric contaminants. The IEC/EN 60529 standard establishes requirements for dustproof and waterproof loudspeakers and it is a goal of this invention to meet the standard without compromising performance. Two shallow subcavities are separated by a waterproof isolating diaphragm **125**.



A miniature loudspeaker drives the first subcavity through an additional cavity and a small air passage. The second subcavity transmits sound to the outside air through a second air passage **113**. By juggling the various acoustical elements, the resulting bandpass filter can cover voice frequencies without introducing loss in the passband.—GLA

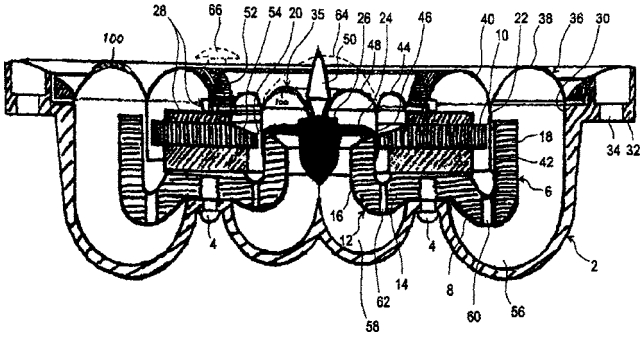
6,980,664

43.38.Ja DOUBLE-DOME SPEAKER

Lars Goller, assignor to Danish Sound Technology A/S

27 December 2005 (Class 381/186); filed in Denmark 4 January 2001

This is a coaxial loudspeaker in which a single ring magnet energizes two concentric gaps housing two concentric voice coils. An "inner membrane part" **35** is driven by the inner voice coil and a "concentrically disposed outer membrane part" **36** is driven by the outer voice coil. The "meet-



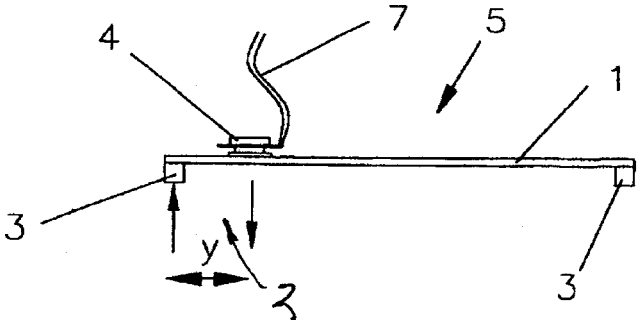
ing area" **28** between the two membrane areas is clamped to the chassis. That brief description is the basis of this patent, but numerous variants are described in the patent claims. In the preferred embodiment shown, both membrane parts take the form of "ring domes," and a central waveguide **50** is included.—GLA

6,985,596

43.38.Ja LOUDSPEAKERS

Graham Bank *et al.*, assignors to New Transducers Limited
10 January 2006 (Class 381/152); filed in the United Kingdom 28 August 1998

During the past few years a considerable amount of work has been done in the development of bending-wave, panel-type loudspeakers. For best overall results, such a panel must be driven in its central area. However, some applications—a transparent display panel, for example—conflict with



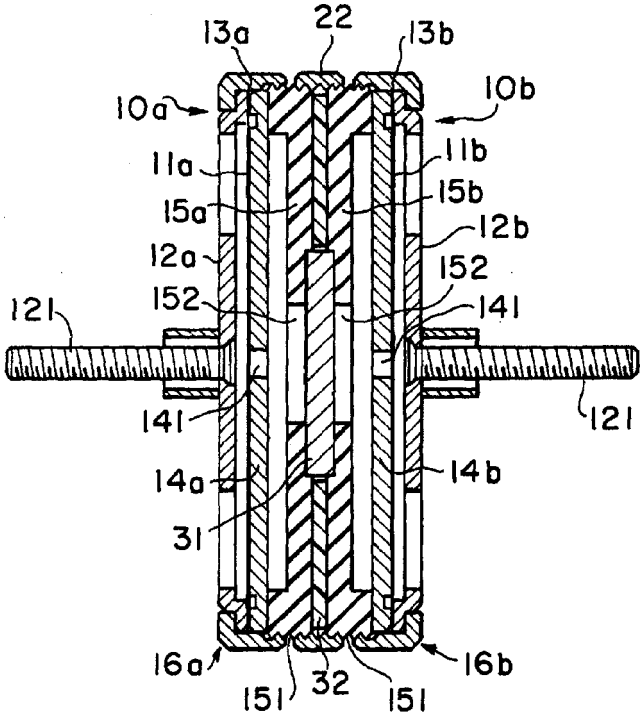
this requirement. This patent describes a number of ways in which bending-wave energy can be efficiently applied near a panel edge. The description is short, interesting, and supplemented by more than two dozen illustrations.—GLA

6,985,597

43.38.Kb VARIABLE DIRECTIONAL CAPACITOR MICROPHONE COMPRISING ELASTIC ACOUSTIC RESISTING MEMBER

Hiroshi Akino, assignor to Kabushiki Kaisha Audio-Technica
10 January 2006 (Class 381/174); filed in Japan 18 December 2003

Today, variable-pattern microphones consist of dual elements whose outputs are electrically combined in different ratios in order to achieve a given pattern. This microphone permits pattern adjustment through the use of an internal acoustical resistance element and the amount of compression



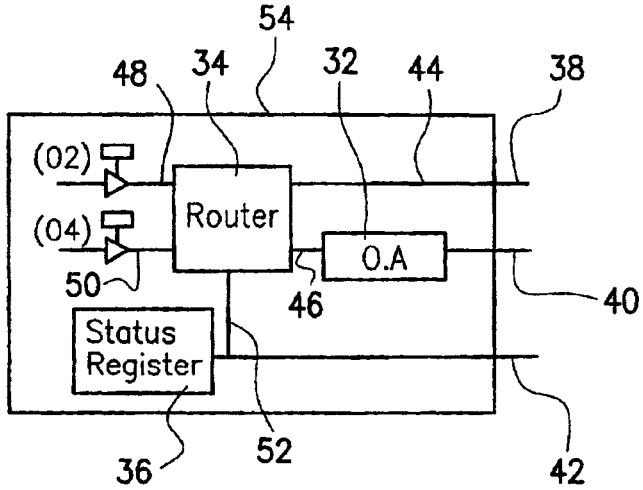
it is subjected to. "The acoustic resisting member **31** is formed with a spongy elastic material which has continuous air bubbles. An acoustic resistance of the spongy material changes with a compressed volume thereof." Based on casual observation, the structure resembles that of the venerable AKG C12 dual element capsule.—JME

6,954,675

43.38.Lc AUDIO CHIP WITH SWITCHABLE AUDIO OUTPUT PATHWAYS

Chia-Chin Chu and Benjamin Ym Pan, assignors to VIA Technologies, Incorporated
11 October 2005 (Class 700/94); filed in Taiwan, Province of China 7 March 2001

The audio chip at the heart of many computer audio cards is supplied to various and sundry audio-card manufacturers. This may mean that the implementation of the controlling program within the audio chip may not



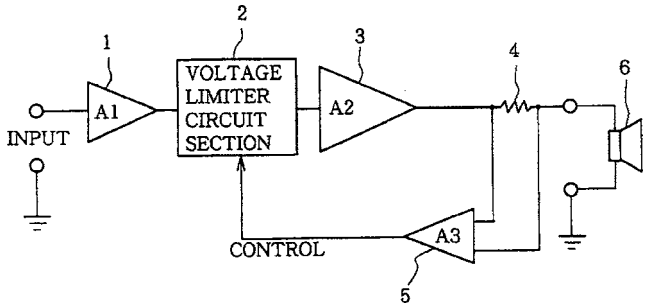
provide the proper functionality, thus requiring modification of the controlling program. As the reader may know, or at least surmise, hardware is easier to fix than software. Thus the present invention.—NAS

6,980,058

43.38.Lc AMPLIFIER CIRCUITRY

Masao Noro, assignor to Yamaha Corporation
27 December 2005 (Class 330/298); filed in Japan 27 December 2002

What is shown is an overload protection circuit for an audio power amplifier. The circuit senses input voltage and output current, and triggers



the voltage limiter section if either of these exceeds a threshold value. This arrangement is said to provide reliable protection over a wide range of loads without introducing unwanted audible artifacts.—GLA

6,975,995

43.38.Md NETWORK BASED MUSIC PLAYING/ SONG ACCOMPANYING SERVICE SYSTEM AND METHOD

Yun-Jong Kim, assignor to Hanseulsoft Company, Limited
13 December 2005 (Class 704/278); filed in the Republic of Korea 20 December 1999

Noraebang (Korean karaoke) to a cell phone is the inspiration for this stew of ideas. First, connect a ''terminal'' (a.k.a. cell phone) to a computer network. Next, since MIDI streams are highly encoded (and bandwidth efficient), download the accompaniment to the cell phone. Provision is also made for transmitting the lyrics to the terminal. The accompanying English explanation is highly accented and the figure text is often misspelled.—MK

6,979,769

43.38.Md DATA REPRODUCING DEVICE, DATA REPRODUCING METHOD, AND INFORMATION TERMINAL

Yoshiyuki Majima *et al.*, assignors to Faith, Incorporated
27 December 2005 (Class 84/645); filed in Japan 8 March 1999

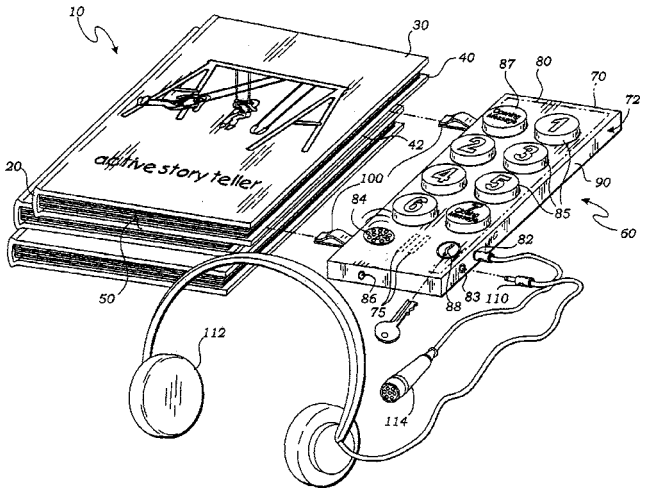
As in United States Patent 6,975,995 (reviewed above), start by taking a bandwidth-restricted cell phone and add karaoke. Again, use MIDI to transmit the note-on and -off commands. During gaps in the MIDI stream, transmit encoded video images. Will karaoke leave the bar and take place on the street? Or worse yet, in the car?—MK

6,985,693

43.38.Md INTERACTIVE PICTURE BOOK WITH VOICE RECORDING FEATURES AND METHOD OF USE

Debra Giampapa Kirwan, Corona Del Mar, California
10 January 2006 (Class 434/317); filed 3 February 2003

Pushbutton electronic books are well known. This marginal invention



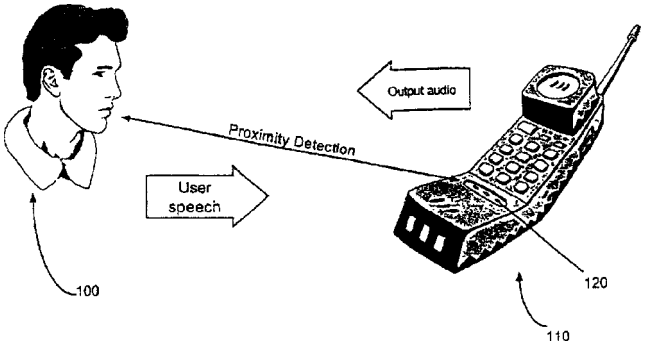
proposes to separate the interface from the actual book (except for a clip).—MK

6,952,672

43.38.Si AUDIO SOURCE POSITION DETECTION AND AUDIO ADJUSTMENT

Bruce A. Smith, assignor to International Business Machines Corporation
4 October 2005 (Class 704/226); filed 25 April 2001

This system for improving operation of a portable device would include an infrared or other distance calibration system. During usage, the



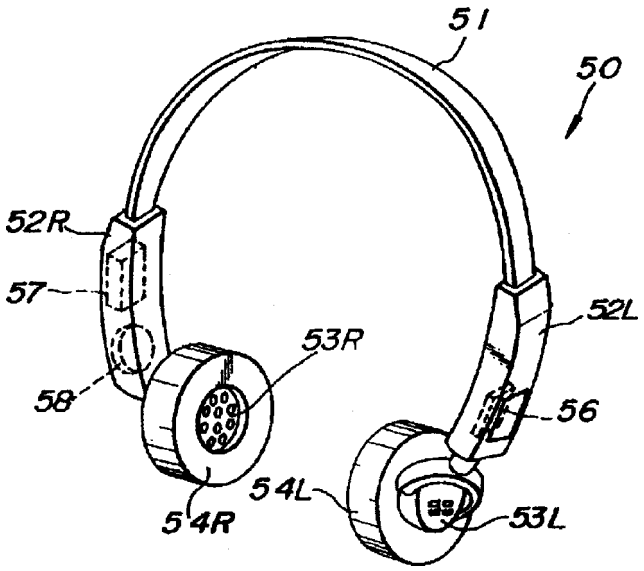
measured speaker-distance information would be used to modify both input and output voice channels so as to maintain a uniform voice quality and suitable audio output.—DLR

6,975,732

43.38.Si AUDIO SIGNAL REPRODUCING APPARATUS

Masayuki Nishiguchi and Yoshihito Fujiwara, assignors to Sony Corporation
13 December 2005 (Class 381/74); filed in Japan 25 October 1989

A small digital memory chip can store many, many compressed audio files. Thus, one's personal library of recorded music can be housed in something the size of a billfold, requiring only a plug-in headset, for many hours of listening. Better yet, the memory chips can be housed in the headset



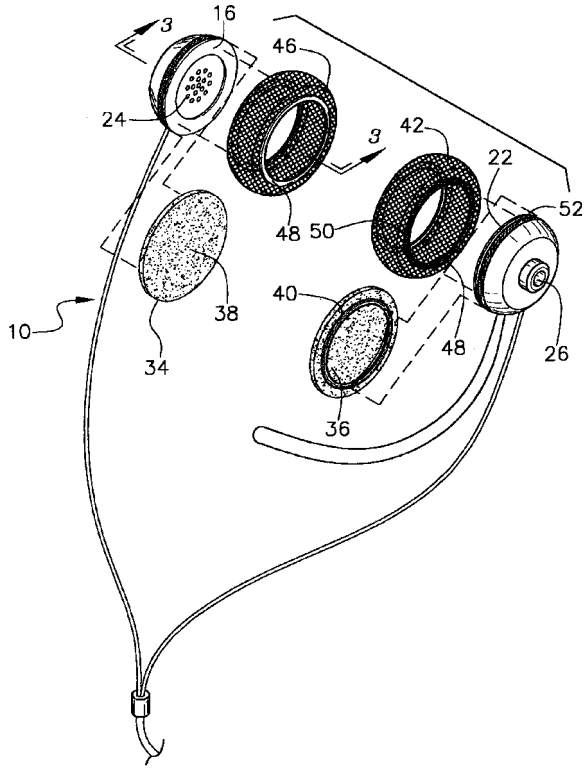
itself. This patent is a division of two previous patents going back to 1996 and devotes most of its space to a "typical example" of the coding and decoding process. This is a parallel bandpass scheme in which allowable noise levels are set progressively higher at higher frequencies.—GLA

6,980,666

43.38.Si HEADPHONE ASSEMBLY

Billie H. Owen, China Grove, North Carolina
27 December 2005 (Class 381/371); filed 3 September 2003

Resilient muffs **42**, **46** fit over the user's ears and support headphones **22**, **24** so that no additional headband is required. For whatever reason, the



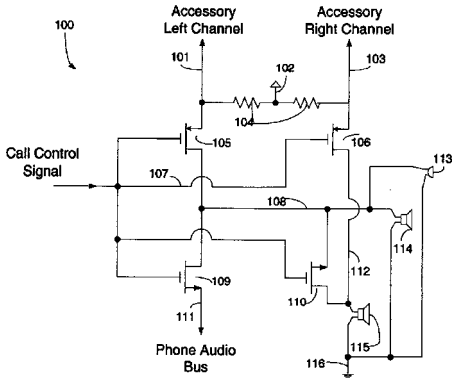
patent does not appear to cover a single unit; all of the claims describe "a pair of headphones, said headphones being operationally coupled together."—GLA

6,980,827

43.38.Si AUDIO ACCESSORY SWITCHING CIRCUIT

Tan Thok Lin *et al.*, assignors to Motorola, Incorporated
27 December 2005 (Class 455/556.1); filed 10 May 2001

This patent relates to multipurpose, portable, electronic devices. The diagram illustrates yet another circuit intended to automatically switch from



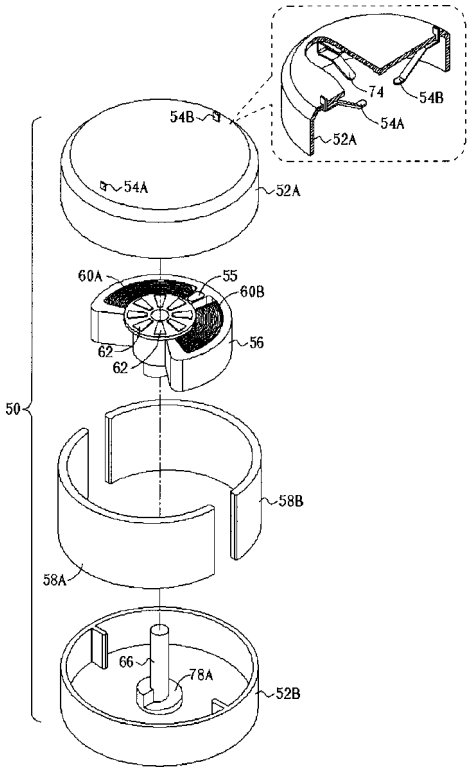
stereo music playback to mono telephone operation whenever a call is transmitted or received. This is accomplished by biasing complementary transistors on and off in response to a single call-control signal.—GLA

6,984,903

43.38.Si VIBRATOR AND RELATED APPARATUS FOR PROVIDING VIBRATION IN DIFFERENT DIRECTIONS

Yu-Chuan Chang and Jen-Cheng Lai, assignors to BenQ Corporation
10 January 2006 (Class 310/36); filed in Taiwan, Province of China 22 October 2003

Suppose your cellular phone vibrates but you don't feel it. This patent



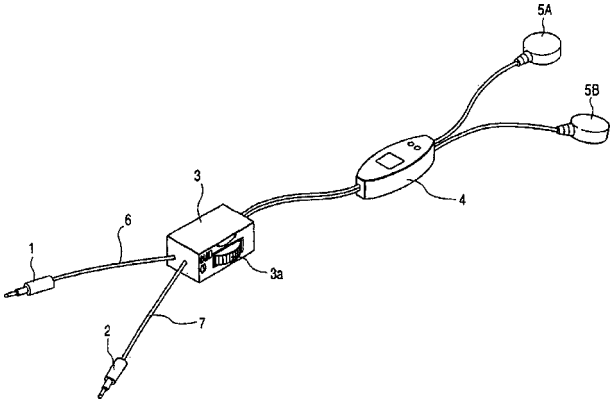
warns us that typical vibrators generate force along a single axis. An improved design is disclosed that combines rotational and pistonlike action.—GLA

6,985,592

43.38.Si MULTIPURPOSE EARPHONE SET

Satoshi Ban and Kiyoshi Hotta, assignors to Matsushita Electric Industrial Company, Limited
10 January 2006 (Class 381/74); filed in Japan 8 January 1997

Why should it be necessary to use separate, dedicated headphones for your cellular phone and portable audio player? Matsushita has corrected this



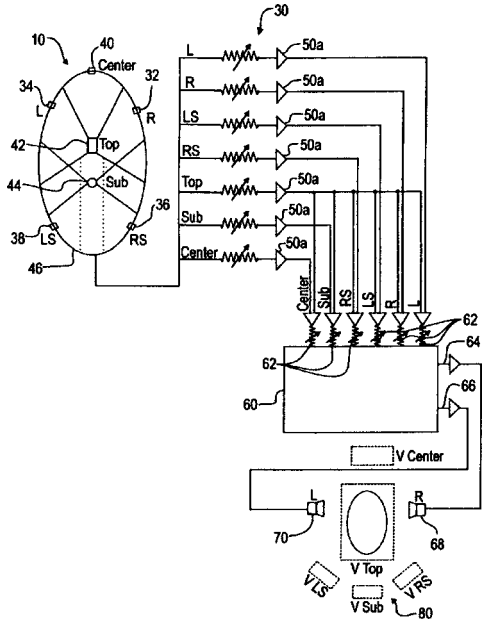
shortcoming by providing a single headset with two cords, two plugs, and a little control box.—GLA

6,980,661

43.38.Vk METHOD OF AND APPARATUS FOR PRODUCING APPARENT MULTIDIMENSIONAL SOUND

Mike Godfrey, Palm Beach, Florida and Antonio Nucaro, Markham, Ontario, Canada
27 December 2005 (Class 381/26); filed 14 November 2001

The patent describes the use of a surround microphone array for feeding a ''virtual surround'' processor. The output of the processor is a two-channel signal embodying combinations of level, equalization, and delay relationships to generate a set of ''transaural'' cues at the ears of an on-axis



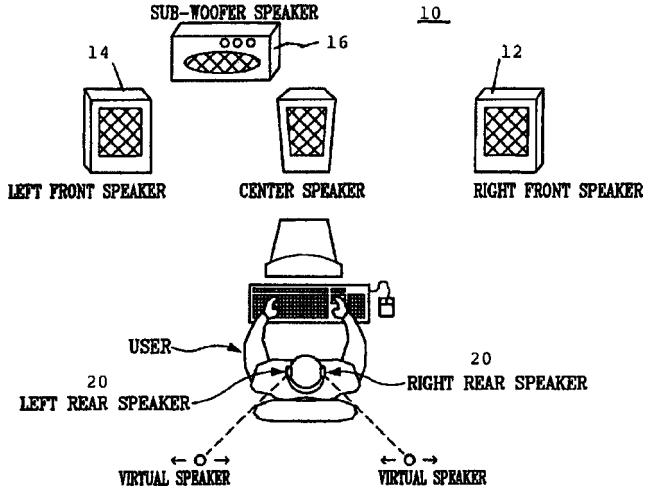
listener. Many modern TV sets and surround processors provide functions of this sort, and they are fairly successful in conveying the general ambience of, for example, sports locales. For most music applications these systems are apt to fall short of ideal response.—JME

6,983,054

43.38.Vk MEANS FOR COMPENSATING REAR SOUND EFFECT

Eric Cheng, assignor to C-Media Electronics Incorporated
3 January 2006 (Class 381/309); filed 10 December 2001

Some commercial television receivers and computer sound systems include three loudspeakers to correctly reproduce front-left, -center, and -right audio channels. More often, left and right channels are combined to



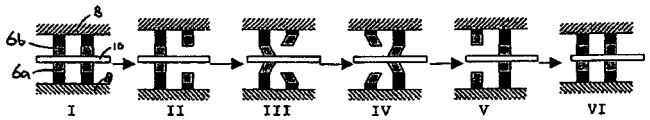
form a phantom center image. Creating phantom sources for rear-left and rear-right channels is much more complicated and involves a number of compromises. This patent suggests that nonsealing headphones can do a better job of reproducing the two surround channels.—GLA

6,977,461

43.40.At SYSTEM AND METHOD FOR MOVING AN OBJECT EMPLOYING PIEZO ACTUATORS

Stefan Gertrud Marie Hendriks *et al.*, assignors to ASML Netherlands B.V.
20 December 2005 (Class 310/328); filed 15 December 2003

This patent discloses a piezoelectric ''grab and translate'' mechanism that can be used for photolithography or (one assumes) anywhere that small, smooth, translation motion is desired. The basic operation is shown in the figure, wherein a plate **10** is translated laterally by suitably sequenced pi-



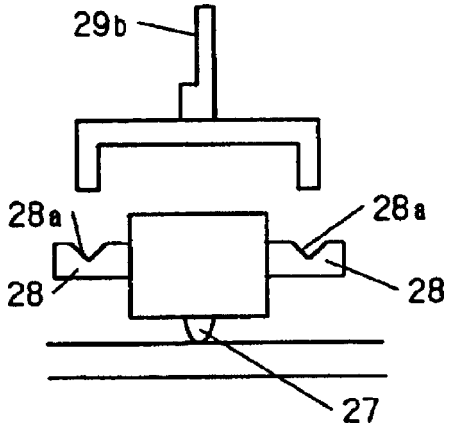
ezoelectric actuators **6a**, **6b**. That being said, there is nothing here that hasn't been done before, although the reduction of hysteresis to nanometer-scale motions is unusual. No discussion of how to design such a thing or deal with the hysteresis and slip problems is given, and the reader is advised to look elsewhere for details.—JAH

6,979,935

43.40.At PIEZOELECTRIC MOTOR AND ELECTRONIC EQUIPMENT WITH PIEZOELECTRIC MOTOR

Akihiro Iino and Haruhiko Hasegawa, assignors to Seiko Instruments Incorporated
27 December 2005 (Class 310/323.09); filed in Japan 12 July 2002

This patent describes various incarnations of a piezoelectric translation device that could be used as a motor. There are some 15 ''modes'' of use, as the inventors call them, depicted, some with very detailed mechanical engagement contacts, one example of which is given in the figure. There is no supporting data to show that any of this works, so perhaps one ought to view it as a portfolio of piezoelectric curiosities.—JAH

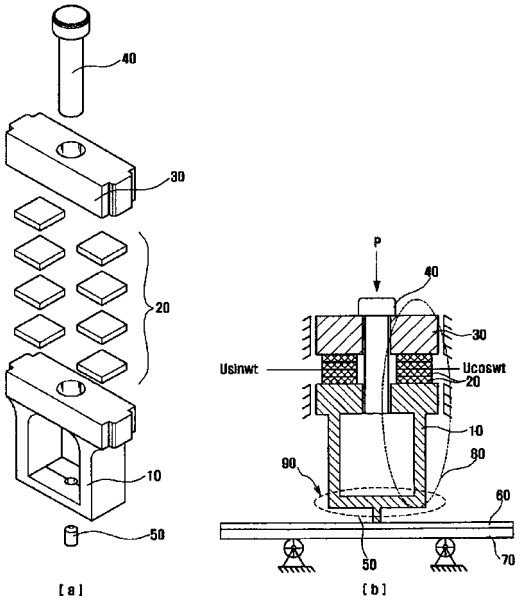


6,984,920

43.40.At LINEAR PIEZOELECTRIC ULTRASONIC MOTOR

Seok Jin Yoon *et al.*, assignors to Korea Institute of Science and Technology
10 January 2006 (Class 310/323.12); filed in the Republic of Korea 22 March 2002

This patent discloses a linear, piezoelectric, friction-drive motor. The left and right halves are driven 90 deg out of phase so that the tip **50** executes an elliptical scrubbing motion. The motion can be used to drive a plate in contact with the tip in either direction, left or right, depending on the phase of the two drive signals. The discussion of the design is brief and lucid.—JAH

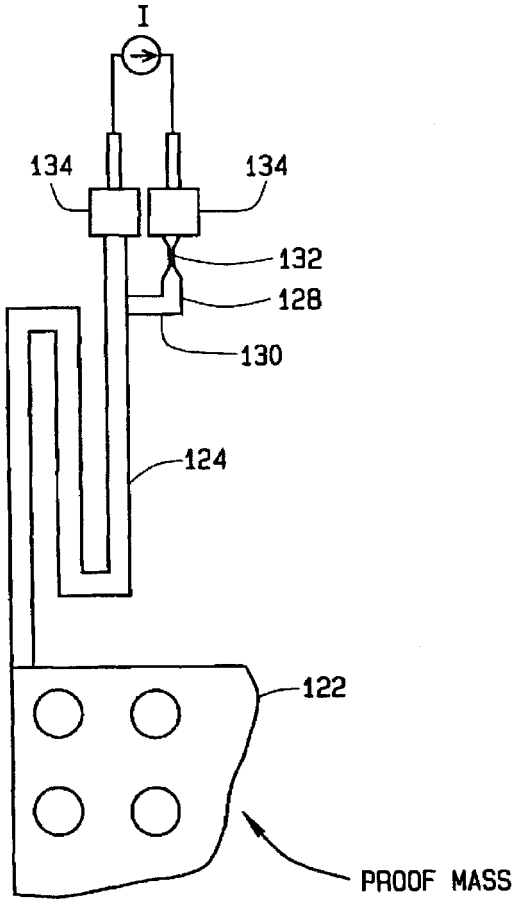


6,978,673

43.40.Cw METHODS AND SYSTEMS FOR SIMULTANEOUSLY FABRICATING MULTI-FREQUENCY MEMS DEVICES

Burgess R. Johnson *et al.*, assignors to Honeywell International, Incorporated
27 December 2005 (Class 73/504.12); filed 7 February 2003

This patent discloses the use of fusible links as frequency tuning elements in MEMS gyros. The inventors present a description of how such gyros are made and they state that three-axis gyros of this type require well-separated lateral resonant frequencies of the suspension **124**. Their solution is to use fusible links **132** to change the suspension resonance in postfabrication tuning. It is not clear, though, why one cannot design the resonances correctly in the first place.—JAH

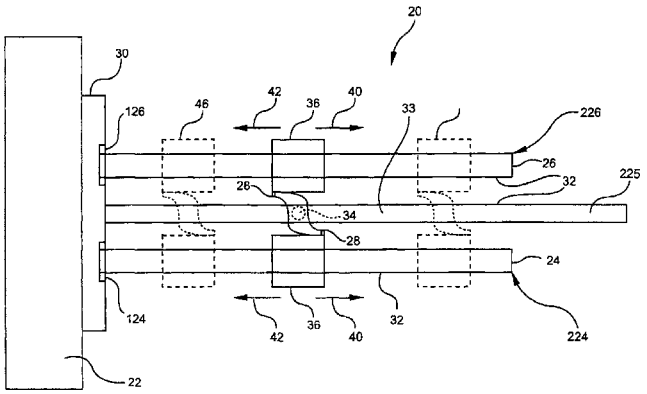


6,983,833

43.40.Tm SELF-TUNING VIBRATION ABSORBER SYSTEM AND METHOD OF ABSORBING VARYING FREQUENCY VEHICLE VIBRATIONS

Douglas E. Ivers and Robert W. Wilson II, assignors to LORD Corporation
10 January 2006 (Class 188/379); filed 16 May 2003

This absorber system, intended to suppress helicopter-rotor vibrations, consists of two spring-mass systems with different amounts of damping, interconnected so as to adjust the natural frequencies of these systems on the



basis of the relative motions of their masses. The more highly damped system consists of mass **36**, relatively free to slide along cantilever beam **24**, and is connected to the base **30** via a high-damping element **124**. The less highly damped system consists of mass **36**, relatively free to slide along a similar cantilever **26**, and is connected to the base via a relatively lightly damped element **126**. The two masses are interconnected by friction or resilient elements at **28** via a pivot arrangement that can slide along a track **225**. If the attached body **22** vibrates at a frequency lower than the natural frequency of the beam-supported masses in their initial configuration, then the relative phase of the motions of the masses results in friction, causing the masses to ''walk'' in the direction **42** toward the support, thus reducing their natural frequency. The converse happens if the body vibrates at a frequency higher than the natural frequency of the masses. Thus, eventually the masses settle at the frequency at which the support vibrates and thus can serve as a vibration absorber at that frequency.—EEU

6,974,000

43.40.Vn SYSTEM AND METHOD FOR LIMITING VIBRATION IN AN APPARATUS DURING A LOSS OF POWER

J. David Carlson *et al.*, assignors to Lord Corporation
13 December 2005 (Class 188/267.2); filed 20 February 2002

A magnetorheological (MR) fluid undergoes an effective viscosity increase when subjected to a magnetic field. Linear dampers that contain a MR fluid are included in a system whose damping is to be controlled. For example, one wants high damping when a washing machine rotor spins at a speed that matches the resonant frequency of the rotor, as supported on its resilient suspension system, but one wants low damping to obtain good vibration isolation when the rotor spins at higher speeds. A controller that actuates the magnetic field of a MR damper can be used to adjust this damping, depending on the rotor speed. A controller can also call for high damping as a machine spins down through resonance when a loss of power occurs. For this application the controller needs to be provided with a source of power that is independent of the primary source—for example, a source using a dynamo actuated by the spinning rotor.—EEU

6,976,403

43.40.Vn STEERING WHEEL DYNAMIC ABSORBER ASSEMBLY

Kenneth L. Oblizajek and Carl T. Wright, assignors to General Motors Corporation
20 December 2005 (Class 74/552); filed 13 September 2002

A ring-shaped mass is located in an annular cavity in the rim of a steering wheel and is supported on an array of flexures. The flexures are designed to permit motion of the mass in the circumferential and radial directions and are tuned to suppress vibrations of the steering wheel at predetermined frequencies.—EEU

6,981,420

43.40.Yq OMNI-DIRECTIONAL MOVEMENT SENSOR

Arlen J. Lowrance, Tulsa, Oklahoma
3 January 2006 (Class 73/652); filed 1 December 2003

As stated in the abstract: ''A microphone is enclosed in a semi-sound-insulated housing with a jingle-bell-type device that generates sound in response to a very slight movement and/or vibration of the housing and/or structure to which the housing is attached.'' While the device can be used for any application where motion is to be sensed, the intended application for this sensor is games, where a player strikes various surfaces, such as a ball hitting a target.—JME

6,983,820

43.50.Gf NOISE AND VIBRATION SUPPRESSORS

David Boast and Donald Milne Turner, assignors to Avon Polymer Products Limited
10 January 2006 (Class 181/249); filed in the United Kingdom 7 September 2001

This patent in essence deals with mufflers of the type used in motor vehicles. The mufflers described here are constructed by connecting modular suppression chambers in series within a duct. The various modules may communicate with each other via apertures in an internal duct and may incorporate Helmholtz resonator and volume resonance elements. Spacers between adjacent elements may be selected to achieve tuning: Resiliently supported elements may act as ''tuned dampers'' to suppress vibrations.—EEU

6,985,593

43.50.Ki BAFFLE VIBRATION REDUCING

George Nichols *et al.*, assignors to Bose Corporation
10 January 2006 (Class 381/89); filed 23 August 2002

By using two loudspeakers out of phase with each other, less vibration energy is transferred into the supporting baffle and, hence, there is less buzzing or rattling of the lightweight baffle. In particular, the patent makes reference to loudspeakers in cars where the baffle might be a lightweight door panel or rear package shelf.—CJR

6,983,822

43.55.Ev SOUND ABSORPTIVE PROTECTIVE MAT

Desmond O'Regan *et al.*, assignors to Rieter Technologies AG
10 January 2006 (Class 181/290); filed 21 May 2002

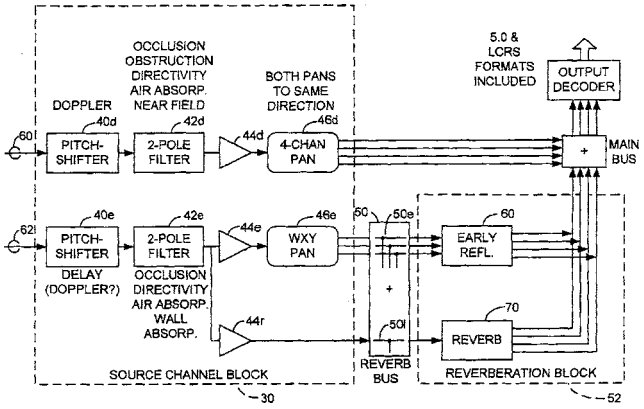
This protective mat for an automobile is also sound absorptive, using a multilayered assembly package with a low air-flow resistant and water-impermeable layer.—CJR

6,917,686

43.55.Lb ENVIRONMENTAL REVERBERATION PROCESSOR

Jean-Marc M. Jot *et al.*, assignors to Creative Technology, Limited
12 July 2005 (Class 381/63); filed 12 February 2001

Jot and colleagues expand Jot's earlier work to include a per-channel delay block for Doppler shifting of moving sources, a two-pole filter to model absorption, and, finally, pan blocks for additional directionality. Note



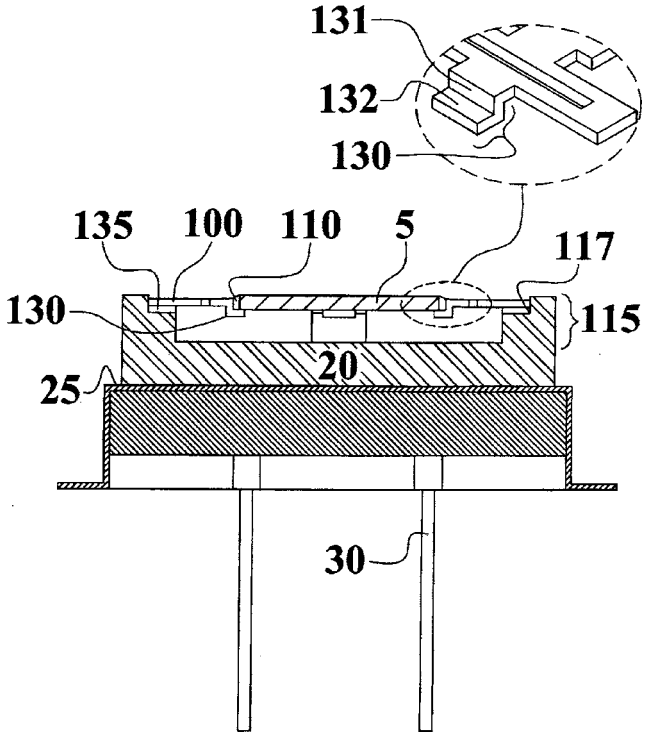
how the reverb is not panned since this aspect adds envelopment, not direction. The last block adds reverb, as shown, implemented with a mix of hardware and software. The patent is clear and includes details and equations.—MK

6,984,925

43.58.Hp LOW ACCELERATION SENSITIVITY MOUNTING STRUCTURES FOR CRYSTAL RESONATORS

Peter E. Morley *et al.*, assignors to Delaware Capital Formation, INC
10 January 2006 (Class 310/348); filed 28 May 2003

This patent will be of interest to those involved with high-precision frequency standards. The inventors describe a means of isolating low-frequency quartz resonators from acceleration and packaging stresses. The basic concept is to support the resonator in the plane of motion using a set



of strain-relieving loop springs. In one implementation of these ideas, quartz element **5** is supported by springs **130** having a shelf on which to rest the crystal, while an adhesive bond is made between the crystal and the electrode. The discussion is informative and measurements on certain resonators and the resulting shock sensitivities are given.—JAH

6,974,902

43.58.Jq MUSIC SYNTHESIZER AND A METHOD OF GENERATING A SYNTHESIZER OUTPUT WITH A CONSTANT BEAT

Chris Adam, assignor to Apple Computer, Incorporated
13 December 2005 (Class 84/671); filed 30 April 2003

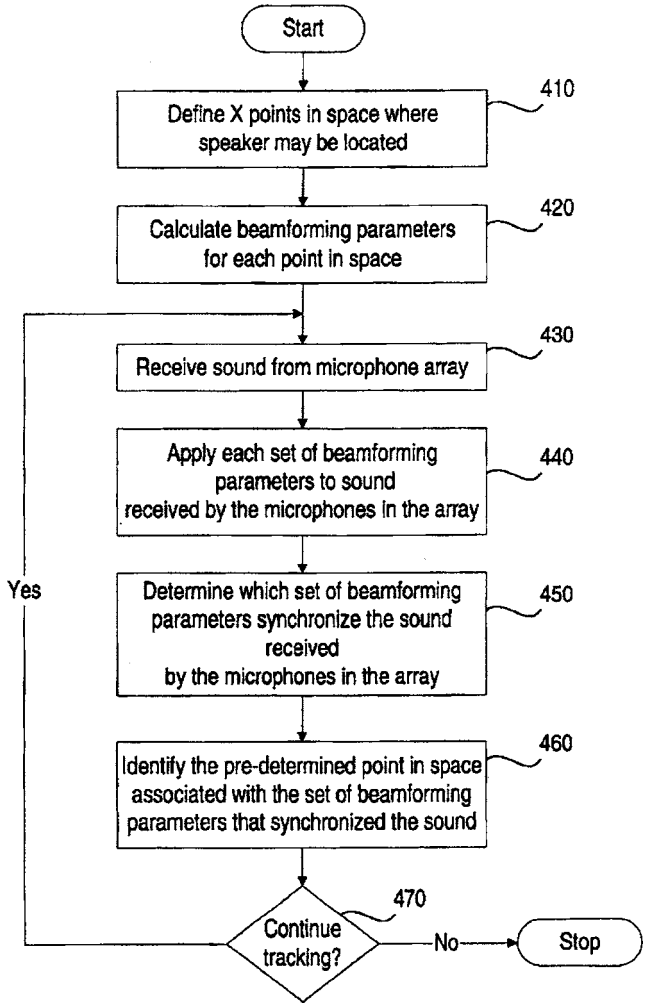
Every small idea deserves a patent. This one-page idea notes that if you mix two detuned oscillators together, the resulting tone will have beats that vary with the pitch and double with each octave. The solution? Add a small bias to the second tone. Game over.—MK

6,980,485

43.60.Fg AUTOMATIC CAMERA TRACKING USING BEAMFORMING

John McCaskill, assignor to Polycom, Incorporated
27 December 2005 (Class 367/119); filed 25 October 2001

This is yet another variation in a large set of recent patents using adaptive beamforming for automatic camera tracking of talkers in remote



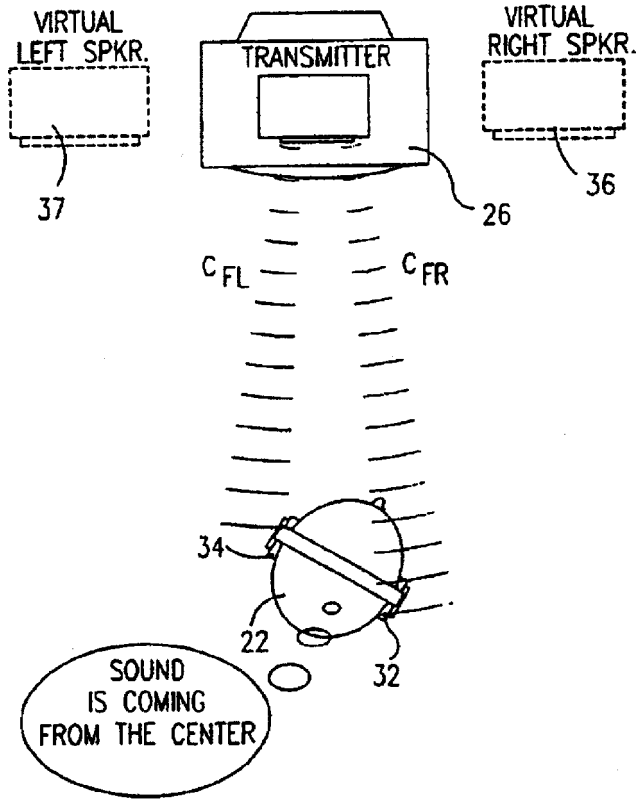
conferencing applications. The iterative signal-flow diagram shows the operation of the system.—JME

6,975,731

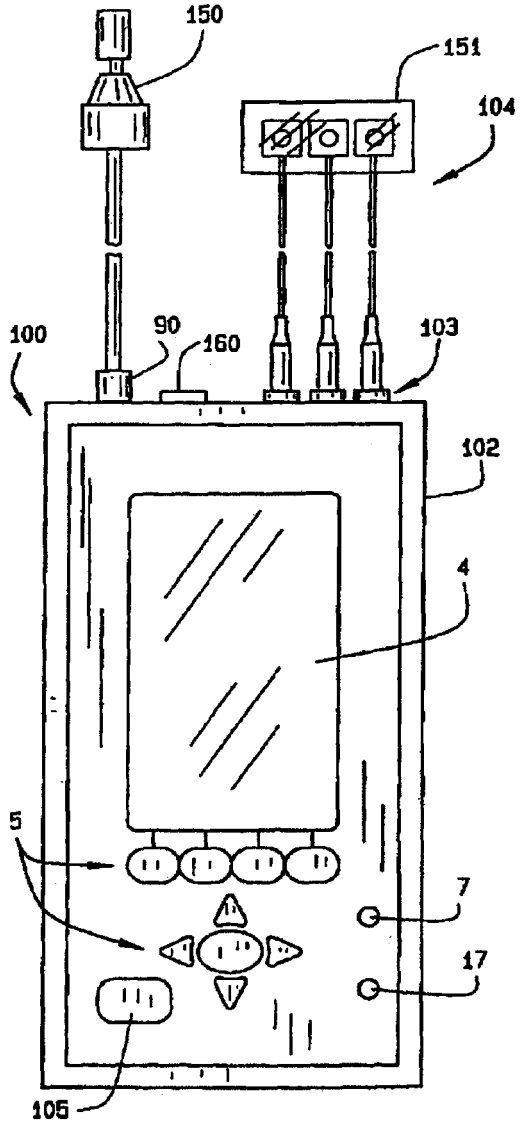
43.66.Pn SYSTEM FOR PRODUCING AN ARTIFICIAL SOUND ENVIRONMENT

Yuval Cohen *et al.*, assignors to Beh Limited
13 December 2005 (Class 381/74); filed in Israel 24 June 1997

The patent describes a stereo headphone arrangement for reproducing five-channel surround sound in a typical home environment. The multichannel input is fed to an ultrasonic modulation transmitter positioned on the center plane atop the video receiver. Microphones at each ear pick up the modulated signal, which is resolved into the five audio signals, which are



then processed to appear at their original positions. The system further provides an azimuthal head tracking function based on arrival time differences at the ears. The listener can thus dispense with loudspeakers altogether—as well as enjoy the immersive realism of images that do not shift with normal head motions.—JME



6,974,421

43.66.Sr HANDHELD AUDIOMETRIC DEVICE AND METHOD OF TESTING HEARING

Elvir Causevic and Eldar Causevic, assignors to Everest Biomedical Instruments Company
13 December 2005 (Class 600/561); filed 28 April 2000

The object of this portable device is to provide a hand-held means for auditory screening which provides otoacoustic emission (OAE), automated screening auditory brainstem response (ABR), tympanometry, otoreflectance, and OAE stimulator operations. Intended for neonatal testing, the integration of OAE and ABR screening devices into a single hand-held instrument enables a user to detect less common sensorineural hearing loss associated with outer hair cell abnormalities and to detect less common sensory hearing loss associated with inner hair cell abnormalities. In the preferred embodiment, the apparatus includes a portable hand-held enclosure containing a digital signal processor. The processor incorporates a computer program capable of conducting both OAE and ABR test procedures. A display device is mounted on the enclosure to display patient information, testing procedures, and graphs of test results. Provisions are included for connecting a probe to the signal processor and accommodating an onboard power supply.—DRR

6,910,013

43.66.Ts METHOD FOR IDENTIFYING A MOMENTARY ACOUSTIC SCENE, APPLICATION OF SAID METHOD, AND A HEARING DEVICE

Sylvia Allegro and Michael Büchler, assignors to Phonak AG
21 June 2005 (Class 704/256); filed 5 January 2001

Hearing aid performance can be improved if the background environment can be recognized and appropriate filtering performed. Rather than having the wearer control the filters, the inventors propose using statistical identification (*á la* hidden Markov models). Details beyond that are conspiculously absent.—MK

6,974,410

43.66.Ts AUDITORY SIGNAL GENERATOR FOR PEOPLE SUFFERING FROM TINNITUS

Christophe Micheyl and Arnaud Norena, assignors to Centre National de la Recherche Scientifique
13 December 2005 (Class 600/25); filed 18 January 2001

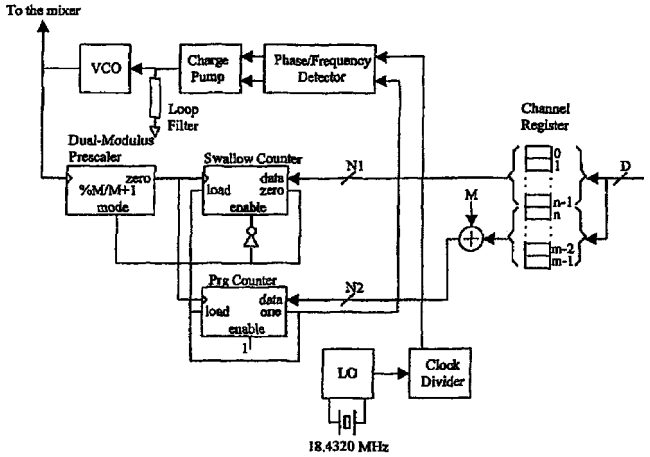
The patent relates to a process for programming a device and the device itself for generating auditory signals for people suffering from tinnitus. The characteristics of the signals generated are adapted specifically to the audiological characteristics of an individual in order to promote the disappearance or alleviation of tinnitus in the patient. To this purpose, the absolute auditory thresholds are measured and the characteristics are calculated (based on the measurements obtained). The spectral envelope of the auditory signal is then adapted to the patient's particular case, in order to program a generator to deliver the auditory signal.—DRR

6,975,739

43.66.Ts HEARING AID WITH A RADIO FREQUENCY RECEIVER

Gudmundur Bogason and Bjarne Klemmensen, assignors to Oticon A/S
13 December 2005 (Class 381/315); filed in Denmark 4 October 2000

A radio receiver is tunable to many receiving frequencies without physically changing crystals. The receiver may be integrated with the hearing aid or added externally to an existing hearing aid. The synthesized frequency is selected, typically with a switch on the hearing aid, via autosearching preprogrammed channels.—DAP

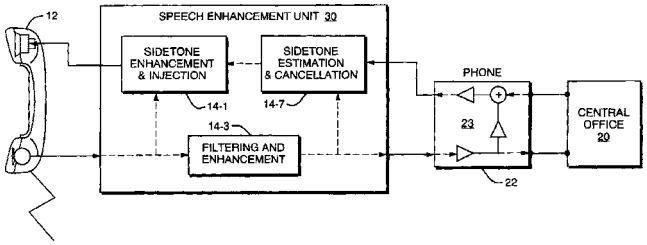


6,975,984

43.66.Ts ELECTROLARYNGEAL SPEECH ENHANCEMENT FOR TELEPHONY

Joel M. MacAuslan *et al.*, assignors to Speech Technology and Applied Research Corporation
13 December 2005 (Class 704/208); filed 7 February 2001

This is an electrolaryngeal (EL) device that provides a means of verbal communication for people who have either undergone a laryngectomy or are otherwise unable to use their larynx (e.g., after a tracheotomy). Current devices cannot be released and restarted quickly to produce the voicing silences in a spoken phrase and, thus, the perceived overall quality of speech is degraded by the presence of the device "buzzing" throughout each phrase. To overcome this defect, a technique was developed for separating an acoustic signal into a voiced component corresponding to an electrolaryngeal source and an unvoiced component corresponding to a turbulence source. This technique is used to improve the quality of EL speech, and it may be adapted for use in a special purpose telephone.—DRR

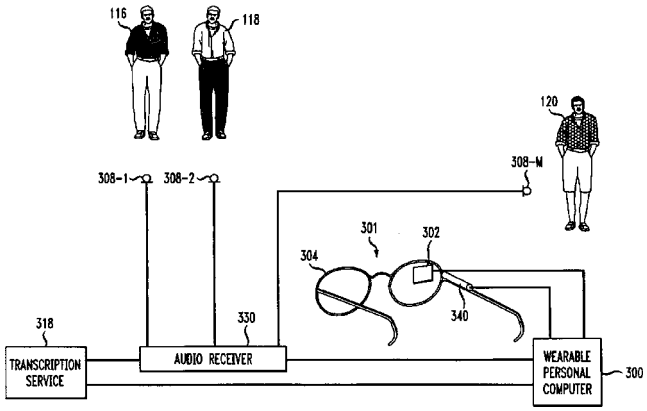


6,975,991

43.66.Ts WEARABLE DISPLAY SYSTEM WITH INDICATORS OF SPEAKERS

Sara H. Basson and Dimitri Kanevsky, assignors to International Business Machines Corporation
13 December 2005 (Class 704/271); filed 31 January 2001

Methods and systems are described for creating a user-friendly environment that permits a hearing-impaired user to identify the speaker and what is being said during interactions with other individuals in a face-to-face meeting. The device provides a visual indicator system with a wearable display (e.g., in the format of eyeglasses) to be worn by an individual with a hearing impediment. The wearable display system produces an indication of who is currently speaking in a conversation or meeting, allowing the user to be able to turn his/her attention to the actual speaker, so that he/she can do some lip reading. In one embodiment, the indication of the speaker involves projecting a marker (e.g., an illuminated circle or dot) onto the display over the head of the individual that is currently speaking.—DRR

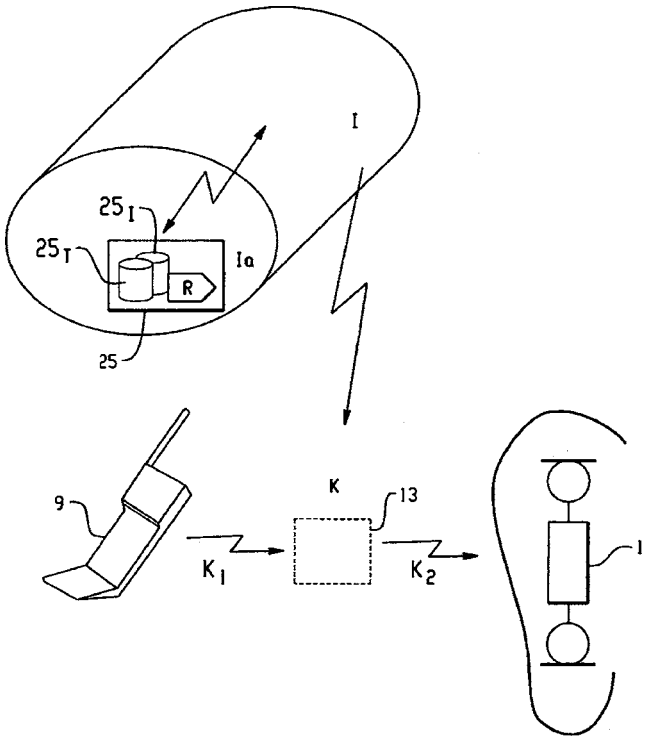


6,978,155

43.66.Ts FITTING-SETUP FOR HEARING DEVICE

Christian Berg, assignor to Phonak AG
20 December 2005 (Class 455/557); filed in the World IPO 18 February 2000

A cell phone is used to communicate either directly or via a converter to fit data and signal processing algorithms wirelessly to a hearing aid. The converter may be connected via wired or wireless connection to the cell



phone. The hearing aid may be connected to a server via the cell phone and programmed either by a hearing-aid-dispensing professional or by the hearing aid wearer.—DAP

6,980,864

43.66.Ts HIGH CONTACT COUNT, SUB-MINIATURE, FULL IMPLANTABLE COCHLEAR PROSTHESIS

Michael A. Faltys *et al.*, assignors to Advanced Bionics Corporation
27 December 2005 (Class 607/116); filed 21 July 2003

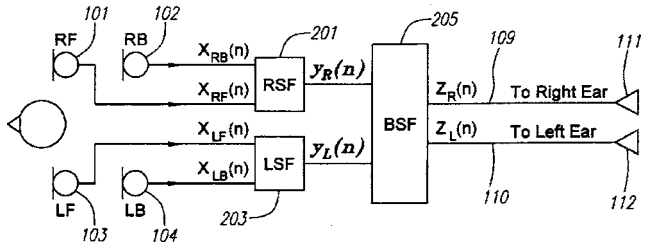
An electrode array, connected to a hermetically sealed, implanted case, is programmable via wireless means for the number of active contacts that supply stimulation to surrounding tissue. Low-voltage, active switching elements within the array provide means for lateral, medial, and radial stimulation in groups.—DAP

6,983,055

43.66.Ts METHOD AND APPARATUS FOR AN ADAPTIVE BINAURAL BEAMFORMING SYSTEM

Fa-Long Luo, assignor to GN Resound North America Corporation
3 January 2006 (Class 381/313); filed 5 December 2001

Each hearing aid of a binaural pair has one pair of microphones whose outputs are processed in a first stage as an end-fire configuration. The outputs from each ear are then combined in a second processing stage as a broadside configuration, resulting in a combination of spatial filtering and binaural cues. Either fixed or adaptive polar patterns may be employed at the two processing stages.—DAP

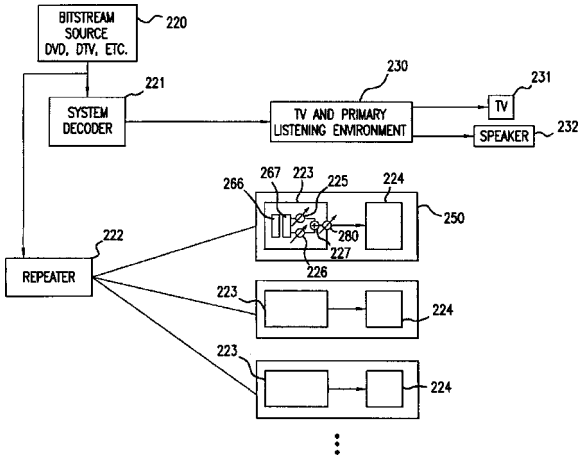


6,985,594

43.66.Ts VOICE-TO-REMAINING AUDIO (VRA) INTERACTIVE HEARING AID AND AUXILIARY EQUIPMENT

Michael A. Vaudrey and William R. Saunders, assignors to Hearing Enhancement Company, LLC.
10 January 2006 (Class 381/96); filed 14 June 2000

A set-top, bit-stream decoder separates a received audio signal into a voice signal and a background signal and independently amplifies each appropriately for one or more hearing-impaired listeners. Transmission of the decoded and mixed analog audio signals to a television set or to individual listening devices may be via wireless means.—DAP

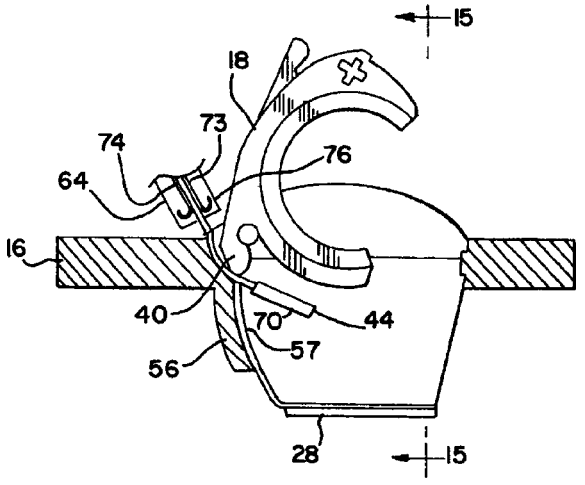


6,985,598

43.66.Ts PROGRAMMING SYSTEM FOR PROGRAMMING HEARING AIDS

Thomas Konrad Joschika, assignor to Knowles Electronics, Incorporated
10 January 2006 (Class 381/323); filed 29 July 1999

A multicontact programming strip is inserted into a slot in a custom hearing aid faceplate while the battery drawer is open. When the battery drawer is closed, the contacts on the programming strip are pushed against mating terminals in the hearing aid.—DAP



6,985,599

43.66.Ts VIBRATOR FOR BONE CONDUCTED HEARING AIDS

Kristian Åsnes, assignor to P&B Research AB
10 January 2006 (Class 381/326); filed in Sweden 2 June 2000

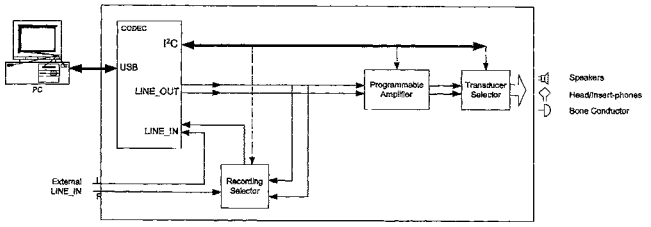
A small, low-power-consumption vibrator contains a two-piece coil and a permanent magnet. The magnet is located between the coil pieces that generate dynamic and static magnetic fields, respectively. The design is based on the assumption that the dynamic magnetic field does not need to pass through the permanent magnet. The static and dynamic magnetic fields superimpose in the air gaps between the coil and magnet to produce an axial force.—DAP

6,980,662

43.66.Yw DEVICE FOR PRESENTING ACOUSTICAL AND VIBRATORY STIMULI AND METHOD OF CALIBRATION

Shawn X. Gao and Daniel J. Freed, assignors to House Ear Institute
27 December 2005 (Class 381/59); filed 5 October 2001

A computer-controlled device automatically self-calibrates and presents stimuli automatically for hearing testing. Frequency-dependent amplitude and phase calibration compensates for impedance variations of air- and bone-conduction transducers to finely control the output levels produced.—DAP



6,947,893

43.72.Ar ACOUSTIC SIGNAL TRANSMISSION WITH INSERTION SIGNAL FOR MACHINE CONTROL

Satoshi Iwaki *et al.*, assignors to Nippon Telegraph & Telephone Corporation
20 September 2005 (Class 704/258); filed in Japan 19 November 1999

The goal of the patented signal transmission system is to inaudibly immerse various coded signals into a voice band. DTMF, for example, is thus ruled out. Several schemes are described. One method involves band-stop filtering of two narrow bands, selectively removing one band from the voice signal to send a zero, removing the other to send a one. The justification is that ''humans are virtually unable distinguish an audible signal uniquely lacking a specific frequency only from the original sound.'' And even when the difference is detectable, ''humans are not able to meaningfully recognize its significance.'' Another scheme would send pulses in narrow bands just outside the voice band, but still within the transmission band (as if that were possible). The justification is similar. The schemes get weirder, supposing that ASCII codes could be inaudibly transmitted without saying how the bits are to be coded. Some schemes involve manipulation of speech parameters, such as pitch, in supposedly undetectable ways. A total of 14 schemes are described, with no frequency details and little timing detail provided.—DLR

6,983,245

43.72.Dv WEIGHTED SPECTRAL DISTANCE CALCULATOR

Alberto Jimenez Felstrom and Jim Rasmusson, assignors to Telefonaktiebolaget LM Ericsson (publ)
3 January 2006 (Class 704/238); filed in Sweden 7 June 1999

A simple method for noise cancellation in, for example, voice-enabled mobile phones is envisioned, though the presentation is hard to follow. The idea seems to be that it is useful to assume that the noise that needs to be compensated for will be either constant or predetermined (such as a known ring tone). This assumption permits a fairly simple approach to noise cancellation by spectral subtraction, without need of more typically complicated adaptive noise-cancellation methods.—SAF

6,985,858

43.72.Dv METHOD AND APPARATUS FOR REMOVING NOISE FROM FEATURE VECTORS

Brendan J. Frey *et al.*, assignors to Microsoft Corporation
10 January 2006 (Class 704/233); filed 20 March 2001

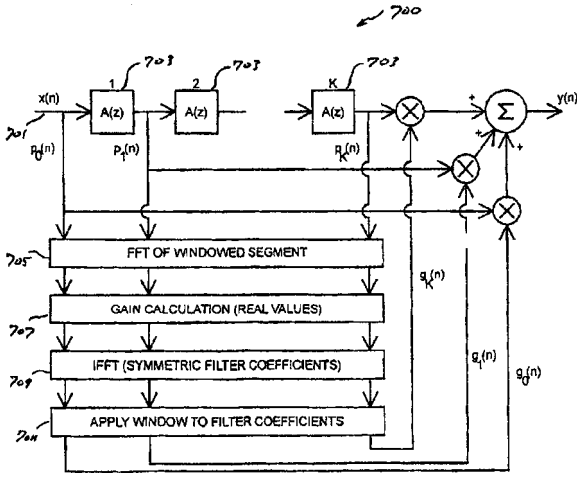
This patent describes in great detail numerous specific embodiments of a general idea for removal of noise and channel distortion by operating on sequences of feature vectors' output by a speech recognition front end. The described idea is intended to generalize across any conceivable kind of feature vectors that would be employed, e.g., in a Gaussian mixture model approach (though other approaches are also covered by extension). A model noisy-signal feature vector is first developed, by any of a number of ways mentioned. This calculated vector is compared with the observed vector to obtain a variance for the noisy-signal vector, which is in turn used in conjunction with the distribution of training feature vectors to extract a clean-signal feature vector.—SAF

6,980,665

43.72.Ew SPECTRAL ENHANCEMENT USING DIGITAL FREQUENCY WARPING

James M. Kates, assignor to GN Resound A/S
27 December 2005 (Class 381/312); filed 1 March 2002

All-pass sections are used in place of FIR sections in a digital filter to produce spectral enhancement via frequency warping. The result, for example, is that for frequencies below and above 2.44 kHz, respectively, the warped system has greater and shorter delays than a conventional FIR filter having the same number of taps.—DAP

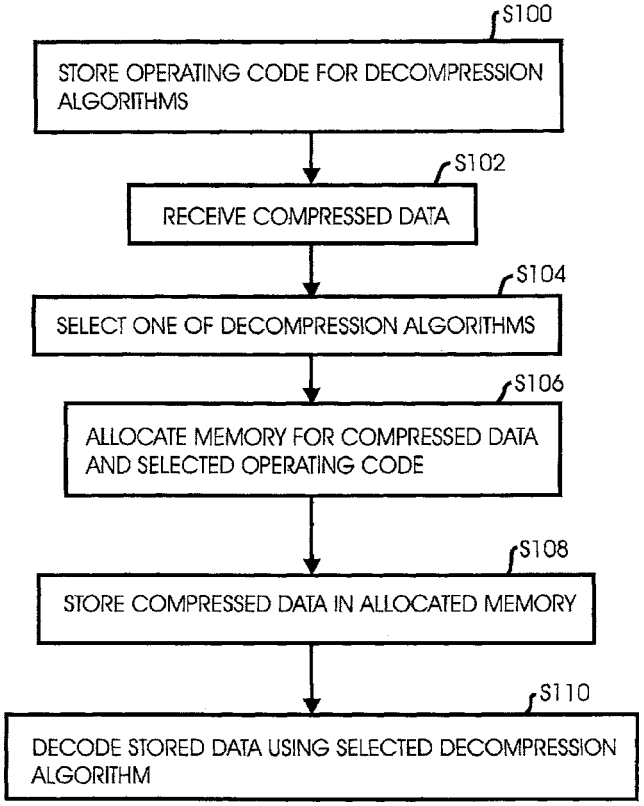


6,985,853

43.72.Gy COMPRESSED AUDIO STREAM DATA DECODER MEMORY SHARING TECHNIQUES

Paul Morton and Darwin Rambo, assignors to Broadcom Corporation
10 January 2006 (Class 704/201); filed 28 February 2002

To prevent wasteful overallocation of memory for storing compressed data, i.e., operating data and operating code in an audio decoder, a processor



dynamically allocates an appropriate amount of memory for the selected decompression algorithm. Thus, jitter buffers may be dynamically resized as the voice decoder algorithm changes.—DAP

6,950,799

43.72.Ja SPEECH CONVERTER UTILIZING PREPROGRAMMED VOICE PROFILES

Ning Bi and Andrew P. DeJaco, assignors to Qualcomm Incorporated
27 September 2005 (Class 704/261); filed 19 February 2002

The described device would alter a speech signal by modifying formant and pitch frequencies, as well as voicing and gain parameters. An LPC analysis of the input speech isolates the formants, which are converted to line-spectral-pair parameters, and a voicing detector produces pitch and gain signals. According to the patent, the device would use a more elaborate system for parameter modification than is known in the prior art. A system of ''voice fonts'' specifies pitch and formant control strategies. Formant alterations may be specified via any of several multiple/offset formulas. A ''voice font,'' as the term is used here, appears to specify formulas for the alteration of a specific input voice, not an absolute output voice quality, as would be suggested by analogy with the specification of printed text fonts.—DLR

6,947,891

43.72.Ne EFFICIENT SPEECH RECOGNITION SYSTEM BASES ON AN AUDITORY MODEL

Soo Young Lee and Chang Min Kim, assignors to Korea Advanced Institute of Science & Technology
20 September 2005 (Class 704/232); filed in the Republic of Korea 26 June 2000

The primary characteristic of the patented speech recognition system is its efficient design, intended for implementation as a custom integrated circuit (i.e., an ASIC), presumably not requiring a separate DSP chip. The front end consists of an iterated power-of-two FIR/downsample filter structure. While very efficient, this architecture is not an ideal match to the frequency characteristics of the ear. Following the filter bank is a series of feature extraction circuits, including word boundary detection, timing normalization, and a word classifier. The classifier is a neural network feeding a radial-basis-function decision system. There is no indication that the system has actually been tested.—DLR

6,947,895

43.72.Ne DISTRIBUTED SPEECH SYSTEM WITH BUFFER FLUSHING ON BARGE-IN

Saravanan Shanmugham, assignor to Cisco Technology, Incorporated
20 September 2005 (Class 704/270); filed 14 August 2001

This patent describes a buffering scheme to be used with voice recognition and synthesis equipment for an Internet browser application, in other words, a sort of two-way version of voice-over-Internet protocols. Barge-in, line turn-around, and buffer management are treated in some detail.—DLR

6,947,896

43.72.Ne TEXT MARKING FOR DEFERRED CORRECTION

Gary Robert Hanson, assignor to International Business Machines Corporation
20 September 2005 (Class 704/270); filed 15 January 2002

This patent describes techniques for processing text in a voice-recognition dictation system. The issue is that, as dictation proceeds, portions of previously entered text may be edited in various ways, such as marking for deletion or modification. Ways of using highlighting and some other marking schemes are described.—DLR

6,950,793

43.72.Ne SYSTEM AND METHOD FOR DERIVING NATURAL LANGUAGE REPRESENTATION OF FORMAL BELIEF STRUCTURES

Steven I. Ross *et al.*, assignors to International Business Machines Corporation
27 September 2005 (Class 704/9); filed 10 January 2002

This speech recognition system is set up for dialogue with a user for database access or manipulation, system control, etc. Incoming speech is recognized, parsed, and understood with respect to a reasoning facility, a formal belief structure, and an ontology constructed for the specific operational domain. A voice response is generated from the belief structure, the syntactic model, and an unspecified synthesizer. Unfortunately, the system description, which could easily take up two or three books, is dealt with by a couple of literature references and about ten columns of patent text.—DLR

6,950,796

43.72.Ne SPEECH RECOGNITION BY DYNAMICAL NOISE MODEL ADAPTATION

Changxue Ma and Yuan-Jun Wei, assignors to Motorola, Incorporated
27 September 2005 (Class 704/244); filed 5 November 2001

This speech recognition system uses a run-of-the-mill mel-cepstral analyzer feeding a hidden Markov model word recognizer. The patented feature is a method of updating a Gaussian-mixture noise model. Between-word intervals are isolated and used, first to select the closest mixture, then to update the resulting noise model. The model update procedures are presented in detail.—DLR

6,950,797

43.72.Ne VOICE REFERENCE APPARATUS, RECORDING MEDIUM RECORDING VOICE REFERENCE CONTROL PROGRAM AND VOICE RECOGNITION NAVIGATION APPARATUS

Yoshikazu Hirayama and Yoshiyuki Kobayashi, assignors to Xanavi Informatics Corporation
27 September 2005 (Class 704/246); filed in Japan 9 September 1999

This speech recognition system, organized as a vehicle navigation system, includes logic intended to improve the accuracy by allowing multiple statements of the target location. Translating the Japanese examples to the United States, it might go something like this. Say you want to go to a ski resort in Colorado. You would first say those two phrases, ''ski resort'' ''Colorado.'' You may not know that the place you want to go is in Eagle County (the Japanese example relies heavily on prefectures), but perhaps you could say ''near Vail.'' The whole point is in narrowing the search by additional location restrictions. That's it.—DLR

6,952,155

43.72.Ne VOICE-CONTROLLED SECURITY SYSTEM WITH PROXIMITY DETECTOR

Richard B. Himmelstein, Kure Beach, North Carolina
4 October 2005 (Class 340/5.52); filed 10 December 2002

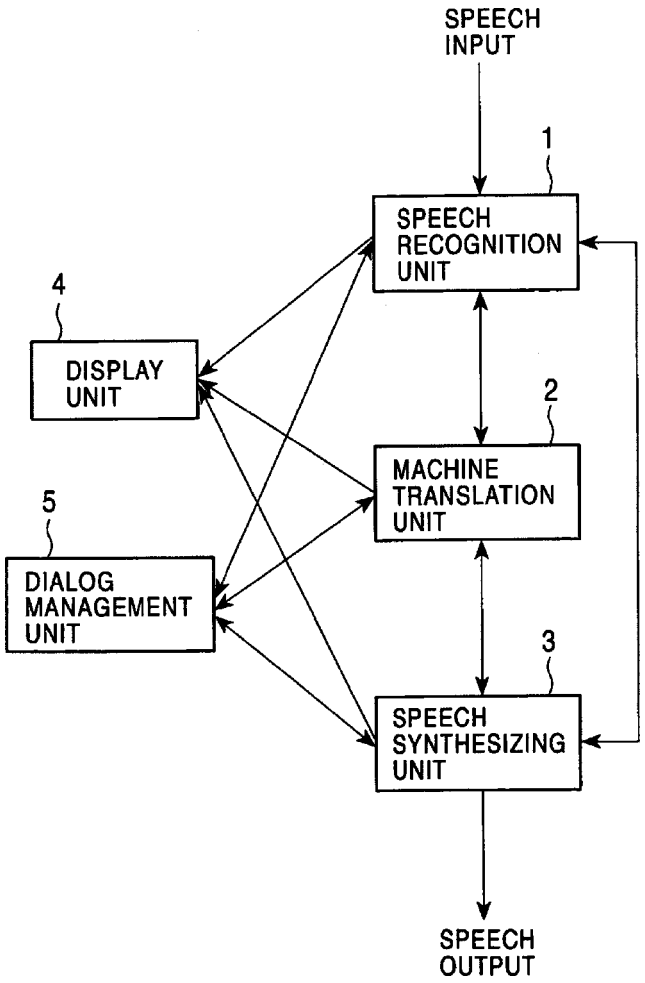
The patent describes a voice-controlled vehicle lock. Improved security is supposedly provided by a combination of interactions. First, the user must approach to within a preset distance of the vehicle and push an ID-code button. If the ID is accepted, a microphone picks up the user's voice request, which may include a command to unlock particular doors, etc. The voice is checked against the user's stored voice pattern. If that test succeeds, then the command is performed.—DLR

6,952,665

43.72.Ne TRANSLATING APPARATUS AND METHOD, AND RECORDING MEDIUM USED THEREWITH

Hideki Shimomura *et al.*, assignors to Sony Corporation
4 October 2005 (Class 704/2); filed in Japan 30 September 1999

The described device would translate speech from one language to another. Incoming speech is recognized in an unspecified manner, parsed into a lexical-functional grammar (LFG) structure, and applied to a vague semantic analyzer. A dialog manager then guides an unspecified synthesizer to produce the target-language output. As dialog descriptions accumulate in



the manager, they may be consulted to help disambiguate new inputs. Inputs and outputs could also be handled in a text format, rather than as speech. Overall, the system is described almost as a two-sided, query/response system. In theory, the LFG grammar structure could provide sufficient mechanism to actually make this work, depending on the semantic structure used. But, not enough information is provided to estimate how much of this is achievable and how much is a pipe dream. For example, there is no discussion of how or whether the past dialog information would be available to influence the parsing process, as, for example, to expand ellipses and anaphoric references.—DLR

6,975,985

43.72.Ne METHOD AND SYSTEM FOR THE AUTOMATIC AMENDMENT OF SPEECH RECOGNITION VOCABULARIES

Werner Kriechbaum and Gerhard Stenzel, assignors to International Business Machines Corporation
13 December 2005 (Class 704/231); filed in the European Patent Office 29 November 2000

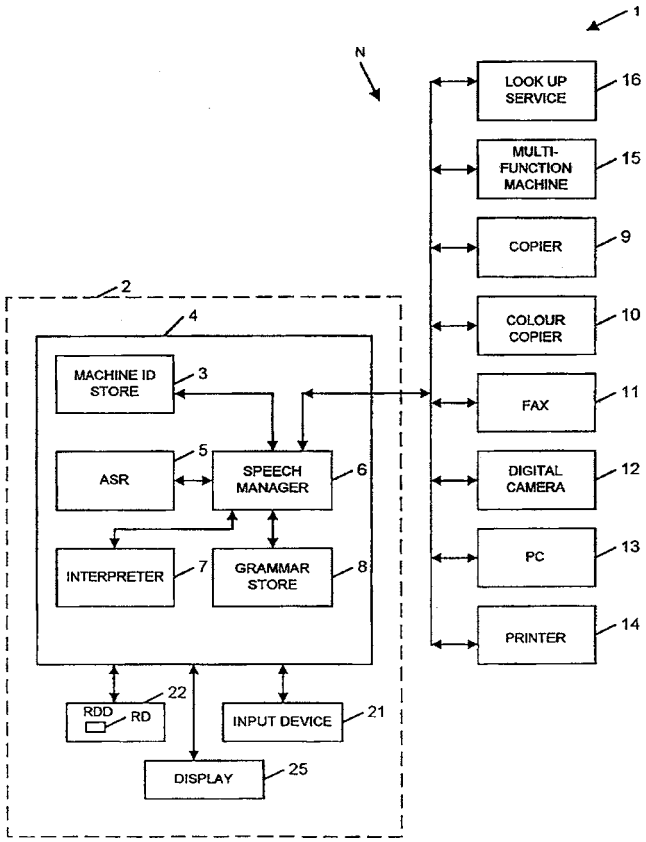
The idea here is quite simple: modeled pronunciations of words in the stored vocabulary of a speech recognition system can be updated to match user-input pronunciations when desired, by time-aligning a user's utterance with a textual representation of the utterance.–SAF

6,975,993

43.72.Ne SYSTEM, A SERVER FOR A SYSTEM AND A MACHINE FOR USE IN A SYSTEM

Robert Alexander Keiller, assignor to Canon Kabushiki Kaisha
13 December 2005 (Class 704/275); filed in the United Kingdom 21 May 1999

Speech recognition is used for controlling, with spoken commands, various types of processor-controlled, networked office equipment. Predefined commands corresponding to functions performed by the machines



are stored in a server for lookup during use. The system may be trained to the voice of the person issuing the commands. This determines whether a particular machine can perform the requested command. Commands may be provided via a telephone system associated with the network.—DAP

6,980,952

43.72.Ne SOURCE NORMALIZATION TRAINING FOR HMM MODELING OF SPEECH

Yifan Gong, assignor to Texas Instruments Incorporated
27 December 2005 (Class 704/234); filed 7 June 2000

To reduce the deleterious effects on speech recognizers of environmental variations (e.g., differences in speakers, handsets, transmission channels, and noise backgrounds), a maximum-likelihood, linear-regression solution is used for environmental normalization. A joint solution is obtained to the canonical distributions, transformations, and biases for source normalization, making it unnecessary to label a database in terms of environment.—DAP

6,980,954

43.72.Ne SEARCH METHOD BASED ON SINGLE TRIPHONE TREE FOR LARGE VOCABULARY CONTINUOUS SPEECH RECOGNIZER

Quingwei Zhao *et al.*, assignors to Intel Corporation
27 December 2005 (Class 704/251); filed 30 September 2000

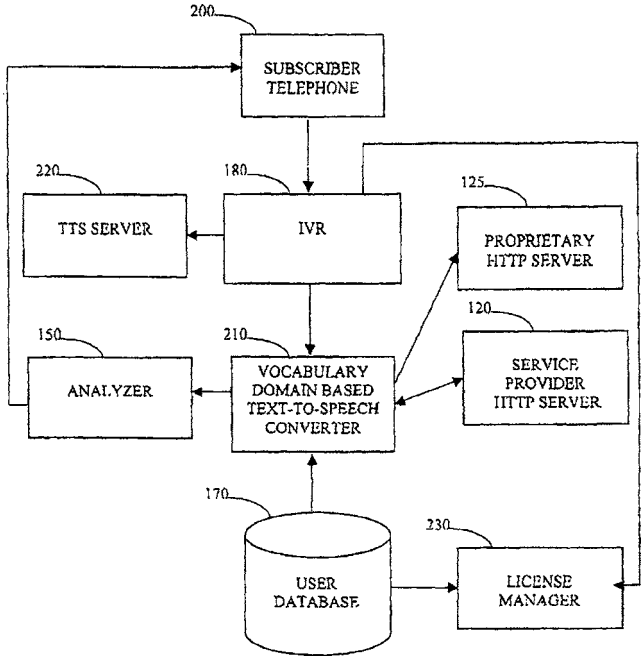
This patent presents, in complete pseudocode, an apparently novel search algorithm that allows a single phonetic tree, expressed using triphones to suffice as a language model for an entire multiword utterance, thereby facilitating a one-pass search. Prior art methods generally require a separate phonetic tree model for each word and an iterated search with tree copying, so the present proposal promises a considerable savings in memory and computational complexity. The core innovation is the propagation of actual tokens through the lexicon tree, with history information recorded in the tokens.—SAF

6,983,250

43.72.Ne METHOD AND SYSTEM FOR ENABLING A USER TO OBTAIN INFORMATION FROM A TEXT-BASED WEB SITE IN AUDIO FORM

David Guedalia and Jacob Guedalia, assignors to NMS Communications Corporation
3 January 2006 (Class 704/260); filed 22 October 2001

Automatic text-to-speech conversion in an interactive voice-response system is performed for text containing menus and content that are used in Internet web sites. Multiple text-to-speech converters are used to define and



utilize at least one vocabulary domain. A content-based closeness metric is used in the analysis to select the vocabulary domains. The resulting speech, for example, local weather or news, is transmitted to the user as audio over a telephone system.—DAP

6,985,859

43.72.Ne ROBUST WORD-SPOTTING SYSTEM USING AN INTELLIGIBILITY CRITERION FOR RELIABLE KEYWORD DETECTION UNDER ADVERSE AND UNKNOWN NOISY ENVIRONMENTS

Philippe R. Morin, assignor to Matsushita Electric Industrial Company, Limited
10 January 2006 (Class 704/240); filed 28 March 2001

A word or multiword recognition approach is very briefly described. The key idea is to generate frame-by-frame the usual kind of recognition score for a word, along with a ''background score'' that is to be obtained by forced matching of a word model to the background signal (a period of no speech input). Based on matching ratios between minimum recognition values and background scores, the system then caclulates frame-by-frame confidence scores which are said to, in essence, track ''noise-corrected likelihoods that the words are in the speech signal.'' Although the method can in theory be used to spot any number of words in the lexicon, it is truly limited to spotting rather than large-vocabulary recognition for reasons of complexity.—SAF

6,985,861

43.72.Ne SYSTEMS AND METHODS FOR COMBINING SUBWORD RECOGNITION AND WHOLE WORD RECOGNITION OF A SPOKEN INPUT

Jean-Manuel Van Thong and Ernest Pusateri, assignors to Hewlett-Packard Development Company, L.P.
10 January 2006 (Class 704/254); filed 12 December 2001

A method is outlined to address the unwanted pruning of intended words from search trees by conventional speech recognition systems, which frequently happens when a user's pronunciation is significantly different from those listed in a reference pronunciation dictionary. The approach here is to perform simultaneous decoding searches over words and subword units, thereby creating with a simple likelihood ranking a ''fused'' *N*-best list of hypotheses arising from both lexical and subword pronunciation. The patent promises improved performance in applications, and further envisions splitting the two searches across platforms, with the vocabulary-independent, subword search being performed client-side, leaving the more intensive word search to a central server.—SAF

6,978,238

43.72.Pf METHOD AND SYSTEM FOR IDENTIFYING A USER BY VOICE

Robert C. Wohlsen, Menlo Park and Sue McNeill, Mendocino, both of California
20 December 2005 (Class 704/246); filed 12 July 1999

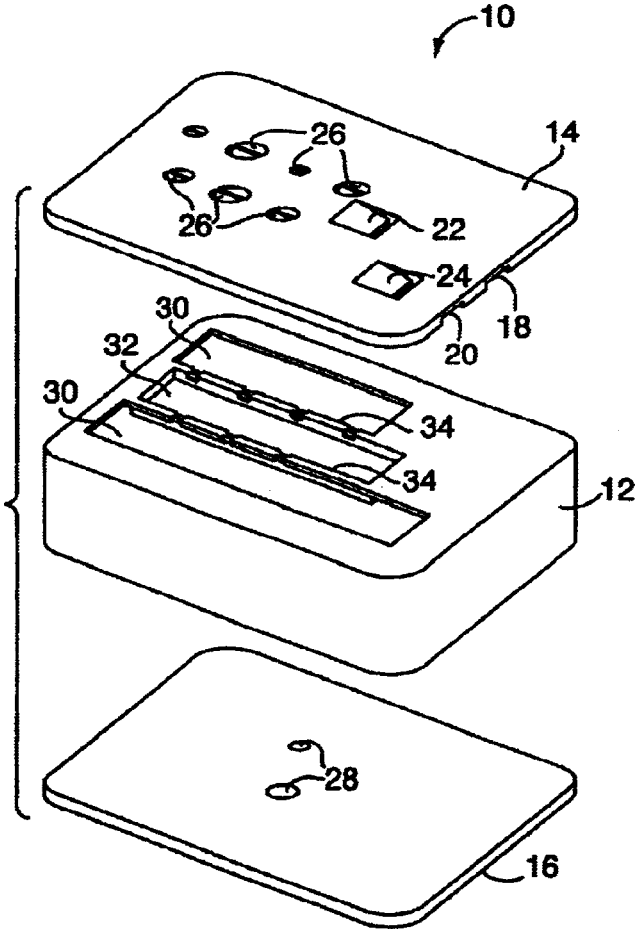
The patented method here is really just a way of employing the sound of a user's name and password in a verification setting with a conventional (not covered by the patent) speaker identification algorithm's output, curiously called a ''voiceprint.'' The idea is simply to pare down the number of users whose secure-access utterances will be compared by the ''voiceprinting'' procedure, by comparing only with those users whose securer-access utterances bear a similar structure according to the speech recognition grammar.—SAF

6,914,179

43.75.Ef FLUTES WITH MULTIPLE CHAMBERS THAT SHARE COMPOUND BISECTED TONEHOLES

Karl P. Ahrens, assignor to Mountain Ocarinas Incorporated
5 July 2005 (Class 84/384); filed 25 April 2003

The ocarina continues to be improved. In this disclosure, the inventor uses multiple cavities, which provide a wider range. The use of ''bisected'' toneholes **26**, i.e., holes that are divided between two chambers **30** and **32**,



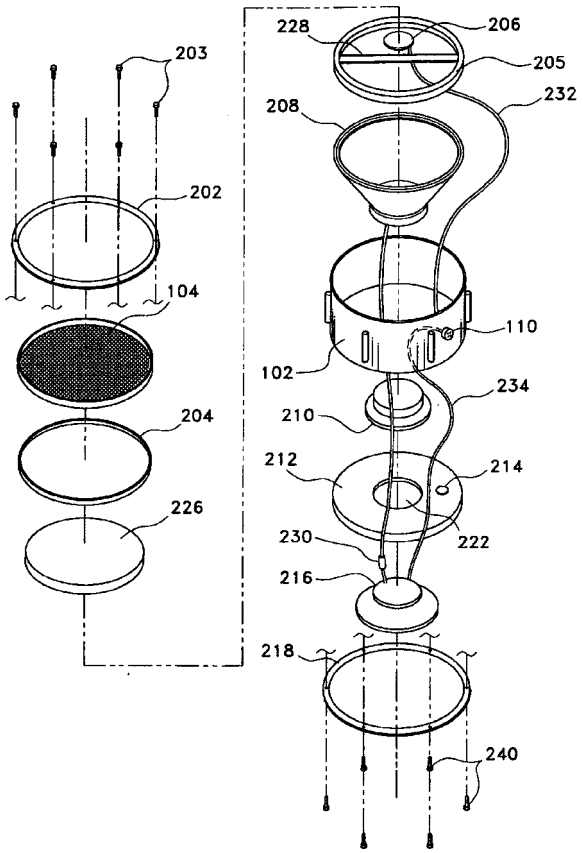
provide pitch control for multiple chambers. Openings **18** and **20** are airways leading to the knife edge fipples **22** and **24**. Lower tone holes **28** are octave holes well known to woodwind players.—MK

6,982,376

43.75.Hi REAL DRUM TRIGGER MONITOR AND AMPLIFIED TONE MODULE

Johnathan R. Wise, Grafton, Ohio
3 January 2006 (Class 84/600); filed 17 October 2003

The idea here is to use a sensor to control the drum attack from a synthetic source. A sensor **206** mounted on a drum head can be connected to



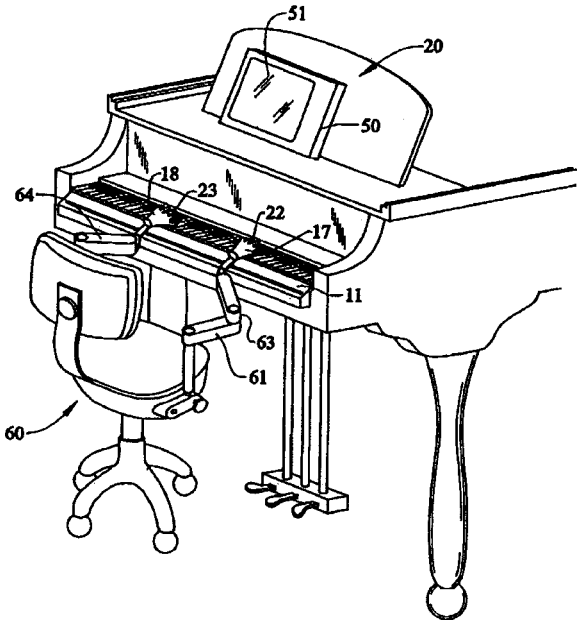
speakers **208** and **216** mounted internally in the shell **102**. Of course MIDI control is also provided. Naturally, this drum is not closed on the bottom.—MK

6,982,375

43.75.St MUSICAL TEACHING DEVICE AND METHOD

Rob McGregor, West Palm Beach, Florida
3 January 2006 (Class 84/478); filed 6 June 2003

This conceptual patent uses movable gloves **17** and **18** that slide along the keyboard. (The learner inserts their hands inside the gloves.) Somehow, "noninjurious electrical pulses" control the fingers. Somehow, the computer **51** runs the whole show. Should be fun to watch—hopefully the gloves are white velvet.—MK

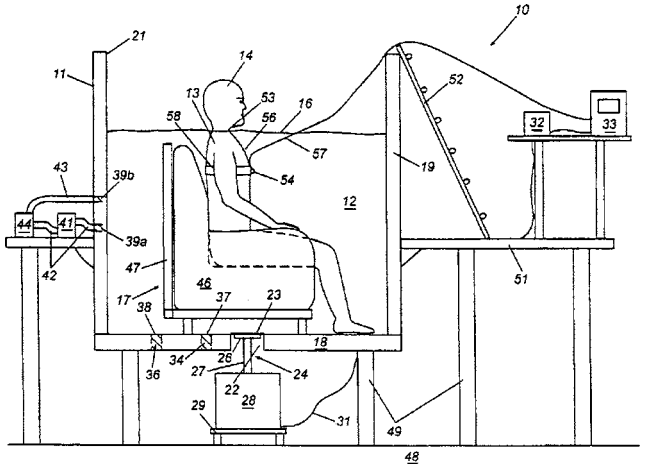


6,974,425

43.80.Qf APPARATUS AND METHOD FOR IMPLEMENTING HYDRO-ACOUSTIC THERAPY FOR THE LUNGS

Peter H. Rogers *et al.*, assignors to Georgia Tech Research Corporation
13 December 2005 (Class 601/46); filed 26 April 2001

The apparatus provides hydro-acoustic therapy (HAT) for a patient's lungs. HAT entails treatment through the use of acoustic vibrations in a liquid environment. The HAT apparatus consists of a chamber filled with a



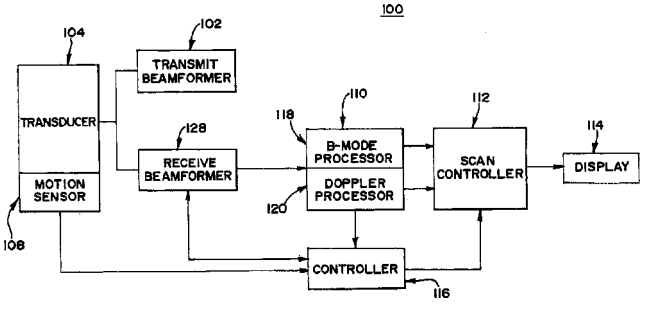
fluid (in particular, water) and an acoustic generator for generating low-frequency acoustic waves in the fluid. For treatment, a patient is partially submerged up to his/her neck in the chamber. The vibrations, if properly tuned, cause the mobilization of respiratory secretions.—DRR

6,976,961

43.80.Qf TISSUE MOTION ANALYSIS MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD

John I. Jackson *et al.*, assignors to Acuson Corporation
20 December 2005 (Class 600/443); filed 13 February 2003

This is a medical diagnostic ultrasound system for tissue motion analysis. The method accounts for the motion of an ultrasonic transducer while providing estimates of tissue motion. Correcting for transducer motion helps to isolate localized tissue contractions or expansions, such as the motion of the myocardial muscle or fibers. Motion estimation is also provided by determining an angle of motion from the ultrasound data. The angle of



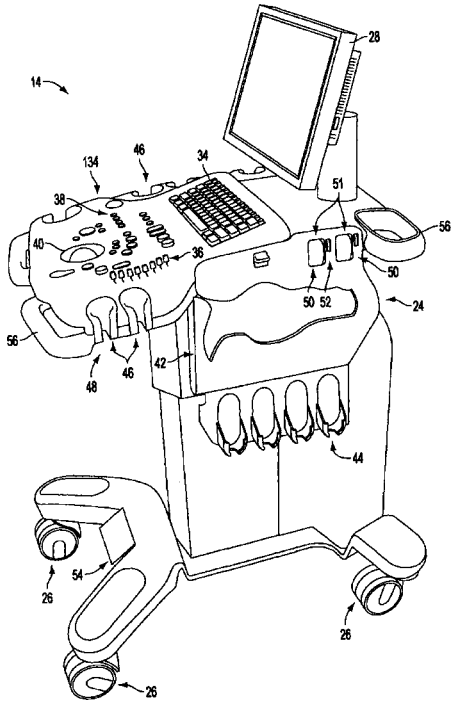
motion is used to adjust velocity estimates, providing two-dimensional velocity vectors (thereby providing estimates of motion in two dimensions). Movement of tissue is established by correlating a speckle or a feature represented by two different sets of ultrasound data obtained at different times. Additional aspects include tracking the location of a tissue of interest, which aids in estimating strain and rate of strain.—DRR

6,980,419

43.80.Qf PORTABLE ULTRASOUND UNIT AND DOCKING STATION

Scott F. Smith *et al.*, assignors to ZONARE Medical Systems, Incorporated
27 December 2005 (Class 361/681); filed 12 March 2003

The principle feature of this ultrasound unit is a portable unit mounted to a docking cart and which can be detached for extreme portability in diagnostic use. The portable unit can be mounted to the cart to transform it into a cart-based system with enhanced features and functionality, such as



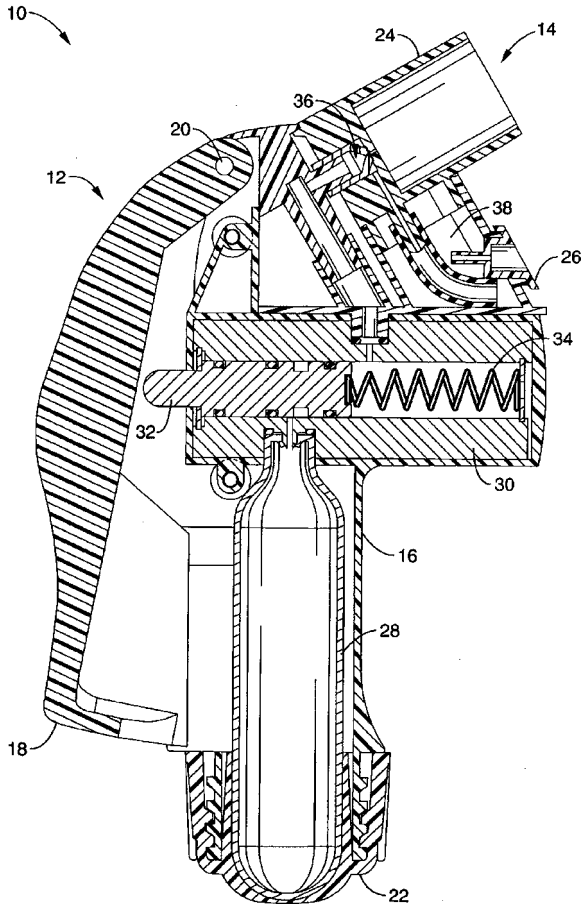
improved ergonomics, ease of use, a larger display format, external communications connectivity, multiple transducer connections, and increased data processing capacity. The docking cart may also contain adjustable interface controls, data handling features, security features, power control functions, and thermal management capabilities.—DRR

6,981,660

43.80.Qf SHOCK WAVE AEROSOLIZATION APPARATUS AND METHOD

Samuel David Piper, assignor to Evit Labs
3 January 2006 (Class 239/433); filed 13 June 2003

We have here a pneumatic inhaler that can deliver a controlled burst or dose of aerosol from a reservoir of liquid or powder medication. A supersonic jet of gas is emitted from a nozzle, creating shock waves in the jet. In one embodiment, the jet is directed into a shock chamber. Liquid or micron-



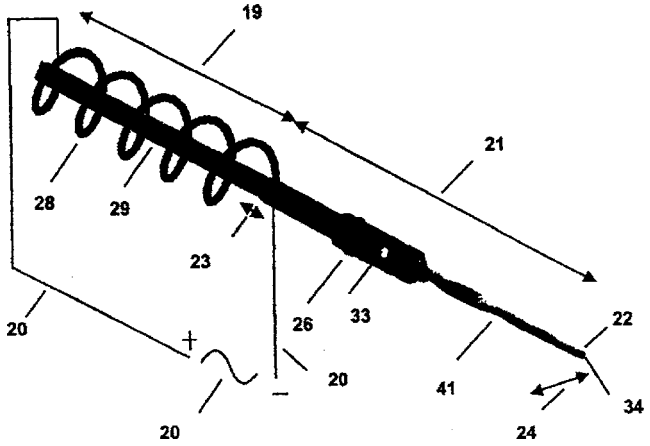
ized powder material is channeled into the jet to form an aerosol. In another embodiment, smaller aerosol particles are separated from larger particles with a separator. In a third embodiment, the aerosol is contained in an aerosol storage chamber prior to inhalation by the user.—DRR

6,984,220

43.80.Sh LONGITUDINAL-TORSIONAL ULTRASONIC TISSUE DISSECTION

David G. Wuchinich, Yonkers, New York
10 January 2006 (Class 604/22); filed 11 April 2001

This device is essentially a mechanically resonant, longitudinal/torsional vibrator that dissects biological tissue. The ultrasound apparatus provides longitudinal and torsion motion of tips, together with means of



irrigation and aspiration. Use can be made of common and fairly inexpensive electro-mechanical transducers for the production of such motion through the use of longitudinal-torsional resonators.—DRR

6,974,415

43.80.Vj ELECTROMAGNETIC-ACOUSTIC IMAGING

Stephen Anthony Cerwin *et al.*, assignors to Magnetus LLC
13 December 2005 (Class 600/437); filed 22 May 2003

A pulsed electromagnetic field is used to produce ultrasonic waves in regions of electrical conductivity gradients or discontinuities. The ultrasonic waves are detected by one or more sensors. Signals corresponding to the detected waves are then displayed or processed to provide information about the regions in which the ultrasound is produced.—RCW

6,974,417

43.80.Vj ULTRASOUND TRANSDUCER ARRAY

Geoffrey R. Lockwood and Christine E. Morton, assignors to Queen's University at Kingston
13 December 2005 (Class 600/459); filed 4 October 2002

This array, which can be used to transmit or receive ultrasound energy, consists of a hard substrate with an array of electrodes without grooves between the electrodes on one side and with a single electrode covering the other side.—RCW

6,979,294

43.80.Vj SPLIT-SCREEN DISPLAY SYSTEM AND STANDARDIZED METHODS FOR ULTRASOUND IMAGE ACQUISITION AND PROCESSING FOR IMPROVED MEASUREMENTS OF VASCULAR STRUCTURES

Robert H. Selzer and Howard N. Hodis, assignors to California Institute of Technology
27 December 2005 (Class 600/450); filed 13 December 2002

A split-screen display is used to show an arterial ultrasound image from an earlier examination on one side of the screen and a real-time ultrasound image from a current examination on the other side of the screen. Echo edge recognition and tracking methods are used to identify ultrasound echo boundaries of the imtima-media complex and to extract arterial dimensions automatically. Measurement accuracy is enhanced using subpixel resolution for the definition of echo boundaries.—RCW

6,979,295

43.80.Vj AUTOMATIC COLOR GAIN ADJUSTMENTS

David Thomas Dubberstein and Lihong Pan, assignors to GE Medical Systems Global Technology Company, LLC
27 December 2005 (Class 600/455); filed 19 November 2003

Values of the time-gain compensation curve in a system are multiplied with corresponding image values to compensate for attenuation with increasing range in the image. The time-gain compensation curve can be averaged to reduce localized gain variations. Values of the time-gain compensation curve can be used to adjust front-end gain.—RCW

6,984,209

43.80.Vj HARMONIC MOTION IMAGING

Kullervo H. Hynynen and Elisa Konofagou, assignors to The Brigham and Women's Hospital, Incorporated
10 January 2006 (Class 600/438); filed 2 July 2003

Two intersecting ultrasonic beams are used to produce vibration in the region of intersection. A third beam is used to transmit energy into the vibrating region and to receive signals from the region. Spectral analysis of the received signals yields information about the vibration of the region.—RCW

6,984,211

43.80.Vj DETECTION OF TUMOR HALOS IN ULTRASOUND IMAGES

Xiaohui Hao *et al.*, assignors to Mayo Foundation for Medical Education and Research
10 January 2006 (Class 600/443); filed 2 January 2004

A halo in an ultrasonic image of a liver tumor is evaluated by detecting the boundary of the tumor, defining two annular regions around the boundary, and comparing the brightness of pixels in the two regions to determine if a halo of darker pixels surrounds the tumor.—RCW

LETTERS TO THE EDITOR

Comments on "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations" [J. Acoust. Soc. Am. 114, 66–69 (2003)] (L)

Renata S. Sousa-Lima[a)]
Bioacoustics Research Program, Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, New York 14850

(Received 6 February 2006; accepted 10 March 2006)

This letter concerns the paper "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations" [Nowacek *et al.*, J. Acoust. Soc. Am. **114**, 66–69 (2003)]. The purpose here is to correct the fundamental frequency range and information on intraindividual variation in the vocalizations of Amazonian manatees reported by Nowacek *et al.* (2003) in citing the paper "Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia)" [Sousa-Lima *et al.*, Anim. Behav. **63**, 301–310 (2002)]. [DOI: 10.1121/1.2195047]

PACS number(s): 43.80.Ka [WWA] Pages: 3537–3537

The Sousa-Lima *et al.* (2002) article concerning identity in the isolation calls of Amazonian manatees (*Trichechus inunguis*) suggested that the fundamental frequency range of this species is broader than previously reported by Evans and Herald (1970) and Sonoda and Takemura (1973). The Sousa-Lima *et al.* (2002) article states that "…the fundamental frequency range in the present study was increased from 6–8 kHz (Evans and Herald, 1970) and 2–4 kHz (Sonoda and Takemura, 1973) to 1.07–8 kHz."

Nowacek *et al.* (2003) incorrectly reported the fundamental frequency range and did not acknowledge the extensive discussion of intraspecific variation that confers identity to Amazonian manatee vocalizations. Nowacek *et al.* (2003) stated: "Manatee sound production has been documented in only four papers (Evans and Herald, 1970; Schevill and Watkins, 1965; Sonoda and Takemura, 1973; Sousa-Lima *et al.*, 2002), including very little discussion of variability inherent in the animals' vocal repertoire. These reports describe several types of sounds, but the primary vocalization recorded was a tonal sound often having several harmonics with the second or third harmonic often stronger than the fundamental frequency; intraspecific variation was not reported. The fundamental frequency ranged from 2.5 to 5 kHz for the West Indian manatee, *Trichechus manatus spp.*, and 2.6 to 5.9 kHz for the Amazonian manatee, *Trichechus inunguis*."

It is important to acknowledge these oversights and correct the information presented for readers. This comment is further warranted by the fact that, similarly, incorrect fundamental frequency range has been reported in a recently published textbook (Berta *et al.* 2006). In this publication, when citing Sousa-Lima *et al.* (2002) and Evans and Herald (1970), the frequency range for manatees is given incorrectly as 2.5–8 kHz, when it should be 1.07–8 kHz. Future printings of Berta *et al.* (2006) are expected to include the correct fundamental frequency range for manatees (Berta, personal communication).

I am most thankful to Jason A. Mobley, Selvino de Kort, Danielle Cholewiak, and Christopher W. Clark for suggestions on this comment.

Berta, A., Sumich, J. L., and Kovacs, K. M., (**2006**). *Marine Mammals: Evolutionary Biology* (Academic, New York), p. 301.

Evans, W. E., and Herald, E. S., (**1970**). "Underwater calls of a captive Amazon manatee, Trichechus inunguis," J. Mammal. **51**, 820–823.

Nowacek, D. P., Casper, B. M., Wells, R. S., Nowacek, S. M., and Mann, D. A., (**2003**). "Intraspecific and geographic variation of West Indian manatee (*Trichechus manatus spp.*) vocalizations," J. Acoust. Soc. Am. **114**, 66–69.

Sonoda, S., and Takemura, A., (1973). "Underwater sounds of the manatees, *Trichechus manatus manatus and Trichechus inunguis* (Trichechidae)," Reports of the Institute for Breeding Research (University of Agriculture, Tokyo), Vol. **4**, pp. 19–24.

Sousa-Lima, R. S., Paglia, A. P., and da Fonseca, G. A. B., (**2002**). "Signature information and individual recognition in the isolation calls of Amazonian manatees, *Trichechus inunguis* (Mammalia: Sirenia)," Anim. Behav. **63**, 301–310.

[a)]Electronic mail: rsl32@cornell.edu

Comparison of Levitt- and Zwislocki-type adaptive procedures for stimulus placement in human listeners (L)[a]

Daniel Rowan,[b] Kathryn Hinton,[c] and Emma Mackenzie
Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, United Kingdom

(Received 14 February 2005; revised 13 March 2006; accepted 13 March 2006)

The "3-down, 1-up" adaptive stimulus placement rules attributed to Levitt and Zwislocki were compared in the context of human interaural time difference discrimination, employing otherwise commonly used procedures (e.g., averaging reversals to estimate threshold). The Zwislocki rule typically accumulated over three more reversals than the Levitt rule but the rules were approximately equally efficient. This may be because the additional reversals did not provide additional information or that any additional information was countered by other factors, such as the lower theoretical asymptotic response probability (0.75 vs 0.794). Relative bias between rules was also explored with the aid of a simulation. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2195088]

PACS number(s): 43.66.Yw, 43.66.Pn [RAL] Pages: 3538–3541

I. INTRODUCTION

Adaptive procedures are commonly used in the measurement of psychophysical thresholds, where a "stimulus placement rule" determines the stimulus magnitude on any one trial according to the values and the listener's responses on previous trials. Levitt (1971) described a class of stimulus placement rule whereby the stimulus magnitude is reduced following one "correct" response on k consecutive trials and increased following an "incorrect" response, denoted $L_{k,1}$. The theoretical asymptotic response probability depends on k, this being 0.707 for $L_{2,1}$ and 0.794 for $L_{3,1}$ ($\sqrt[k]{0.5}$). This remains the most popular class of stimulus placement rule despite evidence that it is less efficient (i.e., is associated with greater variance with repeated measurements over a given number of trials) than alternative rules (e.g., Marvit *et al.*, 2003; Leek, 2001) largely because of its simplicity.

Zwislocki *et al.* (1958) described a stimulus placement rule that differs from $L_{k,1}$ only in that it does not require the k trials to be consecutive, denoted $Z_{k,1}$. The $Z_{3,1}$ and $L_{3,1}$ rules are illustrated in Fig. 1. The theoretical asymptotic probability for $Z_{3,1}$ is 0.75, corresponding to the classical definition of threshold for two-alternative forced choice (2AFC) tasks. Although $Z_{3,1}$ has been used only sporadically since, it may offer advantages over $L_{k,1}$ with little additional complexity. For example, Zwislocki and Relkin (2001) suggested that, in counting nonconsecutive correct responses, $Z_{k,1}$ may, for a given value of k, accrue a larger number of reversals than $L_{k,1}$ for a given number of trials. Since reversals are typically averaged to estimate threshold, this may enhance efficiency. Two studies have evaluated $Z_{k,1}$ in terms of efficiency and bias using computer simulations (García-Pérez, 2002; Schlauch and Carney, 2003). However, it has been shown that results from simulations may deviate markedly from results with human listeners (Marvit *et al.*, 2003). The primary aims of the present study were to demonstrate in humans that $Z_{k,1}$ yields more reversals than $L_{k,1}$ for $k=3$ and evaluate whether this enhances efficiency given the typical approach to calculating threshold. This was conducted in the context of interaural time difference (ITD) discrimination (see Hafter and Trahiotis, 1997). While a previous simulation study has investigated $L_{k,1}$ in this context (Saberi and Green, 1997), we were unaware of such studies in humans. A secondary aim of this study was to gain further insight into the bias associated with $Z_{k,1}$ and $L_{k,1}$ by comparing the human data to computer simulations.

II. METHOD

Stimuli were generated digitally, with a 11025-Hz sampling rate and 16-bit amplitude resolution, and procedures controlled using MATLAB (MathWorks, Inc., r.12). Stimuli were played-out using a soundcard (Creative, Extigy) connected to an audiometer (Kamplex, KC50) and delivered to listeners via insert earphones (Etymotic Research Inc., ER-3A). Stimuli were 400-ms-long tones at 500 Hz presented at 70 dB SPL. An ITD was applied to one channel prior to imposing 40-ms onset and offset cosine-squared ramps in both channels simultaneously. The general approach to measuring ITD thresholds was comparable to many previous experiments (e.g., Bernstein *et al.*, 1998; Bernstein and Trahiotis, 2002; Koehnke *et al.*, 1995). The paradigm was 4-interval, 2-cue, 2AFC: the tone containing the ITD appeared in either the second or third interval, selected randomly; the remaining intervals contained diotic tones. Listeners were instructed to identify the "odd-one-out" from the middle two intervals. Visual feedback was provided after each trial. The mean initial ITD for each adaptive block was a factor of 4 (three large step sizes) above preliminary estimates of ITD threshold obtained during a "training" session,

[a]Portions of this work were presented at the British Society of Audiology Short Papers Meeting on Experimental Studies of Hearing and Deafness in London, UK, 16–17 September 2004.
[b]Electronic mail: dr@isvr.soton.ac.uk
[c]Current address: City Hospital, Dudley Rd, Birmingham, UK, B18 7QH.

 0001-4966/2006/119(6)/3538/4/$22.50

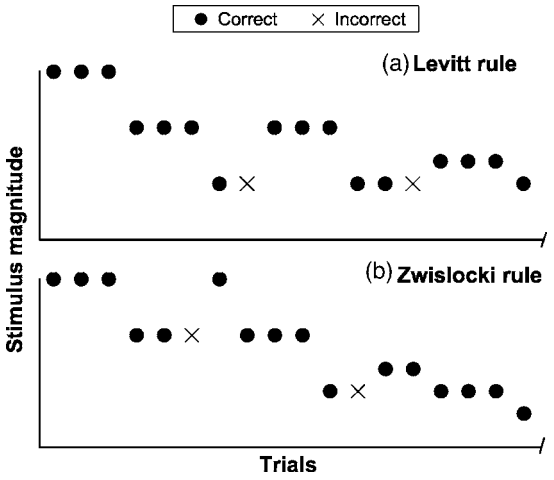


FIG. 1. Portions of the adaptive tracks for the Levitt (a) and Zwislocki (b) 3-down, 1-up stimulus placement rules. The Zwislocki rule differs from the Levitt rule in not requiring the three correct responses to be consecutive before reducing the stimulus magnitude.

with a random component of up to ±100 μs. The step size was a factor of 1.584 for the first three reversals and 1.122 thereafter (Saberi, 1995). The stimulus placement rule was either $Z_{3,1}$ or $L_{3,1}$ and blocks were terminated after 60 trials. The first four reversals were discarded, as was the fifth if necessary to leave an even number. The threshold estimate was defined as the geometric mean of the remaining ("scored") reversals. Twenty-four normal-hearing 18–30 -year-olds (mean age 25.6 yr; 10 females) were recruited from the University population, most having experience of psychoacoustic experiments. Following training, listeners were tested in two sessions spaced a week apart. Within each session, listeners completed one familiarization block with each rule in random order (data discarded). Next, two test blocks were completed with one rule and then two with the other. Half of the listeners received $Z_{3,1}$ first. The sessions were identical except that the order of testing was reversed.

III. RESULTS

Taken across the four test blocks, $L_{3,1}$ yielded 2-12 reversals with a mean of 6.7 whereas $Z_{3,1}$ yielded 4-14 reversals with a mean of 9.8. The number of reversals did not vary significantly across blocks or sessions for either rule. An unrelated-samples *t*-test (homogeneity of variance assumed) indicated that the difference in overall mean reversals between rules was statistically significant ($t_{46}=8.64$, $p<0.001$).

A logarithmic transformation was applied to the ITD thresholds to produce normally distributed data. A repeated-measures ANOVA with factors *rule* and *block* (four per rule) indicated that $L_{3,1}$ yielded significantly higher thresholds than $Z_{3,1}$ ($F_{1,23}=8.49$, $p=0.008$), consistent with the higher theoretical asymptotic probability for $L_{3,1}$. It also indicated that thresholds were stable across the experiment for both rules, the main effect of *block* ($F_{3,69}=8.49$, $p=0.26$) and the *rule-block* interaction ($F_{3,69}=0.25$, $p=0.87$) being nonsignificant. Figure 2 presents the distribution of the geometric mean ITD thresholds (taken across the four replications) for

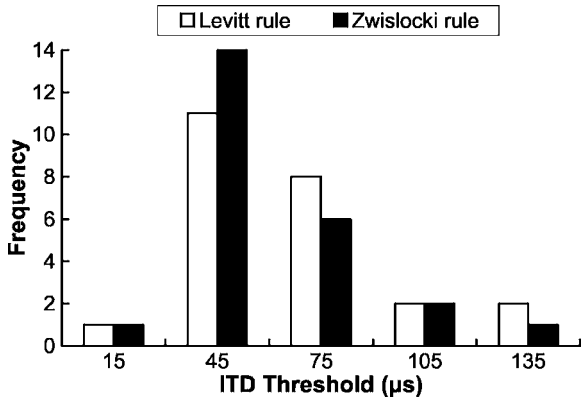


FIG. 2. Distributions of ITD thresholds measured with Levitt and Zwislocki rules ($n=24$); geometric means were 44.0 and 38.6 μs, respectively. The values on the abscissa indicate the midpoints of the bins.

each stimulus placement rule. The overall geometric mean ITD thresholds with $L_{3,1}$ and $Z_{3,1}$ were 44.0 and 38.6 μs, respectively, which compare well with previous studies (e.g., Bernstein *et al.*, 1998; Hafter and Trahiotis, 1997).

The measure of efficiency used here was essentially test-retest variability over 60 trials. To calculate this, the threshold from one block was subtracted from that from another block (with the same rule) for each listener, and then the variance calculated across listeners. Within-session variance was calculated separately for the two sessions and between-session variance was calculated separately for the two blocks in each session, producing four variances for each rule. The ratios of these for $L_{3,1}$ and $Z_{3,1}$ were obtained and subjected to *F*-tests, the results of which are given in Table I. While $Z_{3,1}$ accrued a greater number of scored reversals than $L_{3,1}$, this did not translate to enhanced efficiency. In fact, $L_{3,1}$ was associated with lower variance in all but one condition.

It was not possible to evaluate the *absolute* bias associated with each rule, as psychometric functions (PFs) were not measured. To gain some insight into potential bias, theoretical PFs were fitted to the overall geometric mean ITD thresholds for each rule, assuming the threshold corresponded to the theoretical asymptotic probability. The PF was the upper half of the cumulative normal function (Saberi, 1995; Saberi and Green, 1997). The standard deviation (s.d.) of this function was found to be 53.7 and 57.2 μs for $L_{3,1}$ and $Z_{3,1}$, respectively. Saberi (1995) described that two PFs of this form "differ only by a multiple of their standard deviations;" the ratio of the s.d.s is therefore a measure of the

TABLE I. The ratio of the within- or between-session variances in the ITD thresholds for Levitt (σ_L^2) and Zwislocki (σ_Z^2) rules and the results of an *F*-test ($df=23, 23$). "Within 1" and "Within 2" refer to test-retest variances for each of two test sessions separated by one week. "Between 1" and "Between 2" refer to test-retest variances for each of the two replications across each session.

	σ_L^2/σ_Z^2	p
Within 1	0.22	<0.001
Within 2	1.16	0.72
Between 1	0.85	0.71
Between 2	0.54	0.15

separation of two PFs. If neither rule was biased or both were biased by comparable amounts the s.d. from each PF would be approximately equal. However, the s.d.s were not equal and the ratio, expressed as a percentage, was 6.5%; this is referred to as "relative" bias. This indicates that the measured threshold for one rule and that predicted by the PF fitted to the other differed by 6.5%. To gain insight as to the origin of this relative bias, Monte Carlo simulations were conducted. The parameters of the adaptive procedures were identical to those used in the human study except 2000 blocks were completed for each rule. The form of the PF was as noted earlier and assumed to be stationary; a simulation of inattention was not included. The data did not show a systematic dependency on the s.d. of the PF when varied over a wide range, since the initial ITD and step sizes were essentially constant relative to the s.d. It was found that the mean threshold differed significantly from that expected from the theoretical asymptotic probability for $L_{3,1}$ ($t_{1999}=8.78$; $p<0.001$) but not $Z_{3,1}$ ($t_{1999}=0.58$; $p=0.57$). The actual asymptotic probabilities for $L_{3,1}$ and $Z_{3,1}$ were 0.780 and 0.751, respectively. When PFs were fitted to the mean (simulation) thresholds for the two rules, as described earlier for the human data, the ratio of the s.d.s indicated a relative bias of 6.4%, almost identical to that measured in the human experiment.

IV. DISCUSSION

This study confirmed Zwislocki and Relkin's (2001) assertion that $Z_{3,1}$ yields more scored reversals than $L_{3,1}$. However, this did not translate to enhanced efficiency (defined as test-retest variability over a given number of trials). One possibility is that these additional reversals did not provide any additional information. It is known that successive reversals within a block with $L_{k,1}$ can be highly correlated (Levitt, 1971), and it seems likely that this also applies to $Z_{k,1}$, at least to some degree. Alternatively, the rules may differ in ways that countered any advantage of the additional reversals, such as asymptotic probability. Previous research (e.g., Saberi and Green, 1996, 1997; Leek, 2001) suggests that efficiency reduces with increasing asymptotic probability between 0.7 and 0.8 for many PFs. This would predict that $L_{3,1}$, which targets the higher probability, would have greater efficiency, all else being equal. The additional reversals with $Z_{3,1}$ may therefore be insufficient to overcome the disadvantage of a lower asymptotic probability, at least under the conditions of this experiment. Whether these considerations apply to alternative conditions is unclear, since many aspects of adaptive procedures may influence efficiency. For example, while the method used to calculate threshold used here (averaging reversals) is the most common, alternatives (such as the mid-point of the most probable PF) may be more efficient and may use any additional information obtained with $Z_{3,1}$ more effectively. [The information obtained with either rule may also be enhanced by setting $k=1$ during initial trials (Levitt, 1971).] It is also possible that, in not requiring consecutive correct responses, efficiency associated with $Z_{3,1}$ may be less sensitive to greater instability in the PF than was apparent in our (trained) listeners.

The computer simulation, which excluded effects of inattention and other complexities of human performance, indicated that $L_{3,1}$ was associated with inherent downward bias, as reported previously (e.g., Saberi and Green, 1996; Leek, 2001), whereas $Z_{3,1}$ had negligible inherent bias. The reasons for this difference are unclear. It is probably unrelated to the different asymptotic probabilities; based on Saberi and Green's analysis (1996), one may have expected inherent bias to be *greater* with $Z_{3,1}$ compared to $L_{3,1}$, rather than less. Nevertheless, more extensive simulations of bias by García-Pérez (2002) indicate that $Z_{3,1}$ is associated with substantial inherent bias under some conditions. The relative bias identified in the human experiment was almost identical to that in the simulations and so probably reflected the inherent bias with $L_{3,1}$. This similarity in relative biases also indicates that $Z_{3,1}$ is no more, or less, susceptible to any additional bias that may occur with trained humans than $L_{3,1}$. However, the possibility that both $L_{3,1}$ and $Z_{3,1}$ yielded biased thresholds with humans cannot be excluded as PFs were not measured and the simulation may not have captured important aspects of human performance. Interestingly, Schlauch and Carney (2003) reported that $Z_{3,1}$ produces less biased thresholds than $L_{3,1}$ with simulated inattention. The Zwislocki rule may also be less susceptible to bias arising from anticipation of the rule, which would presumably be more difficult with $Z_{3,1}$.

V. CONCLUSIONS

(1) The Zwislocki 3-down, 1-up rule was not found to enhance test-retest variability over the corresponding Levitt rule despite the accumulation of additional reversals, at least under the conditions used here.
(2) In computer simulations based on current data, the Zwislocki rule yielded ITD thresholds that were less biased compared to the Levitt rule, the reasons for which were unclear.
(3) On current evidence, there are minimal grounds for recommending the Zwislocki over the corresponding Levitt rule, except where it is desirable to target a response probability of 0.75.

ACKNOWLEDGMENTS

D.R. was supported by a grant from the Royal National Institute for Deaf People. The authors thank Mark E. Lutman and two anonymous reviewers for constructive comments.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using transposed stimuli," J. Acoust. Soc. Am. **112**, 1026–1036.

Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). "Inter-individual differences in binaural detection of low-frequency or high-frequency tonal signals masked by narrow-band or broadband noise," J. Acoust. Soc. Am. **103**, 2069–2078.

García-Pérez, M. A. (**2002**). "Properties of some variants of adaptive staircases with fixed step sizes," Spatial Vis. **15**, 303–321.

Hafter, E. R., and Trahiotis, C. (**1997**). "Functions of the binaural system," in *Encyclopedia of Acoustics, Volume 3*, edited by M. J. Crocker (Wiley, New York), pp. 293–302.

Koehnke, J., Culotta, C. P., Hawley, M. L., and Colburn, H. S. (**1995**). "Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing," Ear Hear. **16**,

331–353.
Leek, M. R. (**2001**). "Adaptive procedures in psychophysical research," Percept. Psychophys. **63**, 1279–1292.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Marvit, P., Florentine, M., and Buus, S. (**2003**). "A comparison of psychophysical procedures for level-discrimination thresholds," J. Acoust. Soc. Am. **113**, 3348–3361.
Saberi, K. (**1995**). "Some considerations on the use of adaptive methods for estimating interaural-delay thresholds," J. Acoust. Soc. Am. **98**, 1803–1806.
Saberi, K., and Green, D. M. (**1996**). "Adaptive psychophysical procedures and imbalance in the psychometric function," J. Acoust. Soc. Am. **100**, 528–536.
Saberi, K., and Green, D. M. (**1997**). "Evaluation of maximum-likelihood estimators in nonintensive auditory psychophysics," Percept. Psychophys. **59**, 867–876.
Schlauch, R. S., and Carney, E. J. (**2003**). "A comparison of Levitt and Zwislocki decision rules for use with forced-choice adaptive procedures," J. Acoust. Soc. Am. **113**, 2250.
Zwislocki, J., Maire, F., Feldman, A. S., and Rubin, H. (**1958**). "On the effect of practice and motivation on the threshold of audibility," J. Acoust. Soc. Am. **30**, 254–262.
Zwislocki, J. J., and Relkin, E. M. (**2001**). "On a psychophysical transformed-rule up and down method converging on a 75% level of correct responses," Proc. Natl. Acad. Sci. U.S.A. **98**, 4811–4814.

No adaptation in the amplitude modulation domain in trained listeners (L)

Laetitia Bruckert, Marion Herrmann, and Christian Lorenzi[a)]
Laboratoire de Psychologie de la Perception- FRE CNRS 2929, Institut de Psychologie, Université René Descartes - Paris 5, 71 Av. Edouard Vaillant, 92774 Boulogne-Billancourt, France

(Received 9 March 2005; revised 1 April 2006; accepted 6 April 2006)

The present study shows that on average, exposure to a 15 min, 5 kHz tone modulated sinusoidally in amplitude at 16 Hz with a 100% depth does not affect significantly amplitude modulation (AM) detection thresholds measured between 4 and 64 Hz when listeners are extensively trained to the AM detection task, with and without adaptor before data collection. These results are compatible with previous work given that a clear 6-dB adaptation effect was observed during the first pilot trials. However, the results reveal that adaptation effects are not robust, and suggest that the mechanisms underlying adaptation to AM must be reevaluated. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200696]

PACS number(s): 43.66.Mk, 43.66.Ed [JHG] Pages: 3542–3545

I. INTRODUCTION

Neurophysiological investigations of the auditory pathways in humans have demonstrated the existence of neural populations selectively responsive to amplitude modulated (AM) signals for AM rates lower than 100–200 Hz (e.g., Liégeois-Chauvel *et al.*, 2004). Psychophysical studies have also provided evidence for the existence of AM channels in humans, using either adaptation (e.g., Kay and Matthews, 1972; Tansley and Suffield, 1983; Richards *et al.*, 1997; Wojtczak and Viemeister, 2003) or modulation masking paradigms (e.g., Bacon and Grantham, 1989).

In typical psychoacoustical adaptation procedures, listeners are exposed to a tone modulated sinusoidally in amplitude at a given rate for a prolonged time (a stimulus called an "adaptor") and then tested with a test tone modulated sinusoidally in amplitude at different rates, including the adaptor AM rate. The adaptor is generally modulated at a suprathreshold level [e.g., ten times the pre-exposure detection threshold (Tansley and Suffield, 1983) or 100% (e.g., Richards *et al.*, 1997; Wojtczak and Viemeister, 2003)], and exposure to the adaptor typically lasts between 12 s (Kay and Matthews, 1972) and 30 min (Tansley and Suffield, 1983). Overall, the studies cited above show that AM detection thresholds are selectively elevated above pre-exposure thresholds when the carrier frequency and the AM rate of the adaptor are similar to those of the test tone, and when exposure time is between 10 and 30 min. When the adaptor and test AM rates are equal to 16 Hz, the selective adaptation effect reaches a maximum of 7–9 dB. Lower amounts of adaptation are found at other AM rates, when the carrier frequency and the AM rate of the adaptor are different from those of the test tone, and when exposure time is lower than 10 min.

The selective temporary threshold elevation reported in adaptation studies is generally interpreted as resulting from fatigued neural units that are selectively tuned to the characteristics of the adapting stimulus (i.e., AM rate). Similar adaptation data were reported by Kay and Matthews (1972) for frequency modulation (FM), leading to the notion that the auditory system is also equipped with channels selectively tuned to FM. However, two psychophysical studies (Moody *et al.*, 1984; Wakefield and Viemeister, 1984) have cast serious doubts on this interpretation in terms of fatigued, modulation channels. Wakefield and Viemeister (1984) investigated adaptation to FM using a forced-choice procedure, in contrast with most adaptation experiments which are based on subjective procedures [except Richards *et al.* (1997) and Wojtczak and Viemeister (2003)]. These authors observed that threshold elevation obtained after exposure to the adaptor corresponded only to a modest change in detection performance, and argued that this change could be explained in terms of nonsensory factors, such as a change in detection criterion or the use of the adaptor rather than a comparison as a reference for signal detection. Moody *et al.* (1984) investigated adaptation to upward frequency sweeps using a subjective procedure. In each of the four listeners that participated in their study, the authors observed that large differences in threshold between adapted and nonadapted conditions decreased with continued testing, and that the final threshold differences were much smaller than those reported in previous studies. Moreover, difference between pre- and postexposure thresholds disappeared in three out of the four listeners after having completed six measurement sessions. Thus, the effect that was initially interpreted as selective adaptation and was considered as a strong support for the existence of FM channels disappeared with training.

A recent adaptation study conducted by Wojtczak and Viemeister (2003) with AM stimuli has suggested that the conclusion drawn by Wakefield and Viemeister (1984) in the FM domain does not apply to the AM domain. This study showed that a large effect of the AM adaptor on AM detection performance can be obtained with a forced choice procedure. This led the authors to conclude that such an effect cannot result from the use of an inappropriate reference.

[a)]Author to whom correspondence should be addressed. Electronic mail: lorenzi@psycho.univ-paris5.fr

 0001-4966/2006/119(6)/3542/4/$22.50

TABLE I. Stimulus parameters, AM detection thresholds, and adaptation effects for the five successive practice sessions. Standard deviations across listeners are given in parentheses for AM detection thresholds obtained before and after exposure to the adaptor. The adaptor and target AM rate was fixed at 8 Hz.

	Adaptor duration		Target and standard duration (s)	Threshold before exposure (dB)	Threshold after exposure (dB)	Adaptation effect (dB)
	Initial adaptor (min)	Between-trial adaptor (s)				
Session 1	5	3	1	−24.1 (2)	−17.8 (2.2)	6.3
Session 2	15	6	1	−23.9 (2.9)	−19.1 (3.9)	4.8
Session 3	15	6	0.5	−21.8 (2.4)	−19.3 (3.1)	2.5
Session 4	15	6	0.5	−21.7 (3.2)	−19.2 (3.8)	2.6
Session 5	15	6	0.5	−21.9 (2.6)	−19.4 (3.5)	2.5

Moreover, in contrast with adaptation studies performed with FM stimuli, the various studies that investigated adaptation to AM did not report any decrease in the difference between thresholds obtained in adapted and nonadapted conditions with continued testing. However, during the pilot trials of an experiment designed to replicate part of the data of Wojtczak and Viemeister (2003), we observed that the threshold differences between nonadapted and adapted conditions changed in an orderly manner as a function of repeated sessions. Listeners were therefore trained extensively to AM detection in various adapted and nonadapted conditions during about 12 h before running this replication of the Wojtczak and Viemeister (2003) experiment. The present study reports the outcome of this adaptation experiment in the AM domain conducted with six extensively trained listeners.

II. EXPERIMENTS

A. Method

1. Listeners

Six listeners (referred to as Observers 1 to 6) participated in these experiments. All listeners had audiometric thresholds less than 20 dB hearing level between 0.25 and 8 kHz, and no history of hearing difficulty. Their ages ranged from 20 to 26 years. Listeners were given 10–12 h of practice (corresponding to five successive sessions of pilot trials) in nonadapted and adapted conditions using a large range of parameters for target, standard, and adaptor signals prior to data collection (see Table I). In these pilot trials, the adaptor and target AM rate was fixed at 8 Hz. Stimulus (target and standard) duration was fixed at 1 or 0.5 s. Adaptor duration was fixed at 5 or 15 min, and between-trial adaptor duration was fixed at 3 or 6 s. The apparatus, experimental setup, detection procedure, and stimulus parameters were otherwise identical to those used in the so-called "final experiment" described below.

2. Stimuli and procedure

All experiments were controlled by a personal computer (PC). Stimuli were generated using a 16-bit digital/analog converter at a sampling frequency of 44.1 kHz. They were attenuated to the desired level by a Yamaha analog amplifier and were delivered diotically via Sennheiser HD 580 earphones. The carrier was a 5-kHz sinusoid presented at 60 dB sound pressure level (SPL) within a notched-noise masker. The masker was obtained by adding a low-pass noise (cutoff frequency =2917 Hz; rolloff =90 dB/oct.) and a high-pass noise (cutoff frequency =7500 Hz; rolloff =90 dB/oct.) presented at 20 dB and 7 dB below the overall carrier level, respectively. Such a noise masker was used to restrict the possibility of off-frequency listening, without appreciably affecting the detection of the carrier. This masker therefore forced listeners to attend to the same auditory filter in the adapted and nonadapted conditions.

The adaptor was a 60 dB SPL, 5 kHz sinusoid that was modulated sinusoidally in amplitude with a 100% depth. In the final experiment, the adaptor was modulated at a fixed modulation rate of 16 Hz (in contrast with pilot trials where the adaptor was modulated at 8 Hz). The duration of the adaptor is specified below. The stimuli (target and standards) used in the detection task were 60 dB SPL, 5 kHz sinusoids that were either modulated in amplitude or were unmodulated. In the final experiment, stimulus duration (target and standards) was 500 ms including 500 ms rise/fall times shaped using a raised-cosine function. The adaptor and stimuli (target and standards) were presented within the notched-noise masker described above. All listeners were tested in a soundproof booth.

Listeners had to detect sinusoidal AM applied to the 5 kHz, sinusoidal carrier. In a typical experimental session of the final experiment corresponding to the "adapted condition," the adaptor was played for 15 min before the testing began. A warning signal was displayed on the PC screen 30 s before the first trial was presented. After each trial, the adaptor was reintroduced for 6 s, and after the first 25 trials, testing was interrupted and the adaptor was played for 30 s to reinforce possible adaptation to AM. On each trial, two standard stimuli and a target stimulus were presented in random order with a silent interstimulus interval of 500 ms. The standards were unmodulated 5 kHz sinusoids. The target was a 5 kHz sinusoid which was modulated sinusoidally in amplitude at a given rate f_m where f_m=4, 8, 16, 32, and 64 Hz and at a given modulation depth m ($0 \leqslant m \leqslant 1$). The root-mean-square power was identical in each interval.

AM detection thresholds were obtained using an adaptive three-interval forced-choice (3IFC) procedure with a two-down one-up stepping rule that estimates the modulation depth, m, necessary for 70.7% correct detection. Listeners received visual feedback concerning the correct interval after each trial. The step size of m variation corresponded initially to 4 dB (in $20 \log m$); it was reduced to 2 dB after the first two reversals. The geometric mean of the value of m at the

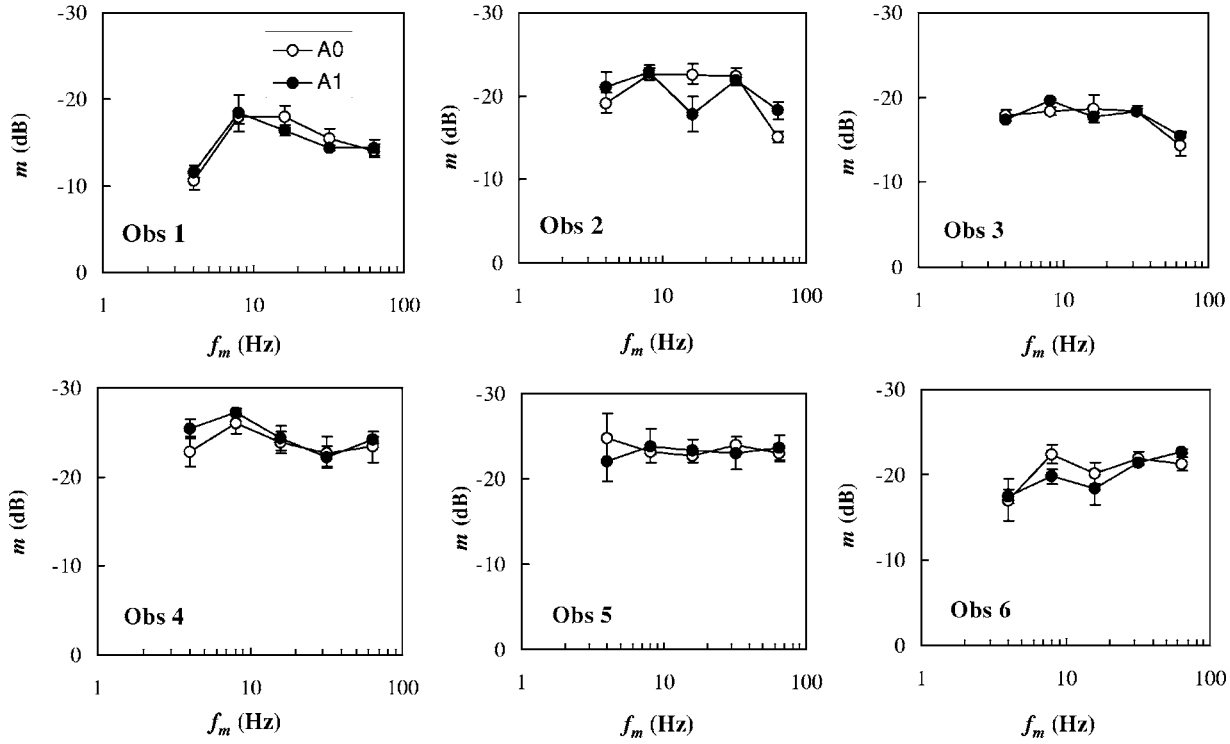


FIG. 1. Individual temporal modulation transfer functions (TMTFs) for the six normal-hearing listeners. Data obtained using a 5 kHz pure tone carrier before (A0: open circles), and after (A1: closed circles) exposure to a 15 min, 16 Hz AM 5 kHz tone. Modulation depth at threshold (m, in dB) is presented as a function of modulation rate f_m. Error bars represent ± one standard deviation from the mean threshold across three repeated measures.

last 10 reversals in a block of 16 reversals was taken as the threshold estimate (in dB) for that block. For each listener and modulation rate, thresholds presented here are based upon three estimates. AM detection thresholds were measured in the "nonadapted condition" using the same procedure and stimulus parameters, except that the adaptor was not played. For each listener, all experimental conditions were run in random order.

B. Results

Mean AM detection thresholds across listeners obtained for adapted and nonadapted conditions during the practice sessions are presented in Table I. The initial AM adaptation effect measured at f_m=8 Hz (target and adaptor) with a 5 min AM adaptor and 1 s targets and standards was relatively strong (6.3 dB) and roughly similar to that reported by Richards *et al.* (1997) or Wojtczak and Viemeister (2003). However, in the following session, the AM adaptation effect measured at f_m=8 Hz with a longer (15 min) adaptor was reduced by 1.5 dB (4.8 dB). This effect decreased even more in the following sessions using shorter targets and standards (500 ms), reaching 2.5 dB.

Individual AM detection thresholds obtained for adapted (A1: filled circles) and nonadapted (A0: open circles) conditions in the final experiment (i.e., after the 10–12 h practice session) are shown in Fig. 1. In each panel, the ordinate indicates the modulation depth at threshold, m, and the abscissa represents the target modulation rate, f_m. The adaptor modulation rate was fixed at 16 Hz in the adapted conditions. In agreement with previous experiments using sinusoidal carriers and 500 ms stimuli (e.g., Viemeister, 1979), AM detection thresholds measured before exposure to the adaptor show an allpass or bandpass characteristic for the range of f_m values used here. In all listeners, thresholds are roughly constant between 8 and 32 Hz and tend to increase at the lowest and highest AM rates tested. On average, AM detection thresholds are about 2–6 dB higher than in earlier studies. This is partly due to slightly poorer thresholds in two listeners (Observers 1 and 3). This is also due to the deleterious effect of the intrinsic random fluctuations of the notched-noise masker (Dau *et al.* 1997), as suggested below (*cf.* Table II). Figure 1 also shows that exposure to the 16 Hz AM adaptor does not affect AM detection thresholds measured between 4 and 64 Hz in most listeners. The mean adaptation effect across listeners is equal to −0.26 dB when calculated across all modulation rates, and 1.53 dB when calculated at f_m=16 Hz only, except for Observer 2 where the detection threshold at f_m=16 Hz is selectively increased by 4.5 dB. An analysis of variance (ANOVA) conducted with factors exposure (two levels: Before/after exposure to the adaptor), modulation rate (five levels), and runs (three levels) showed

TABLE II. Individual AM detection thresholds obtained at f_m=16 Hz with and without notched-noise masker. Standard deviations across repeated measures are given in parentheses.

	Before exposure to adaptor		After exposure to adaptor	
	No masker	Masker	No masker	Masker
Observer 3	−20.2 (1.2)	−18.6 (1.3)	−21.7 (1.2)	−17.6 (0.4)
Observer 4	−24.5 (0.7)	−23.9 (1.2)	−24.1 (1)	−24.2 (1.4)
Observer 5	−25.8 (1)	−22.8 (1.3)	−25.3 (0.9)	−23.2 (0.2)

a significant main effect of modulation rate $[F(4,20)=2.905, p<0.05]$ only. However, the effects of exposure $[F(1,5)=0.015, p=0.908]$, runs $[F(2,10)=1.551, p=0.25]$, and the interaction between exposure and modulation rate $[F(4,20)=2.501, p=0.075]$ and between exposure and runs $[F(2,10)=3.749, p=0.061]$ were not significant.

The absence of clear adaptation effects contrasts with the 7–8 dB adaptation effect found by Richards *et al.* (1997) and Wojtczak and Viemeister (2003) using a shorter adaptor (10 min). These authors did not attempt to prevent off-frequency listening by using a notched-noise masker. The aim of this second experiment was to clarify the contribution of off-frequency listening to adaptation. This was tested by measuring AM detection thresholds in "adapted" and "non-adapted" conditions in the absence of the notched-noise masker. For each listener, thresholds were obtained at $f_m = 16$ Hz only, using an identical 3IFC adaptive procedure with feedback to that used to determine AM detection in the first experiment. The three listeners (Observers 3, 4, and 5) that participated in this control experiment were those showing no adaptation effect in the previous experiment. They were not given any additional practice prior to data collection. Individual data are presented in Table II. Removing the masker yields a quasi systematic decrease in AM detection thresholds, ranging from 0.5 to 4 dB in both the adapted and nonadapted conditions. However, in each listener, similar but negligible adaptation effects are obtained with and without the noise masker (adaptation effects differing inconsistently by 0.5–3 dB), suggesting that off-frequency listening is not involved in AM adaptation.

III. SUMMARY AND CONCLUSIONS

The present results indicate that on average, exposure to a 15 min, 16 Hz AM adaptor does not affect consistently and substantially AM detection thresholds measured between 4 and 64 Hz when listeners are extensively trained to the detection task (with and without adaptor) before data collection. These results are not incompatible with previous work given that a clear, 6 dB adaptation effect was observed during the first pilot trials. However, the results reported here limit the generality of previous findings and emphasize the need to train listeners extensively. Indeed, it is noteworthy that neither Richards *et al.* (1997) nor Wojtczak and Viemeister (2003) provided more than 5 h of training to their listeners (compared to 10–12 h in the present study). Moreover, only two out of four listeners were trained in the study conducted by Richards *et al.* (1997) and among these two listeners, one showed adaptation and the other did not. It is important to acknowledge that in the present study, practice effects were confounded with changes in stimulus conditions (for instance, changes in adaptor and stimulus duration; see Table I). However, as demonstrated by Tansley and Suffield (1983), increasing exposure to the adaptor (and also, decreasing stimulus duration) should yield increased estimates of adaptation instead of decreased estimates, as shown in the present study. This strongly suggests that practice per se is responsible for the absence of adaptation effects reported in the final experiment. Nevertheless, future work assessing systematically practice effects over time is warranted. These results also cast doubts on the notion of "hard wired" AM channels, and on an interpretation of AM adaptation effects solely in terms of fatigued AM channels. It is now crucial to clarify the central mechanisms underlying adaptation to AM such as the use of the adaptor rather than a comparison as a reference for signal detection, as suggested initially by Wakefield and Viemeister (1984).

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking patterns: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I: Modulation detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Kay, R. H., and Matthews, D. R. (**1972**). "On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones," J. Physiol. (London) **225**, 657–677.

Liégeois-Chauvel, C., Lorenzi, C., Trébuchon, A., Régis, J., and Chauvel, P. (**2004**). "Temporal envelope processing in the human left and right auditory cortices," Cereb. Cortex **14**, 731–740.

Moody, D. B., Cole, D., Davidson, L. M., and Stebbins, W. C. (**1984**). "Evidence for a reappraisal of the psychophysical selective adaptation paradigm," J. Acoust. Soc. Am. **76**, 1076–1079.

Richards, V. M., Buss, E., and Tian, L. (**1997**). "Effects of modulator phase for comodulation masking release and modulation detection interference," J. Acoust. Soc. Am. **102**, 468–476.

Tansley, B. W., and Suffield, J. B. (**1983**). "Time-course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation," J. Acoust. Soc. Am. **74**, 765–775.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Wakefield, G. H., and Viemeister, N. F. (**1984**). "Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?," J. Acoust. Soc. Am. **75**, 1588–1592.

Wojtczak, M., and Viemeister, N. F. (**2003**). "Suprathreshold effects of adaptation produced by amplitude modulation," J. Acoust. Soc. Am. **114**, 991–997.

On the sound absorption of quadratic residue diffuser groups with various shapes and combinations (L)

Jianbin Yang, Yong Shen,[a] and Han Wang
Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 8 February 2006; revised 29 March 2006; accepted 1 April 2006)

Quadratic residue diffuser (QRD) groups are used to form a large diffusion surface and extra sound absorption is introduced. In this paper we discuss how the installation pattern of QRD groups affects the sound absorption. The absorption coefficients of QRD in groups with different shapes and combinations were measured in a reverberation room. Notable variations in absorption were observed when the shapes of QRD groups change, especially at the low frequency range below the design frequency of QRD. Two peaks of absorption are observed. When a certain shape is chosen, the detailed combinations of QRDs cause little change in absorption. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200047]

PACS number(s): 43.55.Ev, 43.55.Dt [NX] Pages: 3546–3548

I. INTRODUCTION

The quadratic-residue diffuser (QRD) has been widely used[1,2] since Schroeder introduced it in the 1970s.[3,4] It was designed to diffuse sound, but great absorption, especially at a low frequency range, was found and measured by Fujiwara and Miyajima.[5,6] Two methods have been used to explain and predict the absorption of the QRD unit. One is based on the constant sound pressure on the diffuser plane and the same average surface admittance by Kuttruff.[7] In Wu's paper, a Fourier wave decomposition model was discussed and applied to a single QRD.[8,9] The prediction shows good agreement with measurement results in the impedance tube with normal incidence. The previous theory studied the one-dimensional QRD unit, assuming the walls between wells were infinitely parallel, which is limited in applications. In practical situations, QRDs are used in groups to form large diffusion surfaces.[2] Different installation patterns of QRD groups make the sound absorption not quite the same as one single QRD and introduce extra absorption. It is important to study their sound absorptions with random incidence in a reverberation room.

In this paper, absorption coefficients of QRD in groups with various shapes and combinations were studied experimentally. Measurements were carried out in a reverberation room of *Nanjing University*. Seven installation patterns of QRD groups were applied to make the comparison more comprehensive. The change of absorption at a low frequency range is investigated, because this feature is quite important for room design.

II. TEST SPECIMEN AND PROCEDURE

The expression of the depth sequence l_n in one period of a one-dimensional QRD is defined as[3]

$$l_n = \frac{c \bmod(n^2, N)}{N(2f_r)}, \quad n = 0, 1, \ldots, N-1,$$

where N is the prime number, f_r is the design frequency, and c is speed of sound in air.

Parameters of QRD samples used in measurement are as follows: size about 66 cm × 66 cm, prime number $N=11$, design frequency $f_r \approx 390$ Hz, the maximum length of wells 36 cm, well width 5 cm, and separation wall thickness 0.9 cm. The length sequence is (in cm)

0 4 16 36 20 12 12 20 36 16 4.

The QRD sample is well constructed and the hard boards blocking off the wells would not vibrate in measurements.

In the experiment, the *Acoustics measurement of sound absorption in a reverberation room* (*ISO/FDIS 354:2003*) was applied. Reverberation times of the empty field and after the QRDs have been introduced were measured. The absorption coefficient of QRD could be calculated. The reverberation room of *Nanjing University* has a size about 7.35 m × 5.9 m × 5.1 m, volume about 221 m^3, and has two curved surface diffusers suspended, each composed of nine sheets and approximating to semiconical shape. They are centrosymmetrically installed.

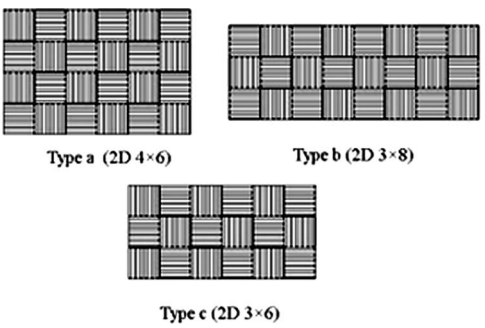


FIG. 1. Three kinds of QRD groups with 2-D installation in measurement.

[a]Electronic mail: yshen@nju.edu.cn

 0001-4966/2006/119(6)/3546/3/$22.50

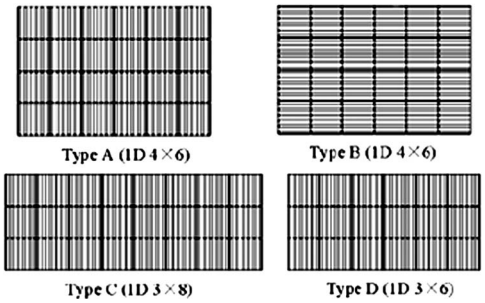


FIG. 2. Four kinds of QRD groups with 1-D installation in measurement.

Norsonic real time analyzer 840 and B&K microphone type 4166 were used in the experiment. QRDs in the previously mentioned above were placed in the center of the room floor. The 24 QRDs' combination had an area of 0.66 m $\times$0.66 m$\times$24=10.45 m^2, and the 18 QRDs' combination had an area of 0.66 m$\times$0.66 m$\times$18=7.84 m^2. To study the change of absorption due to different installation patterns, three two–dimensional (2-D) QRD installations and four one-dimensional (1-D) periodic QRD installations were measured, illustrated in Figs. 1 and 2.

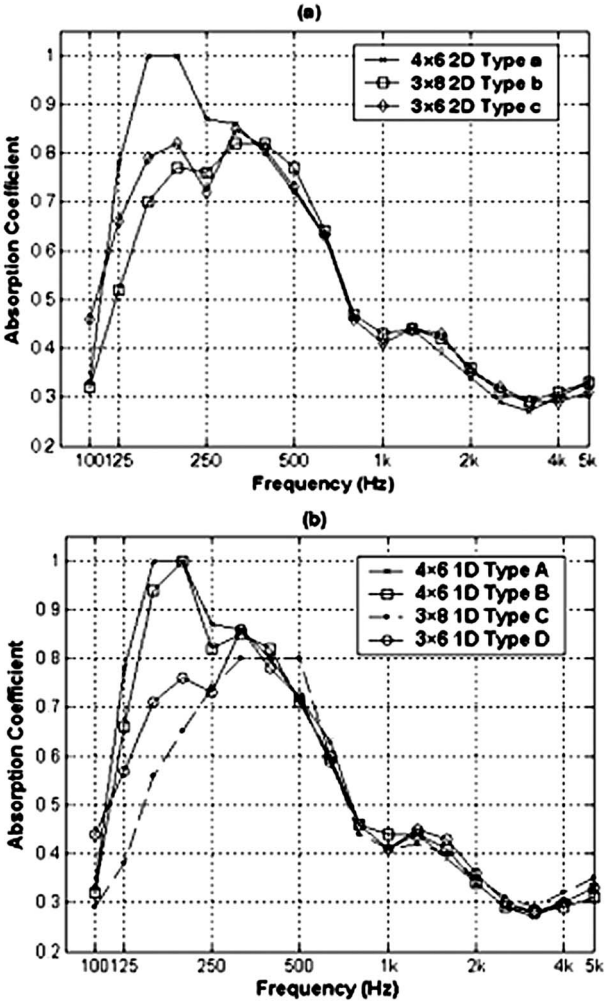


FIG. 3. A comparison of absorption coefficient versus frequency of QRD in groups with the same installations of different shapes: (a) three shapes with 2-D installation; (b) four shapes with 1-D installation.

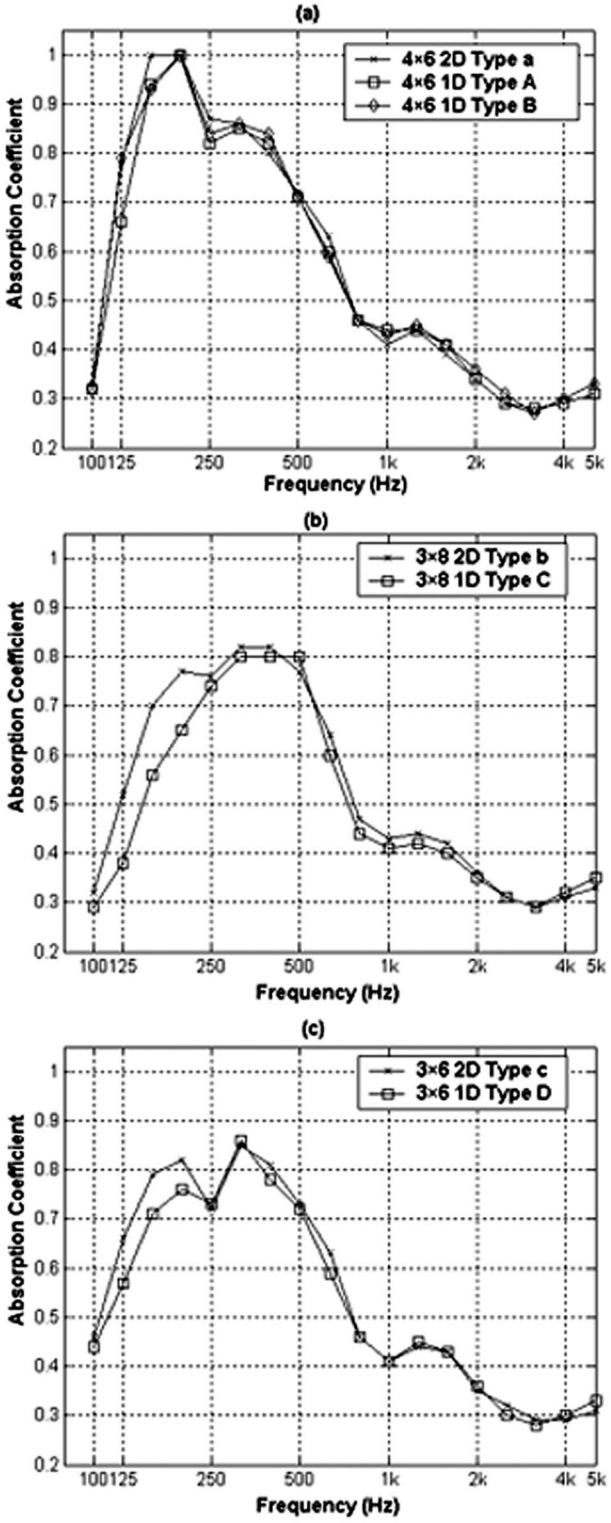


FIG. 4. A comparison of the absorption coefficient versus frequency of QRD in groups of the same shapes with 1-D or 2-D installations: (a) in 4 $\times$6 shape; (b) in 3$\times$8 shape; (c) in 3$\times$6 shape.

III. MEASUREMENT RESULTS

Figures 3 and 4 illustrate the absorption coefficient of QRD in seven groups measured in the reverberation room. Figure 3 shows how the shapes of the QRD groups affect the sound absorption, while Fig. 4 demonstrates the effect on sound absorption due to different installations within the same shapes.

Figure 3 shows that the shape of the QRD groups greatly affects the absorption coefficient. In Fig. 3(a), although all are with 2-D installations, at a frequency range from 125 to 200 Hz, QRD in the 4×6 shape has a larger absorption coefficient than in the 3×8 shape and the 3×6 shape, while the latter two have close values. Comparing the four kinds of 1-D installation in Fig. 3(b), when the length number remains 6 and the width number increases from 3 to 4, absorption coefficients of the QRD change greatly, still between 125 and 200 Hz. While the width number remains 3 and the length number increases from 6 to 8, a little change in absorption could be observed around the first peak. This effect may be caused by two factors: the change of surface admittance distribution and sound diffraction caused by width of the sample. As the wavelength in this frequency range is about 1.7 to 2.7 m, the width of sample with 4×6 shape is 2.64 m, just about this one wavelength range, and would introduce extra sound absorption.

In Fig. 4(a), with the same 4×6 shape, QRD in three groups have almost the same absorption capabilities, which also could be seen in Fig. 4(b) demonstrating the 3×8 shape and in Fig. 4(c) demonstrating the 3×6 shape. It shows that when the shape of the QRD group is chosen, sound absorption is little affected by the detailed distributions of QRDs this shape.

In the absorption coefficient curves, two peaks of absorption coefficient are observed, both in Figs. 3 and 4. The first peak is at frequency of 160 to 200 Hz, whose value could be influenced by the shape of QRDs; the frequency of the second peak at about 400 Hz is around the design frequency of QRD. And there is a small fluctuation around 1200 Hz, triple of the design frequency.

IV. CONCLUSION

The paper has studied how the installation pattern of QRD groups affects the sound absorption coefficient of QRD, especially at a low frequency range below the design frequency of the one-dimensional QRD unit. Seven samples were measured. In the curves of QRD's absorption coefficient versus frequency, two peaks are observed. The first one is relevant to the shape of QRDs, which may introduce extra sound absorption by change of surface impedance distribution and sound diffraction. The second peak is around the design frequency of QRD. The detailed combinations of QRDs inside a chosen shape cause little change in absorption. Large low frequency absorption should be treated carefully when QRD groups are used in a room. Further work is being done to extend the theory to QRD groups.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (10474046).

[1]L. L. Beranek, "Concert hall acoustics," J. Acoust. Soc. Am. **92**, 1–40 (1992).
[2]P. D'Antonio and T. Cox, "Diffuser application in rooms," Appl. Acoust. **60**, 113–142 (2000).
[3]M. R. Schroeder, "Diffuse sound reflection by maximum length sequences," J. Acoust. Soc. Am. **57**, 149–150 (1975).
[4]M. R. Schroeder, "Binaural dissimilarity and optimum ceilings for concert halls: more lateral sound diffusion," J. Acoust. Soc. Am. **65**, 958–963 (1979).
[5]K. Fujiwara and T. Miyajima, "Absorption characteristics of a practically constructed Schroeder diffusor of quadratic-residue type," Appl. Acoust. **35**, 149–152 (1992).
[6]K. Fujiwara, K. Nakai, and H. Torihara, "Visualization of the sound field around a Schroeder diffuser," Appl. Acoust. **60**, 225–235 (2000).
[7]H. Kuttruff, "Sound absorption by pseudo stochastic diffusers (Schroeder diffusers)," Appl. Acoust. **42**, 215–231 (1994).
[8]T. Wu, T. J. Cox, and Y. W. Lam, "From a profiled diffuser to an optimized absorber," J. Acoust. Soc. Am. **108**, 643–650 (2000).
[9]T. Wu, T. J. Cox, and Y. W. Lam, "A profiled structure with improved low frequency absorption," J. Acoust. Soc. Am. **110**, 3064–3070 (2001).

Passive correlation imaging of a buried scatterer (L)

Eric Larose,[a] Oleg I. Lobkis, and Richard L. Weaver
Department of Theoretical and Applied Mechanics, University of Illinois, Urbana, Illinois 61801

(Received 28 November 2005; revised 4 April 2006; accepted 4 April 2006)

Waveforms obtained by correlation of diffuse fields are now widely used to estimate times of flight and ballistic Green's functions. Theory, however, predicts that such passively obtained waveforms should equal the full Green function of a medium, and thereby convey information on the details of a structure and the presence of isolated scatterers. Here we investigate diffuse field correlations in a reverberant elastic body and demonstrate that their shear and Rayleigh waves allow us to image internal and surface features. Resolution in the resulting images is comparable to standard medical and seismic images. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200049]

PACS number(s): 43.40.Hb, 43.40.Qi, 43.35.Yb [JGM] Pages: 3549–3552

The use of diffuse waves for estimating the properties of heterogeneous or complex media has been proposed in several fields of wave physics. Absorption and diffusion are perhaps the most obvious. It is also possible, in principle, to measure modal densities. The coherent backscattering effect was shown to be an accurate way to estimate the transport parameters of optical, ultrasonic, and even seismic waves.[1–3] Dynamic fluctuations in multiply scattered waves have been shown in Diffusing Optical Wave Spectroscopy[4] and Diffusing Acoustic Wave Spectroscopy[5] to provide information on a changing medium, for example, the Brownian motion of scatterers. All these measurements give access to global, or system-averaged, quantities; they provide no details.

Another idea is to use the time correlations of noise for passively recovering the Green's function. This technique gives access to the elastic or acoustic impulse response between two receivers without the need for a controlled source. This was conjectured by Claerbout[6] in the 1960s and confirmed more recently by different numerical works from other geophysicists (see, for instance, Schuster *et al.*[7] and references therein), but found feeble real experimental evidence. Correlations of noise were also developed in helioseismology to reconstruct the time of flight (not the full waveform) of acoustic waves between two points at the sun's surface.[8] This idea was reexplored more recently using an original approach based on wave scattering[9] and perfectly diffuse noise,[10] and was fruitfully applied to real experiments in ultrasound,[9–12] ocean acoustics[13–15] and most especially in seismology[16–19] Experimental works to date[8–13,16–18,20,21] have used the identity between diffuse field correlations and Green's functions to estimate direct propagation and times of flight. The technique has not yet been applied to imaging or locating scatterers (except the double echo seen by Larose *et al.*,[22] but with a perfect distribution of sources, and other numerical works[7]). Here we attempt to address that lack by using a highly reverberant body and a few active distant sources. The resulting high-quality diffuse ultrasonic field allows us to construct good images. We expect the principles to apply also to less perfect diffuse fields, as might be obtained from ambient sound.

Here we present two experiments. In the first of these we image an isolated scatterer with Rayleigh waves (Fig. 1). A 12 mm diameter cylindrical hole was drilled through an aluminum block of dimensions 125 mm × 125 mm × 90 mm. To excite elastic waves in the reverberant body, we employ a Q-switched Nd:YAG laser mounted on a step motor. The laser beam has a Gaussian profile 1.5 mm half-width, an energy of 30 mJ, and a duration of ≈7 ns. Peak power densities were ≈200 MW/cm^2. In this ablation regime the source is mainly equivalent to a normal impulsive surface force. The resulting wave field is recorded by two pin transducers R_1 and R_2 coupled to the aluminum block with a thin layer of oil. The laser triggers the waveform acquisition. For each position of the source S_i and receiver k, the record is noted $\varphi_{ik}(t)=G(S_i,R_k,t)\otimes r^k(t)$, where G is the elastic Green's function, $\otimes$ is convolution, and $r^k(t)$ is the transfer function of the receiver k. Each record is filtered in the 0.05–0.9 MHz frequency band, where the absorption time of the block is about 30 ms. Diffuse field decay is therefore slow enough to permit record lengths greater than 100 ms, which represent thousands of reverberation within the cavity. After each acquisition, the laser is moved to another position. The 1 mm steps are used to mimic a linear array of 60 points. By reciprocity, the sources and receivers can be interchanged. Our experimental setup is therefore analogous to a conventional seismic experiment, where a linear array of 60 geophones would sense the seismic diffuse wavefield generated by two distant sources. The Green's function between any couple of points S_i,S_j of the array is recovered by processing the following time correlation:

$$C_{ij}^k(\tau>0)=\int \varphi_{ik}(t)\varphi_{jk}(t+\tau)dt = G(S_i,R_k,t) \times G(S_j,R_k,t)\otimes r^k(t)\otimes r^k(-t). \quad (1)$$

We only estimate the causal part ($\tau>0$) of the correlation because $C_{ij}(-\tau)=C_{ji}(\tau)$. To remove the receiver functions r^k, we deconvolve the cross-correlations by the averaged autocorrelations $\langle C_{ii}^k(\tau)\rangle_i \approx r^k(t)\otimes r^k(-t)$. This procedure has the additional virtue of removing contamina-

[a]Electronic mail: elarose@uiuc.edu

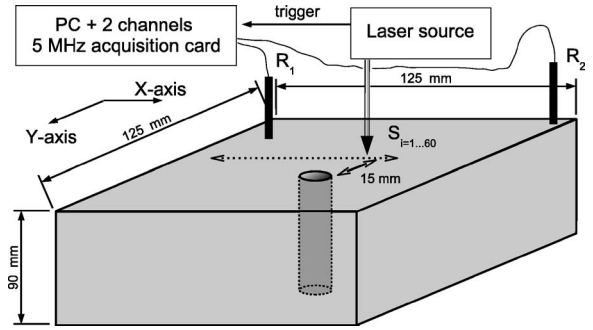


FIG. 1. Experimental setup. A cylindrical hole is passively imaged with Rayleigh waves.

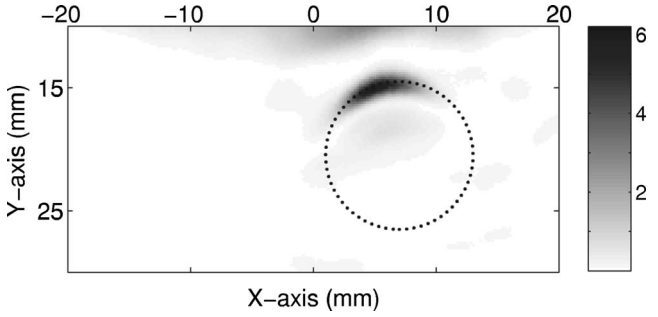


FIG. 3. Reflectivity ϕ (linear scale, arbitrary unit) of the aluminum block probed by passive Rayleigh waves. Black indicates a high reflectivity. The actual position of the cylinder is displayed in a dotted line.

tions $G(R_k,R_k,t)$ or *ghosts*[23,24] related to the environment of the receivers R. Then these correlations are averaged over the available distant sources $R_{k=1,2}$ to get $C_{ij}(\tau)$. As noted by several authors[9,10,14,23–27], this correlation is essentially the Green's function $G_{ij}(\tau)$ and therefore should contain the deterministic signature of the isolated scatterer.

In Fig. 2 we display the time-distance wavefield obtained for all the available autocorrelations $C_{ii}(\tau)$. This autocorrelation is the field sensed in i if i were both source and receiver. Position 0 marks the center of the array. The hyperbolic feature first arriving around $\tau=10$ μs at $i=7$ mm corresponds to the reflection of a Rayleigh wave on the empty cylinder, it is labeled (b). The Rayleigh wave velocity in aluminum is $v_R=2.9$ mm/μs. This corresponds well with the round-trip distance $2d=30$ mm (Fig. 1). (a) and (c) are Rayleigh waves reflected at the lateral sides of the block. Another hyperbolic shaped arrival can be seen in light white before (b). It arrives around $\tau=7.5$ μs and corresponds to a compressional-to-Rayleigh wave converted at the scatterer. Noteworthly, the passive reconstruction of any $G_{ii}(\tau)$ remains imperfect. The averaging used to construct the correlation is finite, leaving visible fluctuations in Fig. 2. To improve the quality of this image, one could increase the record length, or employ additional receivers R_k.[28] Alternatively, we could perform beamforming in order to take advantage of all the $C_{i\neq j}$ cross-correlations. This is a standard procedure to obtain any medical or seismic (migrated) image, except here

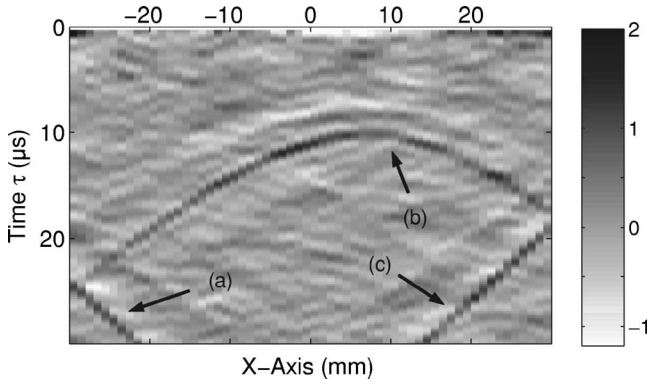


FIG. 2. The time-distance wavefield displayed in a linear scale (arbitrary units). Each record i (x axis) is the autocorrelation $C_{ii}(\tau)$ and is equivalent to the impulse response that would be obtained with a pulse-echo system, (source and receiver at i). (b) Reflection from the hole. (a), (c) Reflections from the sides of the block.

the impulse responses are obtained *passively*.

The image we now process is a 2-D image of the reflectivity of the medium. The first step is to apply beamforming to the forward propagation to focus the wave on any point (x,y) in the medium. This is achieved by summing the time-delayed impulse responses $C_{ij}(\tau)$. The same beamforming technique is then applied to the wave backpropagation (from the focal point to the receivers), the reflectivity ϕ of the medium is then

$$\phi(x,y)=\left(\sum_{ij} C_{ij}(\tau_i+\tau_j)\right)^2, \tag{2}$$

where v_R is the Rayleigh wave velocity and $\tau_i=1/v_R\sqrt{(x-x_i)^2+y^2}$. The 2-D reflectivity map of the medium is displayed in Fig. 3. The array of 60 laser sources is at $y=0$. The top of the hole is clearly visible, as its shape. Because the array has a finite aperture, the sides of the cylinder cannot be imaged; this would be the case with any standard pulse-echo technique. So, the limitations of our passive imaging technique are basically the same as standard imaging technique, except that we only need an array of receivers (resp. sources) and a distant source (resp. receiver).

This image was obtained using Rayleigh waves propagating along the 2-D free surface of the aluminum block. In terms of medical or seismic application, an image obtained using bulk waves throughout a 3-D medium would be more realistic. This problem is addressed in our second experiment. To create an image of a buried scatterer using bulk waves, we scanned the 90 mm$\times$125 mm lateral surface (see Fig 4). The hole is 25 mm beneath the surface. To increase the amplitude of bulk waves relative to Rayleigh waves generated by the laser pulse, a lens is placed in the beam line to expand its size to a diameter of 9 mm. In this configuration the source emits mainly shear waves[29] (see the directivity in Fig. 4). The data acquisition procedure is then identical to that used in the Rayleigh wave experiment. The source was repeated over 60 laser spot positions, and waveforms sampled by the two pin transducers at R_k.

In Fig. 5 we display the time-distance image $C_{ii}(\tau)$ obtained by autocorrelation. As with the first experiment, the hyperbolic feature is the signature of the buried scatterer: the wave labeled (a) is the wave reflected by the top of the cylindrical hole placed at $z=25$ mm and $x=4$ mm. The arrival times along the array correspond to a shear (transverse) wave

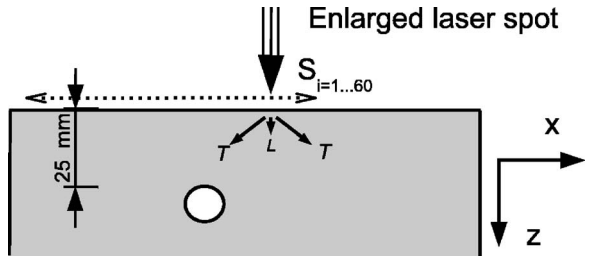


FIG. 4. The second (bulk wave, buried scatterer) imaging experiment. The source scans the surface with 1 mm steps along a line of 60 mm. The directivity (Ref. 29) of one laser shot is shown for shear (transverse T) waves, and weaker compressional (longitudinal L) waves.

(v_T=3.1 mm/μs). (c) is a compressional-to-Rayleigh reflected by the lateral edge of the cavity, and (d) is a Rayleigh-to-Rayleigh reflected by the same edge. The shear wave directivity of the laser generation is clearly visible in the null at (b). Because longitudinal wave generation is much weaker than that of shear waves,[29] its reflection (e) is hardly visible. To obtain a 2-D image of the reflectivity of the medium, we perform the same data processing discussed above in Eq. (2), but now use v_T instead of v_R. This image is plotted in Fig. 6. The top of the reflector is clearly visible. Because of the finite size of the linear array, the sides and bottom of the cylindrical hole cannot be imaged. The shear wave directivity of the laser spot induces a preferential reflection for oblique incidences; additionally, the images in Figs. 5 and 6 show a null at the epicenter. Speckle fluctuations are noticeable, making this figure a little more noisy than the ones obtained with Rayleigh waves. This is expected since the field at the free surface is dominated by Rayleigh waves.

To conclude, we have shown here the feasibility of imaging small details of the medium (like a buried isolated scatterer) by means of the passive time-correlation technique. We have conducted two experiments in an aluminum block in which the elastic excitation is provided by a series of laser shots. A few distant receivers sensed the wavefield. By reciprocity, this experimental setup is analogous to an array of seismic geophones sensing the diffuse wavefield originating from distant sources. We therefore believe this technique could be transposed and applied to prospection geophysics,

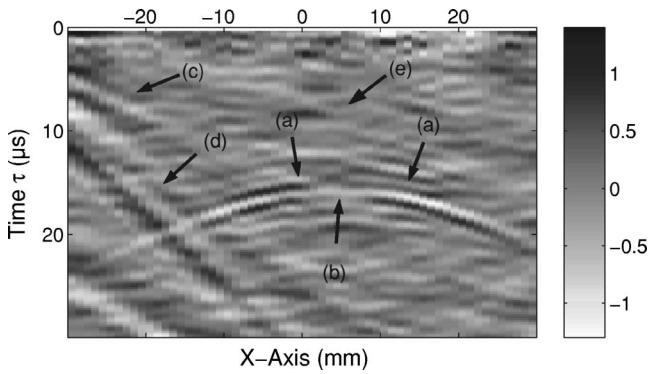


FIG. 5. The time-distance wavefield (linear scale, arbitrary units). Each autocorrelation $C_{ii}(\tau)$ is plotted for different position i along the X axis, and time τ. Position 0 marks the center of the array; (a)–(e) are different reflections (see the text).

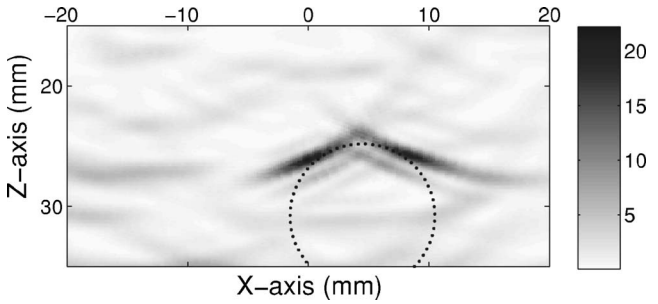


FIG. 6. Reflectivity (linear scale, arbitrary units) of the aluminum block as probed by bulk shear waves. The array of 60 laser sources is at y=0. Black indicates a high reflectivity. The top of the hole is clearly visible. The actual position of the cylinder is displayed in the dotted line.

as to medical imaging. In the first experiment we passively obtained an image of a scatterer with Rayleigh waves. In the second, we used bulk (shear) waves to passively localize a buried scatterer. The use of a fully developed diffuse field in a closed cavity is not a rigorous requirement for this imaging technique. It could, in principle, be replaced by any other diffuse field, like diffuse waves in an open medium, or ambient noise. The application to ambient seismic noise is particularly promising for improving images of the deep Earth or performing passive images of the near surface.

ACKNOWLEDGMENTS

This work was funded by the National Science Foundation Grant No. EAR 0543328 and by the University of Illinois Critical Research Initiatives. The authors wish to thank A. Derode, M. Campillo, and B. van Tiggelen for a fruitful discussion and support. E. Larose acknowledges a Lavoisier grant from the French Ministry of Foreign Affairs.

[1] M. P. Van Albada and A. Lagendijk, "Observation of weak localization of light in a random medium," Phys. Rev. Lett. **55**, 2692 (1985).
[2] P.-E. Wolf and G. Maret, "Weak localization and coherent backscattering of photons in disordered media," Phys. Rev. Lett. **55**, 2696 (1985).
[3] A. Tourin, A. Derode, P. Roux, B. A. van Tiggelen, and M. Fink, "Time-dependent coherent backscattering of acoustic waves," Phys. Rev. Lett. **79**, 3637 (1997).
[4] D. J. Pine, D. A. Weitz, P. M. Chaikin, and E. Herbolzheimer, "Diffusing-wave spectroscopy," Phys. Rev. Lett. **60**, 1134 (1988).
[5] M. L. Cowan, I. P. Jones, J. H. Page, and D. A. Weitz, "Diffusing acoustic wave spectroscopy," Phys. Rev. E **65**, 066605 (2002).
[6] J. F. Claerbout, "Synthesis of a layered medium from its acoustics transmission response," Geophysics **33**, 264–269 (1968).
[7] G. T. Schuster, J. Yu, J. Sheng, and J. Rickett, "Interferometric/daylight seismic imaging," Geophys. J. Int. **157**, 838–852 (2004).
[8] T. L. Duvall, S. M. Jefferies, J. W. Harvey, and M. A. Pomerantz, "Time-distance helioseismology," Nature **362**, 430–432 (1993).
[9] O. I. Lobkis and R. L. Weaver, "On the emergence of the Green's function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[10] R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[11] E. Larose, A. Derode, M. Campillo, and M. Fink, "Imaging from one-bit correlations of wideband diffuse wave fields," J. Appl. Phys. **95**, 8393–8399 (2004).
[12] A. E. Malcolm, J. A. Scales, and B. A. van Tiggelen, "Extracting the Green function from diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601(R) (2004).
[13] P. Roux, W. Kuperman, and the NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).

[14]K. G. Sabra, P. Roux, and W. A. Kuperman, "Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide," J. Acoust. Soc. Am. **117**, 164–174 (2005).
[15]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[16]N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High resolution surface wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[17]P. Roux, K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, "P-waves from cross-correlation of seismic noise," Geophys. Res. Lett. **32**, L19303 (2005).
[18]E. Larose, A. Khan, Y. Nakamura, and M. Campillo, "Lunar subsurface investigated from correlation of seismic noise," Geophys. Res. Lett. **32**, L16201 (2005).
[19]K. G. Sabra, P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, "Surface wave tomography from microseisms in southern California," Geophys. Res. Lett. **32**, L14311 (2005).
[20]M. Campillo and A. Paul, "Long range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[21]A. Paul, M. Campillo, L. Margerin, E. Larose, and A. Derode, "Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves," J. Geophys. Res. **110**, B08302 (2005).
[22]E. Larose, G. Montaldo, A. Derode, and M. Campillo, "Passive imaging of localized reflectors and interfaces in open media," Appl. Phys. Lett. **88**, 104103 (2006).
[23]A. Derode, E. Larose, M. Tanter, J. de Rosny, A. Tourin, M. Campillo, and M. Fink, "Recovering the Green's function from field-field correlations in an open scattering medium," J. Acoust. Soc. Am. **113**, 2973–2976 (2003).
[24]R. L. Weaver and O. I. Lobkis, "Diffuse fields in ultrasonics and seismology," Geophysics (in press).
[25]R. L. Weaver, "Information from seismic noise," Science **307**, 1568–1569 (2005).
[26]R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[27]B. A. van Tiggelen, "Green function retrieval and time reversal in a disordered world," Phys. Rev. Lett. **91**, 243904 (2003).
[28]R. L. Weaver and O. I. Lobkis, "The mean and variance of diffuse field correlations in finite bodies," J. Acoust. Soc. Am. **118**, 3447–3456 (2005).
[29]W. P. Mason and R. N. Thurston, *Physical Acoustics* (Academic, New York, 1988) Vol. **18**.

Expressions for direct evaluation of wave number in cylindrical shell vibration studies using the Flügge equations of motion

Denis G. Karczub
SVT Engineering Consultants, 433 Vincent Street West, Leederville, Western Australia, 6007, Australia

(Received 8 July 2006; revised 13 March 2006; accepted 13 March 2006)

Algebraic expressions for direct evaluation of structural wave number in cylindrical shells are presented for vibration described by the Flügge equations of motion. Applications of the expressions presented here include the evaluation of nondimensional correlation ratios for the estimation of dynamic stress from vibration velocity, and the development of more accurate transmission loss calculations. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2193814]

PACS number(s): 43.20.Bi, 43.20.Mv [TDM] Pages: 3553–3557

I. INTRODUCTION

Wave number calculations for cylindrical shell vibration are of importance in calculations of sound transmission and radiation from industrial piping and ducting (Fagerlund and Chou[1]), and the estimation of dynamic stress due to acoustically induced vibration (Karczub and Norton[2]). Transmission loss studies performed in the 1970s, such as those of Fagerlund and Chou,[1] typically used simplified formulations to evaluate wave number due to limited computing capabilities at that time, and the extra complexity of incorporating iterative numerical techniques (Forsberg[3]) to solve for wave number when using the more accurate Flügge equations of motion. In the more recent work of Karczub and Norton[2] on the relationships between dynamic stress and vibration velocity for cylindrical shell vibration, exact algebraic expressions for direct evaluation of structural wave number based on the Flügge equations of motion were utilized. As these expressions have not been published previously, and are of potential use in the development of fast and accurate transmission loss predictions for industrial pipes and ducts, they are presented here for reference along with a basic outline of their derivation. These algebraic expressions for wave number have the advantage of being accurate, robust, fast and simple to program in comparison with other alternatives.

The expressions presented here for wave number were derived directly from the Flügge equations of motion. As the wave numbers are the roots of an eighth order polynomial, the symbolic algebra package *Mathematica* was used to derive algebraic expressions for these roots and to simplify the algebraic expressions via a series of substitutions. The expressions presented here are valid for circumferential modes $n \geqslant 1$. The interested reader is referred to Leissa[4] for detailed information on the Flügge equations of motion and their derivation.

II. THEORY

A. Equations of motion

The coordinate system used for describing the vibrational response of a thin cylindrical shell is defined in Fig. 1. The orthogonal displacement coordinates u, v, and w describe the vibrational motion of the shell wall midsurface at axial position x and circumferential position θ. The coordinates u and v represent in-plane axial and circumferential displacements of the shell wall midsurface, respectively, and w represents the out-of-plane transverse displacement of the shell wall. Using the coordinate system defined in Fig. 1, the Flügge equations of motion for the free vibration of circular cylindrical shells are (Fuller[5])

$$\left[a^2\frac{\partial^2}{\partial x^2}+\frac{1-\mu}{2}\frac{\partial^2}{\partial\theta^2}-\rho a^2\frac{1-\mu^2}{E}\frac{\partial^2}{\partial t^2}+\beta^2\frac{1-\mu}{2}\frac{\partial^2}{\partial\theta^2}\right]\mathbf{u}+\left[a\frac{1+\mu}{2}\frac{\partial^2}{\partial x\partial\theta}\right]\mathbf{v}+\left[a\mu\frac{\partial}{\partial x}-\beta^2a^3\frac{\partial^3}{\partial x^3}+\beta^2a\frac{1-\mu}{2}\frac{\partial^3}{\partial x\partial\theta^2}\right]\mathbf{w}=0, \tag{1a}$$

$$\left[a\frac{1+\mu}{2}\frac{\partial^2}{\partial x\partial\theta}\right]\mathbf{u}+\left[a^2\frac{1-\mu}{2}\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial\theta^2}-\rho a^2\frac{1-\mu^2}{E}\frac{\partial^2}{\partial t^2}+3\beta^2a^2\frac{1-\mu}{2}\frac{\partial^2}{\partial x^2}\right]\mathbf{v}+\left[\left(\frac{\partial}{\partial\theta}-\beta^2a^2\frac{3-\mu}{2}\frac{\partial^3}{\partial x^2\partial\theta}\right)\right]\mathbf{w}=0, \tag{1b}$$

and

$$\left[a\mu\frac{\partial}{\partial x}-\beta^2a^3\frac{\partial^3}{\partial x^3}+\beta^2a\frac{1-\mu}{2}\frac{\partial^3}{\partial x\partial\theta^2}\right]\mathbf{u}+\left[\frac{\partial}{\partial\theta}-\beta^2a^2\frac{3-\mu}{2}\frac{\partial^3}{\partial x^2\partial\theta}\right]\mathbf{v}+\left[1+\beta^2\left(a^4\frac{\partial^4}{\partial x^4}+2a^2\frac{\partial^4}{\partial x^2\partial\theta^2}+\frac{\partial^4}{\partial\theta^4}\right)+\rho a^2\frac{1-\mu^2}{E}\frac{\partial^2}{\partial t^2}+\beta^2\left(1+2\frac{\partial^2}{\partial\theta^2}\right)\right]\mathbf{w}=0, \tag{1c}$$

where $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are the complex displacements, a is the shell mean radius, μ is Poisson's ratio, E is the modulus of elasticity, ρ is density, h is the shell wall thickness, and β is the nondimensional thickness parameter defined by

$$\beta=\frac{h}{a\sqrt{12}}. \tag{2}$$

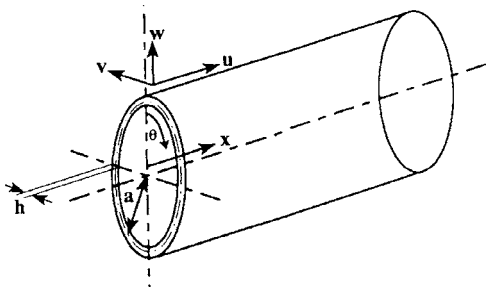


FIG. 1. Cylindrical shell coordinate system.

B. Wave solution

Cylindrical shell wave propagation can be expressed in terms of two orthogonal wave components, one in the circumferential direction and the other in the axial direction. Only propagating waves with wave number $k_c = n/a$ occur in the circumferential direction resulting in circumferential mode shapes of the form $\cos(n\theta)$, where n is the integral circumferential mode number and θ is the shell angular position. The circumferential modes are independent of each other and can be analyzed separately. In the axial direction, for a given circumferential mode n, the axial wave number is given by k_{ns} and these waves have the form $e^{k_{ns}x}$, where x is the axial position along the cylindrical shell axis and s is the particular axial wave. It will be seen in the next section that there are eight admissible axial waves in a cylindrical shell ($s=1,\ldots,8$). The resultant wave number at frequency f for circumferential mode n and particular wave s is $k_{R,ns} = \sqrt{k_c^2 + k_{ns}^2}$. The propagating waves in a cylindrical shell are helical waves.

Using the separated wave components described above, the complete travelling wave solution for $n \geqslant 1$ may be expressed as

$$\mathbf{u}(x,\theta,\omega) = \sum_{n=0}^{\infty}\sum_{s=1}^{8} \mathbf{U}_{ns}\cos(n\theta)e^{k_{ns}x}e^{i\omega t}, \tag{3a}$$

$$\mathbf{v}(x,\theta,\omega) = \sum_{n=0}^{\infty}\sum_{s=1}^{8} \mathbf{V}_{ns}\sin(n\theta)e^{k_{ns}x}e^{i\omega t}, \tag{3b}$$

and

$$\mathbf{w}(x,\theta,\omega) = \sum_{n=0}^{\infty}\sum_{s=1}^{8} \mathbf{W}_{ns}\cos(n\theta)e^{k_{ns}x}e^{i\omega t}, \tag{3c}$$

where $\mathbf{U}_{ns}$, $\mathbf{V}_{ns}$, and $\mathbf{W}_{ns}$ are the complex wave amplitudes for circumferential mode n and axial wave s, and $\omega = 2\pi f$ is the angular frequency.

Since the complete response at frequency f is a linear summation over each circumferential mode n, the response for each circumferential mode can be calculated separately and then summed to give the total response at frequency f. This reduces the analysis of cylindrical shells to a one-dimensional problem in the axial direction for each circumferential mode n.

C. Characteristic equation

The characteristic equation for cylindrical shell vibration at nondimensional frequency Ω is derived by substituting the travelling wave solution from Eq. (3) into the equations of motion [Eq. (1)] to give a 3×3 homogeneous matrix equation with unknowns $\mathbf{U}_{ns}$, $\mathbf{V}_{ns}$, and $\mathbf{W}_{ns}$, and then equating the determinant of the coefficient matrix to zero for nonvanishing $\mathbf{U}_{ns}$, $\mathbf{V}_{ns}$, and $\mathbf{W}_{ns}$. The resulting equation is the system characteristic equation for $n \geqslant 1$, and is a polynomial of order four in terms of the nondimensional wave number $k_{ns}a$ squared. The system characteristic equation is simply

$$g_8\{(k_{ns}a)^2\}^4 + g_6\{(k_{ns}a)^2\}^3 + g_4\{(k_{ns}a)^2\}^2 + g_2\{(k_{ns}a)^2\} + g_0 = 0 \tag{4}$$

with real coefficients g_0, g_2, g_4, g_6, and g_8 given by

$$\begin{aligned} g_0 = {} & \mu_1\{\beta^2 n^4 - 2\beta^2 n^6 + \beta^2 n^8 + \beta^4 n^4 - 2\beta^4 n^6 + \beta^4 n^8\} \\ & + \Omega^2\{-n^2\mu_1 - n^4\mu_1 - \beta^2 n^2 - 2\mu_1\beta^2 n^2 + 2\beta^2 n^4 \\ & + \mu_1\beta^2 n^4 - \beta^2 n^6 - \mu_1\beta^2 n^6 - \mu_1\beta^4 n^2 + 2\mu_1\beta^4 n^4 \\ & - \mu_1\beta^4 n^6\} + \Omega^4\{1 + \beta^2 + n^2 + n^2\mu_1 - 2\beta^2 n^2 \\ & + \mu_1\beta^2 n^2 + \beta^2 n^4\} - \Omega^6, \end{aligned} \tag{5a}$$

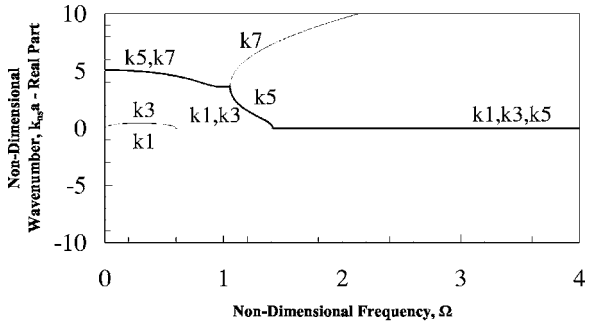
$$\begin{aligned} g_2 = {} & n^2\mu^2 - n^2\mu_1^2 - 2n^2\mu\mu_3 + n^2\mu_3^2 + \beta^2 n^2(-1 - 5\mu_1^2 \\ & + \mu_3^2) + \beta^2 n^4(2 - 2\mu\mu_1 + 2\mu_1^2 + 2\mu_1\mu_3 - 2\mu_3^2 \\ & + 2\mu_1\mu_4) + \beta^2 n^6(-1 - 2\mu_1 - \mu_1^2 + \mu_3^2) - 7\beta^4 n^2\mu_1^2 \\ & + \beta^4 n^4(8\mu_1^2 + 2\mu_1\mu_4) + \beta^4 n^6(-2\mu_1 - 3\mu_1^2) \\ & - 3\beta^6 n^2\mu_1^2 + 6\beta^6 n^4\mu_1^2 - 3\beta^6 n^6\mu_1^2 + \Omega^2\{1 + \beta^2 + n^2 \\ & - \mu^2 + \mu_1 + 4\beta^2\mu_1 + 3\beta^4\mu_1 + n^2\mu_1^2 - n^2\mu_3^2 \\ & + \beta^2 n^2(-2 - 2\mu_1 + 2\mu\mu_1 + 4\mu_1^2 - 2\mu_4) \\ & + \beta^2 n^4(3 + 3\mu_1) + \beta^4 n^2(-6\mu_1 + 3\mu_1^2) + \beta^4 n^4(5\mu_1 \\ & - \mu_1^2)\} + \Omega^4\{-1 - \mu_1 - 3\beta^2\mu_1 - 2\beta^2 n^2\}, \end{aligned} \tag{5b}$$

$$\begin{aligned} g_4 = {} & \mu_1 + 4\beta^2\mu_1 + 3\beta^4\mu_1 - \mu^2\mu_1 - 3\beta^2\mu^2\mu_1 \\ & + \beta^2 n^2(-2\mu - 2\mu_1 + 2\mu\mu_1^2 + 2\mu_3 - 2\mu_4 + 2\mu\mu_3\mu_4) \\ & + \beta^2 n^4(2 + 2\mu_1 + 2\mu_1^2 - 2\mu_3^2) + \beta^4 n^2(-6\mu_1 + 6\mu\mu_1^2) \\ & + \beta^4 n^4(6\mu_1 + 8\mu_1^2 - \mu_1^3 - 2\mu_1\mu_3\mu_4 - \mu_1\mu_4^2) \\ & + \beta^6 n^4(6\mu_1^2 - 3\mu_1^3 - \mu_1\mu_4^2) + \Omega^2\{2\beta^2\mu - \mu_1 \\ & - 3\beta^2\mu_1 + \beta^2 n^2(-3 - 3\mu_1) + \beta^4 n^2(-9\mu_1 + \mu_4^2)\} \\ & + \Omega^4\beta^2, \end{aligned} \tag{5c}$$

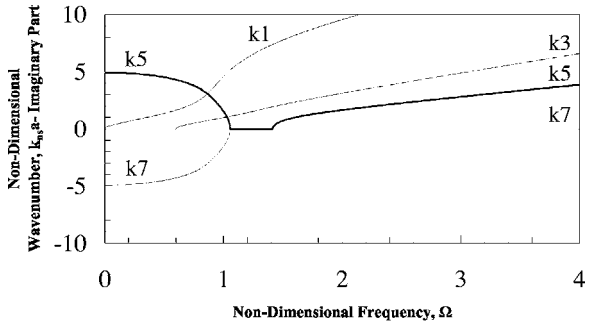
$$\begin{aligned} g_6 = {} & 2\beta^2\mu\mu_1 + 6\beta^4\mu\mu_1 + \beta^2 n^2(-1 - 2\mu_1 - \mu_1^2 + \mu_3^2) \\ & + \beta^4 n^2(1 - 6\mu_1 - 6\mu_1^2 - 2\mu_3\mu_4 + u_4^2) - 9\beta^6 n^2\mu_1^2 \\ & + \Omega^2\{\beta^2(1 - \beta^2 + \mu_1 + 3\beta^2\mu_1)\}, \end{aligned} \tag{5d}$$

$$g_8 = \mu_1(\beta^2 + 2\beta^4 - 3\beta^6). \tag{5e}$$

The constants μ_1, μ_2, μ_3, and μ_4 have been used to simplify the above equations and are defined by



(a)



(b)

FIG. 2. Wave-number curve for circumferential mode $n=1$ ($\beta=0.0192$, $\mu=0.28$): (a) real part; (b) imaginary part.

$$\mu_1 = (1-\mu)/2, \tag{6}$$

$$\mu_2 = 1-\mu^2, \tag{7}$$

$$\mu_3 = (1+\mu)/2, \tag{8}$$

and

$$\mu_4 = (3-\mu)/2. \tag{9}$$

The nondimensional frequency Ω is defined as

$$\Omega = \omega a\sqrt{\frac{\rho(1-\mu^2)}{E}} = \frac{\omega a}{c_L}, \tag{10}$$

where a is the shell mean radius and c_L is the longitudinal wave speed for a thin plate.

D. Roots of the characteristic equation

The roots of the characteristic equation for circumferential mode n at each frequency Ω give the admissible nondimensional wave numbers $k_{ns}a$, $s=1,\ldots,8$, which consist of four positive wave numbers and four negative wave numbers. The numeric form of the wave-number solutions varies depending on the type of wave and may be categorized as follows:

$k_{ns}a = \pm$ Imaginary — Type 1,
$k_{ns}a = \pm$ Real — Type 2,
$k_{ns}a = \pm$ (Real + Imaginary) — Type 3,
$k_{ns}a = \pm$ (Real − Imaginary) — Type 4.

Type 1 wave numbers give propagating waves, and the other wave number types represent various types of evanescent waves. At frequencies below the cutoff frequency of a particular circumferential mode n there are no propagating waves, and above the cutoff frequency there may be one, two or three pairs of propagating waves.

Simplified closed form solutions for the roots of the characteristic equation in Eq. (4), defining the nondimensional wave numbers $k_{ns}a$ at nondimensional frequency Ω, are

$$s=1,2: \quad (k_{ns}a)^2 = \Gamma_{16} - \frac{\sqrt{\Lambda_1}}{2} - \frac{1}{2}\sqrt{\Lambda_2 - \frac{\Lambda_3}{4\sqrt{\Lambda_1}}}, \tag{11a}$$

$$s=3,4: \quad (k_{ns}a)^2 = \Gamma_{16} - \frac{\sqrt{\Lambda_1}}{2} + \frac{1}{2}\sqrt{\Lambda_2 - \frac{\Lambda_3}{4\sqrt{\Lambda_1}}}, \tag{11b}$$

$$s=5,6: \quad (k_{ns}a)^2 = \Gamma_{16} + \frac{\sqrt{\Lambda_1}}{2} - \frac{1}{2}\sqrt{\Lambda_2 + \frac{\Lambda_3}{4\sqrt{\Lambda_1}}}, \tag{11c}$$

$$s=7,8: \quad (k_{ns}a)^2 = \Gamma_{16} + \frac{\sqrt{\Lambda_1}}{2} + \frac{1}{2}\sqrt{\Lambda_2 + \frac{\Lambda_3}{4\sqrt{\Lambda_1}}}, \tag{11d}$$

where

$$\Gamma_1 = g_4^2 - 3g_2g_6 + 12g_0g_8,$$

$$\Gamma_2 = 2g_4^3 - 9g_2g_4g_6 + 27g_0g_6^2 + 27g_2^2g_8 - 72g_0g_4g_8,$$

$$\Gamma_3 = (-4\Gamma_1^3 + \Gamma_2^2)^{1/2},$$

$$\Gamma_4 = \Gamma_2 + \Gamma_3,$$

$$\Gamma_5 = \Gamma_4^{1/3},$$

$$\Gamma_6 = 2^{1/3}\Gamma_1,$$

$$\Gamma_7 = g_6^2/4g_8^2,$$

$$\Gamma_8 = \Gamma_5/3g_8 2^{1/3},$$

$$\Gamma_9 = \Gamma_6/3\Gamma_4^{1/3}g_8, \tag{12a}$$

$$\Gamma_{10} = -8g_2/g_8,$$

$$\Gamma_{11} = g_6^3/g_8^3,$$

$$\Gamma_{12} = 4g_4g_6/g_8^2,$$

$$\Gamma_{13} = -2g_4/3g_8,$$

$$\Gamma_{14} = g_6^2/2g_8^2,$$

$$\Gamma_{15} = -1/3 g_8 2^{1/3},$$

$$\Gamma_{16} = -g_6/4g_8,$$

$$\Gamma_{17} = -4g_4/3g_8,$$

$$\Gamma_{18} = -\Gamma_6/3\Gamma_4^{1/3} g_8,$$

and

$$\Lambda_1 = \Gamma_{13} + \Gamma_7 + \Gamma_8 + \Gamma_9,$$

$$\Lambda_2 = \Gamma_{14} + \Gamma_{17} + \Gamma_{18} + \Gamma_{15}\Gamma_5, \tag{12b}$$

$$\Lambda_3 = \Gamma_{10} - \Gamma_{11} + \Gamma_{12}.$$

The constants g_0, g_2, g_4, g_6, and g_8 are the coefficients of the characteristic equation. These constants are given by Eq. (5) and are a function of frequency.

The equations presented here were derived using the symbolic algebra package *Mathematica*, and then simplified within the same package by performing the substitutions in Eq. (12). The above equations can be programmed directly for calculation of the eight admissible wave numbers at each frequency of interest.

III. CALCULATED RESULTS

A. Nondimensional wave-number plots

Curves of nondimensional wave number versus frequency for the four positive wave numbers of a cylindrical shell with thickness factor $\beta=0.0192$ are plotted in Figs. 2–4

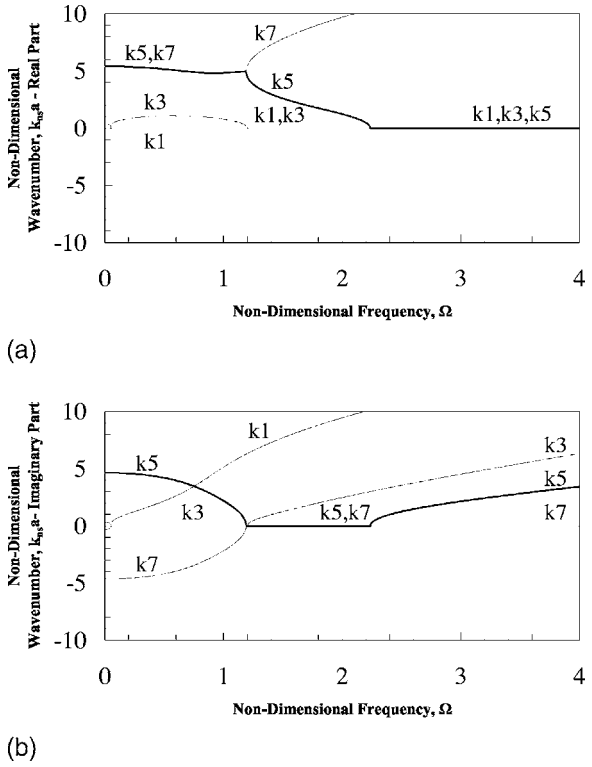


FIG. 3. Wave-number curve for circumferential mode $n=2$ ($\beta=0.0192$, $\mu=0.28$): (a) real part; (b) imaginary part.

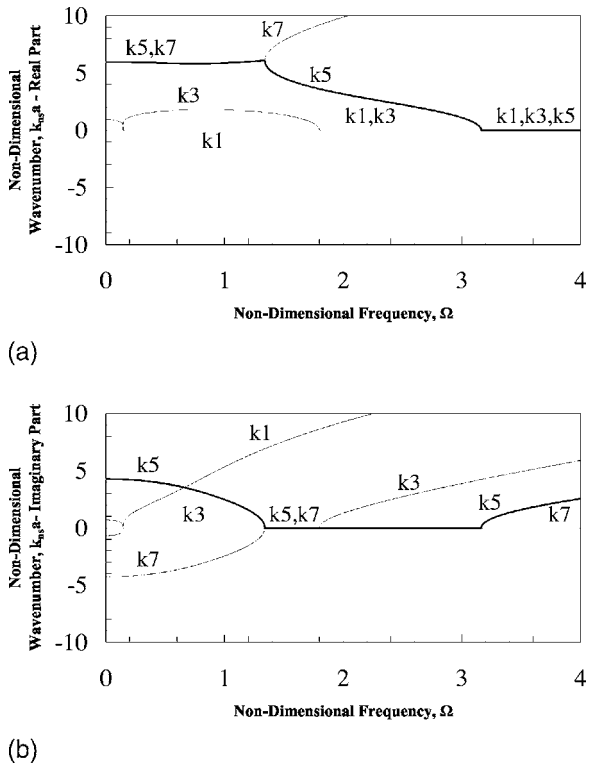


FIG. 4. Wave-number curve for circumferential mode $n=3$ ($\beta=0.0192$, $\mu=0.28$): (a) real part; (b) imaginary part.

for the circumferential modes $n=1$, $n=2$, and $n=3$, respectively. The four positive wave numbers are ordered using sorting procedures and have been labelled $k1$, $k3$, $k5$, and $k7$ for identification corresponding to the positive wave numbers k_{n1}, k_{n3}, k_{n5}, and k_{n7}, respectively. The curves are plotted as nondimensional wave number versus nondimensional frequency and are a function of only the circumferential mode number n and the thickness factor β. The wave number type is seen to vary with frequency.

B. Testing

The wave-number expressions presented in Sec. II have been tested by comparison with numerical results presented in the literature by Fuller[5] and Lee *et al.*[6]

Fuller presents charts of nondimensional wave number versus nondimensional frequency for each pair of waves for the first few circumferential modes. The wave-number calculations were required as part of a study into wave transmission at discontinuities in cylindrical shells using Flügge's equations of motions. The curves identify the various wave types and the transitions in wave type as the nondimensional frequency is increased. The results of Fuller were obtained

TABLE I. Comparison of numerical results for wave number from Lee *et al.* (Ref. 2) against calculations using Eq. (11) for $n=1$.

s	Lee *et al.* (Ref. 2)	Equation (11)
1,2	121.99	121.87
3,4	121.49	121.61
5,6	29.13	29.12
7,8	14.93	14.92

TABLE II. Comparison of numerical results for natural frequencies as given by Lee *et al.* (Ref. 2) against calculations based on Eq. (11).

	Source	Natural frequency (Hz)			
$n=1$	Lee *et al.* (Ref. 2)	14 328	14 628	15 815	16 246
	Equation (11)	14 305	14 630	15 817	16 223
$n=5$	Lee *et al.* (Ref. 2)	13 681	15 615	17 740	25 314
	Equation (11)	13 629	15 565	17 692	25 272
$n=8$	Lee *et al.* (Ref. 2)	12 824	14 545	15 788	16 434
	Equation (11)	12 737	14 462	15 694	16 355
$n=12$	Lee *et al.* (Ref. 2)	14 270	14 966	15 866	16 968
	Equation (11)	14 176	14 870	15 768	16 869
$n=13$	Lee *et al.* (Ref. 2)	15 715	15 812	16 106	17 288
	Equation (11)	15 721	16 014	16 503	17 192

using iterative numerical methods to solve for the roots of the characteristic equation, whereas the current work uses direct evaluation of Eq. (11). The numerical results of Fuller and Eq. (11) were found to be in agreement, within the limitations imposed in reading numerical values from the curves given by Fuller. It should be noted that when interpreting the curves given by Fuller, a different definition for nondimensional wave number was used.

In the case of Lee *et al.*, direct numerical comparisons of wave number were possible for circumferential mode $n=1$ (refer to Table I), as well as comparisons of natural frequency for a finite cylindrical shell for several values of n from $n=1$ to $n=13$ (refer to Table II). The agreement in each case confirms the correctness of the equations presented in Sec. II. Small differences in the results are attributed to Lee *et al.* using Donnell's shell equations, whereas Flügge's equations were used to derive the equations in Sec. II. The comparisons with the data in Lee are for a cylindrical shell with simple support (axially constrained), length of 304.8 mm, shell mean radius of 99.187 mm, wall thickness of 3.81 mm, Poisson's ratio of 0.28, density of 7700 $\mathrm{kg/m^3}$, and modulus of elasticity of 19.5×10^{10} $\mathrm{N/m^2}$. The natural frequencies presented by Lee *et al.* are for nondimensional frequencies Ω close to 2.

IV. CONCLUSIONS

Algebraic expressions for direct evaluation of wave number have been derived for cylindrical shell vibration described by the Flügge equations of motion. These expressions avoid the need to use simplified formulations when calculating wave number for cylindrical shell vibration, and avoid the need to program iterative numerical solutions when using the more accurate Flügge equations of motion for cylindrical shell vibration. The expressions presented for evaluation of wave number have been tested by comparison with calculations of wave number given by Fuller[5] and Lee *et al.*[6] Applications of the expressions presented here include the evaluation of nondimensional correlation ratios for the estimation of dynamic stress from vibration velocity (Karczub and Norton[2]), and calculations of transmission loss based on the procedures of Fagerlund.[1]

ACKNOWLEDGMENTS

Preparation of this paper was completed while the first author was on a technical visit to the R.A. Engel Technical Center at Fisher Controls, Marshalltown, Iowa for development of new piping system noise prediction methods. The author would like to acknowledge the technical support and hospitality of Dr. Al Fagerlund and test engineers Fred Catron and Charley DePenning during these visits. The contributions to this paper and associated work by Professor Michael Norton, who passed away during preparation of this paper, is gratefully acknowledged. Professor Norton was a substantial contributor to the Noise and Vibration community throughout his career.

[1] A. C. Fagerlund and D. C. Chou, "Sound transmission through a cylindrical pipe wall," J. Eng. Ind. **103**, 355–360 (1981).

[2] D. G. Karczub and M. P. Norton, "Correlations between dynamic strain and velocity in randomly excited plates and cylindrical shells with clamped boundaries," J. Sound Vib. **230**, 1069–1101 (2000).

[3] K. Forsberg, "Influence of boundary conditions on the modal characteristics of thin cylindrical shells," AIAA J. **2**, 2150–2157 (1964).

[4] A. W. Leissa, "Vibration of shells," NASA SP-288, 1973.

[5] C. R. Fuller, "The effects of wall discontinuities on the propagation of flexural waves in cylindrical shells," J. Sound Vib. **75**, 207–228 (1981).

[6] D. Lee, J. Jarzynski, and Y. H. Berthelot, "A study of wave propagation on a cylindrical shell using fiber optic laser doppler velocimetry," J. Acoust. Soc. Am. **94**, 196–212 (1993).

Sound absorption by viscoelastic coatings with periodically distributed cavities

Sven M. Ivansson
Division of Systems Technology, Swedish Defence Research Establishment, SE-164 90 Stockholm, Sweden

(Received 7 January 2005; revised 1 March 2006; accepted 6 March 2006)

Thin rubber layers with air-filled cavities can be used as anechoic submarine coatings. Normally incident sonar energy is redistributed in the lateral direction and absorbed. In this paper, the anechoic effect is studied theoretically and numerically by adapting techniques used in electron scattering and band-gap computations for photonic and phononic crystals. Reflection and transmission matrices are computed recursively, from basic ones for layers containing periodic arrays of spherical cavities. A method to locate zeroes of analytical functions is applied to prove the existence of, and to specify, thin coatings with vanishing reflectance at isolated frequencies. Coatings much thinner than quarter-wavelength ones are found. Most of the absorption loss takes place close to the cavities and scattering of compressional spherically symmetric waves is important. The viscoelastic shear-wave properties of the rubber are crucial for generating this loss. The requirements for vanishing reflectance are specified using a simplified model with normal plane waves and spherically symmetric waves, that includes effects of multiple scattering among the cavities. An energy relation is derived, relating the anelastic loss in the rubber coating to loss by monopole resonance scattering from isolated cavities. The noticeable effects of multiple scattering are incorporated by a modulating factor.
[DOI: 10.1121/1.2190165]

PACS number(s): 43.20.Fn, 43.20.Bi, 43.30.Gv, 43.30.Ky [ADP] Pages: 3558–3567

I. INTRODUCTION

Submarines are often equipped with anechoic coatings, in order to prevent detection by active sonar. Rubber coatings with air-filled cavities[1] were used already during the second world war. Such coatings are said to be of Alberich type, and an illustration is provided in Fig. 1. When sound is incident on the coating, part of it is scattered by the cavities. The scattered energy can subsequently be absorbed by the rubber material, and the amount of energy that is reflected back can be significantly reduced.

The echo reduction has been discussed and modeled by Gaunaurd *et al.*[2,3] in connection with RST (resonance scattering theory). The case of spherical scatterers, as in Fig. 1, is well adapted for an analytical treatment, since scattering by spheres in a solid medium is a classical topic.[4] Multiple scattering among the cavities was ignored in the mentioned studies, but the need for research concerning multiple-scattering corrections has been pointed out.[5]

At sufficiently low frequency, the coating may be modeled approximately as a homogeneous viscoelastic layer, with effective sound velocity and attenuation parameters.[6,7] The accuracy of the approximation needs to be evaluated, however.

The finite element method (FEM) has been used to model the echo reduction.[8,9] An important advantage is the flexibility to model different types of Alberich coatings. Utilizing periodicity, the computer intensive FEM modeling can be restricted to an individual unit cell. With coating layer reflection and transmission coefficients determined in this way, simple one-dimensional plane-wave computations have been shown feasible.[10]

In the present paper, echo reduction by Alberich anechoic coatings is studied with a faster semianalytical method, borrowed from atomic physics and applied in recent years to studies of band gaps for photonic and phononic crystals.[11,12] Indeed, sound scattering by cavities in the rubber layer is formally related to electron scattering by atoms in a lattice. Sound propagation through a sequence of layers may be handled recursively by the invariant embedding or Riccati method.[13,14] The wave field scattered by each cavity is expanded in spherical wave functions. An equation system is obtained for the coefficients, with multiple scattering among the cavities taken into account.[15] The theory is reviewed in Sec. II.

Reflections at a particular frequency can be eliminated by a classical quarter-wavelength coating.[13] Zero-finding algorithms for analytical functions are applied in Sec. III to show reflection elimination with much thinner Alberich coatings. Effects of multiple scattering among the cavities are observed, and the absorption process in the rubber layer is clarified.

The mechanism for the Alberich anechoic effect is studied in Sec. IV by wave-field expansions restricted to normal plane waves and spherically symmetrical waves. The absorption loss in the rubber layer is expressed as the loss induced by isolated cavities modified by a correction factor that, indeed, incorporates the effects of multiple scattering. The single-cavity loss is enhanced when the cavity circumference is about two shear-wave wavelengths.

Some conclusions and ideas for future work are provided in Sec. V.

 0001-4966/2006/119(6)/3558/10/$22.50

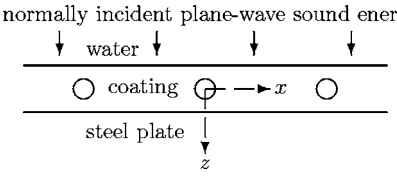


FIG. 1. A horizontal steel plate is covered with an Alberich rubber coating with spherical cavities in a quadratic lattice with period d (the distance between adjacent cavity centers). Plane-wave sound energy is normally incident from the water above. The x and z axes of a Cartesian xyz coordinate system are also indicated (dashed), and one of the cavities is centered at the origin.

II. NUMERICAL APPROACH

A right-hand Cartesian xyz coordinate system is introduced, with xy in the horizontal interface plane between two homogeneous solid or fluid half spaces and the z axis pointing downward. Sound waves with time dependence $\exp(-i\omega t)$ are considered, where ω denotes the angular frequency. In the solid case, because of three possible polarizations: P, SV, and SH, a plane wave incident from above gives rise to three reflected and three transmitted waves.[14] Including the three incident-wave polarizations, a 3×3 reflection-coefficient matrix $\mathcal{R}_B$ and a 3×3 transmission-coefficient matrix $\mathcal{T}_B$ may be formed. These matrices depend on the horizontal wave number, which is the same for all waves according to Snell's law. For a plane wave incident from below, the matrices $\mathcal{R}_A$ and $\mathcal{T}_A$ are introduced analogously.

A two-dimensional periodic array of similar cavities, interspersed in a homogeneous medium, is considered next. The centers of the cavities $\mathbf{R}$ are in the xy plane with

$$\mathbf{R} = (x,y,0) = m(d,0,0) + n(0,d,0), \tag{1}$$

where m,n are integers and d is the lattice period, cf. Fig. 1. For this cavity interface, R/T (reflection/transmission) matrices $\mathcal{R}_B,\mathcal{T}_B$ and $\mathcal{R}_A,\mathcal{T}_A$ can still be defined. However, reflected and transmitted waves appear with horizontal wave-number vectors different from that of the incident wave $\mathbf{k}_\parallel$. By a Fourier expansion in x,y, the appearing horizontal wave-number vectors are $\mathbf{k}_\parallel+\mathbf{g}$, where $\mathbf{g}$ belongs to the reciprocal lattice

$$\mathbf{g} = (k_x,k_y,0) = m(2\pi/d,0,0) + n(0,2\pi/d,0). \tag{2}$$

Displacement coefficients are used for the R/T matrices. With $\mathbf{r}=(x,y,z)$, the displacement vector for an incident, reflected, or transmitted plane wave is a multiple of

$$\mathbf{u}(\mathbf{r}) = \exp(i\mathbf{K}^s_{\mathbf{g}j}\cdot\mathbf{r})\mathbf{e}_j. \tag{3}$$

The time dependence $\exp(-i\omega t)$ is suppressed, and $j=1,2,3$ for a wave of type P, SV, SH, respectively. Furthermore, $s=+(-)$ for a wave going downward (upward), and

$$\mathbf{K}^\pm_{\mathbf{g}j} = \mathbf{k}_\parallel + \mathbf{g} \pm [\omega^2/c_j^2 - |k_\parallel+\mathbf{g}|^2]^{1/2}(0,0,1) \tag{4}$$

$$= \frac{\omega}{c_j}(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \tag{5}$$

where c_j is the compressional-wave velocity α when $j=1$ and the shear-wave velocity β when $j=2,3$. The angular variables θ and ϕ of $\mathbf{K}^\pm_{\mathbf{g}j}$ are introduced by Eq. (5), with a possibly complex $\cos\theta$. The vectors $\mathbf{e}_j=\mathbf{e}_j(\mathbf{K}^\pm_{\mathbf{g}j})$, finally, are defined by $\mathbf{e}_1=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)$, $\mathbf{e}_2=(\cos\theta\cos\phi,\cos\theta\sin\phi,-\sin\theta)$, $\mathbf{e}_3=(-\sin\phi,\cos\phi,0)$.

Including one cavity interface within the rubber layer, three interfaces are involved in Fig. 1. Individual R/T matrices can be combined recursively. With R/T matrices $\mathcal{R}_{B1},\mathcal{T}_{B1},\mathcal{R}_{A1},\mathcal{T}_{A1}$ and $\mathcal{R}_{B2},\mathcal{T}_{B2},\mathcal{R}_{A2},\mathcal{T}_{A2}$ for two interfaces given in the order of increasing z, with an intermediate level in between and phase shifts included to account for layer thicknesses, the following formulas are established for the total R/T matrices $\mathcal{R}_B,\mathcal{T}_B$:[15,14]

$$\mathcal{R}_B = \mathcal{R}_{B1} + \mathcal{T}_{A1}\cdot\mathcal{R}_{B2}\cdot[\mathbf{I}-\mathcal{R}_{A1}\cdot\mathcal{R}_{B2}]^{-1}\cdot\mathcal{T}_{B1}, \tag{6}$$

$$\mathcal{T}_B = \mathcal{T}_{B2}\cdot[\mathbf{I}-\mathcal{R}_{A1}\cdot\mathcal{R}_{B2}]^{-1}\cdot\mathcal{T}_{B1}, \tag{7}$$

where each $\mathbf{I}$ denotes the appropriate identity matrix. Formulas for $\mathcal{R}_A,\mathcal{T}_A$ are analogous.

The solution at the intermediate level can be determined when the wave fields at the upper and lower levels (interfaces) have been computed. Complex displacement amplitudes are denoted by $\mathbf{D}$ (downgoing at the upper level), $\mathbf{U}$ (upgoing at the lower level), $\mathbf{\Phi}$ (upgoing at the intermediate level), and $\mathbf{\Psi}$ (downgoing at the intermediate level). The following relations are readily established:

$$\mathbf{\Phi} = \mathcal{R}_{B2}\cdot\mathbf{\Psi} + \mathcal{T}_{A2}\cdot\mathbf{U}, \tag{8}$$

$$\mathbf{\Psi} = \mathcal{R}_{A1}\cdot\mathbf{\Phi} + \mathcal{T}_{B1}\cdot\mathbf{D}. \tag{9}$$

Thus, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ can be determined by solving the linear equation systems

$$(\mathbf{I}-\mathcal{R}_{B2}\cdot\mathcal{R}_{A1})\cdot\mathbf{\Phi} = \mathcal{T}_{A2}\cdot\mathbf{U} + \mathcal{R}_{B2}\cdot\mathcal{T}_{B1}\cdot\mathbf{D}, \tag{10}$$

$$(\mathbf{I}-\mathcal{R}_{A1}\cdot\mathcal{R}_{B2})\cdot\mathbf{\Psi} = \mathcal{T}_{B1}\cdot\mathbf{D} + \mathcal{R}_{A1}\cdot\mathcal{T}_{A2}\cdot\mathbf{U}. \tag{11}$$

A. R/T matrices for an interface with periodically distributed cavities

Explicit expressions for the R/T matrices are well known for an interface between two homogeneous half spaces.[14] To handle an interface with periodically distributed cavities, the following spherical vector solutions to the wave equations can be used:[12]

$$\mathbf{u}^L_{lm}(\mathbf{r}) = \frac{\alpha}{\omega}\nabla[f_l(\omega r/\alpha)Y^m_l(\theta,\phi)], \tag{12}$$

$$\mathbf{u}^M_{lm}(\mathbf{r}) = if_l(\omega r/\beta)\frac{1}{\sqrt{l(l+1)}} \times\left(\frac{1}{\sin\theta}\frac{\partial Y^m_l(\theta,\phi)}{\partial\phi}\mathbf{e}_2(\mathbf{r}) - \frac{\partial Y^m_l(\theta,\phi)}{\partial\theta}\mathbf{e}_3(\mathbf{r})\right), \tag{13}$$

$$\mathbf{u}_{lm}^{N}(\mathbf{r}) = \frac{\mathrm{i}\beta}{\omega}\nabla \times \mathbf{u}_{lm}^{M}(\mathbf{r}), \tag{14}$$

where $r=|\mathbf{r}|$ and θ,ϕ are the angular variables of $\mathbf{r}$. $Y_l^m = Y_l^m(\theta,\phi)$ are the surface spherical harmonics. The index $l = 0,1,2,\ldots$, with $m=-l,\ldots,l$, but it is understood that $\mathbf{u}_{00}^{M} \equiv \mathbf{u}_{00}^{N} \equiv \mathbf{0}$. The notation $\mathbf{u}_{lm}^{0L}, \mathbf{u}_{lm}^{0M}, \mathbf{u}_{lm}^{0N}$ and $\mathbf{u}_{lm}^{+L}, \mathbf{u}_{lm}^{+M}, \mathbf{u}_{lm}^{+N}$ is used for the two basic cases with f_l as the spherical Bessel function j_l and f_l as the spherical Hankel function h_l^+, respectively.

For an incident plane wave as in Eq. (3), the total scattered field $\mathbf{u}_{\mathrm{sc}}$ can be written[12]

$$\mathbf{u}_{\mathrm{sc}}(\mathbf{r}) = \sum_{Plm}\left(b_{lm}^{+P}\sum_{\mathbf{R}} \mathrm{e}^{\mathrm{i}\mathbf{k}_\parallel\cdot\mathbf{R}}\mathbf{u}_{lm}^{+P}(\mathbf{r}-\mathbf{R})\right), \tag{15}$$

where the index P runs over $P=L,M,N$. The $\mathbf{R}=\mathbf{0}$ terms provide the field

$$\mathbf{u}_{\mathrm{sc},\mathbf{0}}(\mathbf{r}) = \sum_{Plm} b_{lm}^{+P}\mathbf{u}_{lm}^{+P}(\mathbf{r}) \tag{16}$$

scattered from the cavity at the origin. The incoming field $\mathbf{u}_{\mathrm{inc},\mathbf{0}}(\mathbf{r})$ on this cavity has two parts: the incident plane wave of type (3) and the scattered field from all other cavities. Expanding both parts in terms of $\mathbf{u}_{lm}^{0P}$, with expansion coefficients a_{lm}^{0P} and $b_{lm}^{\prime P}$, respectively,

$$\mathbf{u}_{\mathrm{inc},\mathbf{0}}(\mathbf{r}) = \sum_{Plm}(a_{lm}^{0P} + b_{lm}^{\prime P})\mathbf{u}_{lm}^{0P}(\mathbf{r}). \tag{17}$$

It follows by a T-matrix (transition matrix) argument that

$$b_{lm}^{+P} = \sum_{P'l'm'} T_{lm;l'm'}^{PP'}(a_{l'm'}^{0P'} + b_{l'm'}^{\prime P'}) \tag{18}$$

with explicit expressions for a_{lm}^{0P} and the T matrix $T_{lm;l'm'}^{PP'}$ for a spherical scatterer.[11,12] A second equation system is derived by translating each wave $b_{lm}^{+P}\mathbf{u}_{lm}^{+P}(\mathbf{r}-\mathbf{R})$ to the origin:[12]

$$b_{lm}^{\prime P} = \sum_{P'l'm'} \Omega_{lm;l'm'}^{PP'} b_{l'm'}^{+P'}, \tag{19}$$

where the computable matrix $\Omega_{lm;l'm'}^{PP'}$ depends on $\mathbf{k}_\parallel$, the lattice, ω/α, and ω/β. Substitution of Eq. (19) into Eq. (18) yields a linear equation system for determination of b_{lm}^{+P}. For spherical scatterers, the equation system splits in two subsystems.[15]

To obtain the R/T matrices, the expansion (15) must be transformed to plane waves, also called beams, of the type (3). The following relation is crucial for this purpose:[12,15]

$$\sum_{\mathbf{R}} \mathrm{e}^{\mathrm{i}\mathbf{k}_\parallel\cdot\mathbf{R}}\mathrm{h}_l^+(\omega|\mathbf{r}-\mathbf{R}|/c_j)Y_l^m(\widehat{\mathbf{r}-\mathbf{R}})$$

$$= \sum_{\mathbf{g}} \frac{2\pi(-\mathrm{i})^l c_j}{\omega d^2 K_{\mathbf{g}jz}^+} Y_l^m(\widehat{\mathbf{K}_{\mathbf{g}j}^{\pm}})\mathrm{e}^{\mathrm{i}\mathbf{K}_{\mathbf{g}j}^{\pm}\cdot\mathbf{r}}. \tag{20}$$

Here, $\mathbf{K}_{\mathbf{g}j}^+$ should be used for $z>0$ while $\mathbf{K}_{\mathbf{g}j}^-$ is needed for $z<0$. A caret indicates the angular variables of the indicated quantity, and $K_{\mathbf{g}jz}^+$ is the z component of $\mathbf{K}_{\mathbf{g}j}^+$. Using Eq. (20), a plane-wave or beam representation of Eq. (15) is easily obtained[11] by expressing derivatives of Y_l^m in terms of itself and $Y_l^{m\pm1}$. As anticipated, it is the reciprocal lattice (2) that provides the changes of the horizontal wave-number vectors.

Equation (15) is valid outside the spherical scatterers in the layer, since all boundary conditions, radiation conditions, and wave equations are fulfilled. Since it is derived from Eq. (15) together with Eq. (20), the plane-wave or beam representation is valid outside the spherical scatterers for nonzero z. A single-scattering approximation is implemented by ignoring the $b_{l'm'}^{\prime P'}$ contribution in Eq. (18). In this case, the expansion coefficients b_{lm}^{+P} are readily obtained from the T matrix without need to solve an equation system.

B. Energy absorption around a cavity

The time-averaged power transferred into the sphere $|\mathbf{r}| \leqslant r$, with $a \leqslant r \leqslant d-2a$ for the case of spherical cavities with radius a, is[16]

$$\mathcal{P}(r) = \int_{|\mathbf{r}|=r}\int \frac{\omega}{2}\,\mathrm{Im}[-\mathbf{T}(\mathbf{r})\cdot\mathbf{u}^*(\mathbf{r})]dS(\mathbf{r}), \tag{21}$$

where $dS(\mathbf{r})$ is the surface area element, $\mathbf{u}^*(\mathbf{r})$ is the conjugate displacement vector, and $\mathbf{T}(\mathbf{r})$ is the traction vector acting on a spherical surface with $|\mathbf{r}|=r$. Within homogeneous spherical shells, the wave field $\mathbf{u}(\mathbf{r})$ can be expanded in spherical waves according to

$$\mathbf{u}(\mathbf{r}) = \sum_{Plm} a_{lm}^{0P}\mathbf{u}_{lm}^{0P}(\mathbf{r}) + \sum_{Plm} b_{lm}^{+P}\mathbf{u}_{lm}^{+P}(\mathbf{r}) \tag{22}$$

$$= \sum_{Plm}\frac{1}{2}a_{lm}^{0P}\mathbf{u}_{lm}^{-P}(\mathbf{r}) + \sum_{Plm}\left(\frac{1}{2}a_{lm}^{0P} + b_{lm}^{+P}\right)\mathbf{u}_{lm}^{+P}(\mathbf{r}), \tag{23}$$

where a_{lm}^{0P} and b_{lm}^{+P} concern the incident (including possible multiple-scattering contributions $b_{lm}^{\prime P}$, cf. Sec. II A) and scattered fields, respectively. The basic vector wave functions $\mathbf{u}_{lm}^{-P}(\mathbf{r})$ are defined according to Eqs. (12)–(14) with f_l chosen as the spherical Hankel function h_l^-.

It is convenient to expand the basic solutions $\mathbf{u}_{lm}^{\pm P}(\mathbf{r})$, and the corresponding traction vectors acting on a spherical surface with $|\mathbf{r}|=r$, in vector spherical harmonics that are complete and orthonormal over the surface of the unit sphere. The r-dependent expansion coefficients are obtained from Eqs. (12)–(14), and Eqs. (8.33)–(8.34) of Ref. 17. The expression for $\mathcal{P}(r)$ is simplified by exploiting the orthonormality properties.

For purposes of reference, it is useful to consider an isolated cavity as well, for an incident plane compressional wave. In this case, $a_{lm}^{0M} = a_{lm}^{0N} = 0$ for all lm. With a real compressional-wave velocity α, the expansion coefficients can be simplified for large r by applying standard asymptotic expressions for spherical Hankel functions. Coupling terms disappear in the expression for $\mathcal{P}(r)$, between "$-L$" and "$+L$" terms, as well as between terms of different $P=L,M,N$ types. For the anelastic case, with a complex nonreal shear-wave velocity β,

$$\lim_{r\to\infty}\mathcal{P}(r) = \frac{\rho\alpha^3}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}|b_{lm}^{+L}|^2\left[-1-\mathrm{Re}\left(\frac{a_{lm}^{0L}}{b_{lm}^{+L}}\right)\right] \tag{24}$$

is obtained, where ρ is the density of the surrounding medium. An absorption cross section is then naturally defined

by division with the intensity of the incident compressional wave. The typical case is a plane wave of unit displacement amplitude, for which the intensity is $\rho\alpha\omega^2/2$. For a single spherical cavity of radius a, a further division with the geometrical cross-section πa^2 provides the normalized absorption cross section.

III. COMPLETE REFLECTION ELIMINATION

For an example of the type in Fig. 1, beams with different wave-number vectors or directions are reflected into the water, according to Eqs. (3)–(5). However, only a finite number of these beams are propagating while the remaining ones are evanescent with exponential decay in the upward (decreasing z) direction, as seen from Eq. (4) with $\mathbf{g}$ according to the reciprocal lattice (2). Henceforth, a normally incident plane wave from the water is assumed, $\mathbf{k}_\parallel=\mathbf{0}$ there, and only the normally reflected beam is propagating at low frequency. Additional propagating beams appear as the frequency passes $c/d, c\sqrt{2}/d, 2c/d$, and so forth, where c is the sound velocity of the water. The reflectance, henceforth denoted $\mathcal{R}$, is defined in this paper as the quotient of the complex displacement amplitudes at the water/rubber interface of the normally reflected and incident beams. The transmittance Γ is analogously defined involving the normal beam transmitted into the steel at the water/steel interface.

Starting from an existing program for photonic crystals,[18] the approach from Sec. II has been implemented for computation of $\mathcal{R}$. The reflectance $\mathcal{R}$ is an analytical function of the complex rubber shear-wave velocity β, since all appearing relationships have this property. The analyticity allows zeroes of $\mathcal{R}$ to be identified by numerical winding-number techniques,[19] whereby the argument variation of $\mathcal{R}$ is determined around search rectangles in the β plane. These search rectangles are split adaptively until no subrectangle contains more than one zero. The secant method is finally used to refine the estimate of each isolated zero.

With carefully implemented error control, the existence of zeroes can actually be proved. The argument variation of $\mathcal{R}$ around a closed path in the β plane is an integral multiple of 2π. A numerical value close to 2π, for example, implies that exactly one zero is enclosed.

A. Example case

An example case as in Fig. 1 is now considered, with a 10 mm thick rubber coating with a square cavity lattice of period $d=44$ mm in the middle. The cavities are spherical with radius $a=3$ mm and evacuated. Air filling produces almost identical results. The sound velocity and density are 1480 m/s and 1.0 kg/dm^3, respectively, for water as well as rubber. Direct reflections at the water/rubber interface are thus avoided at normal incidence, which simplifies the analytical treatment in Sec. IV. The steel half space below the coating has density 7.7 kg/dm^3, compressional-wave velocity 5850 m/s, and shear-wave velocity 3230 m/s. Only the rubber is anelastic, a viscoelastic solid with absorption for shear waves only.

The described technique was used to obtain vanishing reflectance for normally incident plane waves at 9 kHz. The computed β corresponds to rubber shear-wave velocity and shear-wave absorption of 83.6 m/s and 14.3 dB/wavelength, respectively. Figure 2 shows the resulting reflectance versus frequency curve. The anechoic effect is due to the cavities and the rubber shear-wave absorption. Without the cavities, $\mathcal{R}$ would be about 0.94, implying a reflection loss of only about 0.6 dB. Without the absorption, virtually all incident energy would also be reflected back into the water by the rubber/steel interface, whether or not any cavities were present.

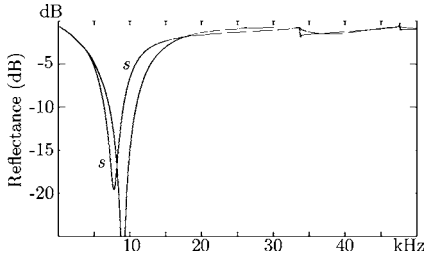


FIG. 2. Reflectance in decibel versus frequency for the example case, a coating as shown in Fig. 1 with parameter values given in the text. In particular, the rubber shear-wave parameters were determined to achieve vanishing reflectance at 9 kHz. The curve marked s concerns the single-scattering approximation result.

Comparison with the curve marked s, computed as indicated at the end of Sec. II A, provides an assessment of the effect of multiple scattering among the cavities. Although the lattice period $d=44$ mm is large in comparison to the cavity radius $a=3$ mm, it is only about 27% of the wavelength for compressional waves (P waves). The single-scattering approximation leads to clearly noticeable errors, with the 9 kHz dip appearing at 7.7 kHz.

Additional propagating beams are introduced at $c/d=33.6$ kHz and $c\sqrt{2}/d=47.6$ kHz. Slight reflectance reductions can be observed above these frequencies, when reflected energy is redistributed from the normal beam to other directions.

Rubber absorption of 10 dB/wavelength or more is used in Ref. 8, for compressional as well as shear waves. However, the compressional-wave absorption is not very important below about 15 kHz here, and the analysis in Sec. IV is simplified by omitting it. Strictly speaking, values of at least 0.06 dB/wavelength would be needed to meet the physical constraint of a bulk modulus in the fourth quadrant,[16,20] but this can safely be ignored.

Figure 3 shows time domain results obtained by Fourier synthesis. A pulse with spectrum in the band 7–11 kHz was selected (upper panel), for which significant echo reduction is anticipated from Fig. 2. The reflected pulse as viewed at the water/rubber interface (middle panel) is weak except in the vicinity of a spherical cavity (the central traces). The corresponding energy is built up by evanescent waves, however. It has disappeared long before the distance of 1 m into the water (lower panel). As can be verified by Fourier analysis, a notch filter suppresses the central part of a narrow-band pulse with symmetrical spectrum around the notch frequency. Indeed, this can be observed in the lower panel.

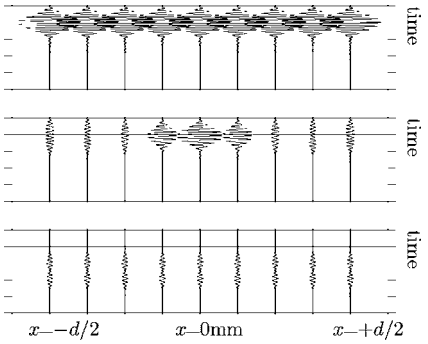


FIG. 3. Results of pulse computations by frequency synthesis, for the example case. The incident pulse is shown (upper panel) along with the reflected pulse as seen at the water/rubber interface (middle panel) and 1 m into the water (lower panel). Nine traces are drawn in each case, covering the overall period d=44 mm along the x axis. Time increases downward with 0.5 ms between tick marks, and the time at which the center of the symmetrical incident pulse has reached the water/rubber interface is indicated by a horizontal line.

B. Absorption loss

At low frequency, before non-normal propagating beams appear in the water and steel half spaces, the fraction of the incident energy that is lost by anelastic absorption is

$$1 - |\mathcal{R}|^2 - \frac{Z_{\text{steel}}}{Z_{\text{rubber}}}|\Gamma|^2, \tag{25}$$

where Z_{steel} and Z_{rubber} are the characteristic impedances[21] of the steel and the rubber, respectively. An effective normalized absorption cross section per cavity is obtained by multiplication with $d^2/\pi a^2$, and shown for the example case by curve (a) of Fig. 4. As expected from Fig. 2, very large values, almost $d^2/\pi a^2$=68.5, appear around 9 kHz.

The complex amplitudes of the plane waves that approach the cavities from above and below, respectively, can be computed using Eqs. (8)–(11). These plane waves can subsequently be expanded in spherical waves, and the wave field around a cavity can be determined using relations from Sec. II A. Using Sec. II B, finally, the anelastic loss within a spherical region enclosing a cavity can be computed. Curve (b) in Fig. 4 shows the contribution to the effective normalized absorption cross section from the part of the rubber layer within 2 mm of a particular cavity. Although cavity surroundings of such size occupy only about 2.1% of the rubber volume, about 71% of the loss at 9 kHz is suffered there.

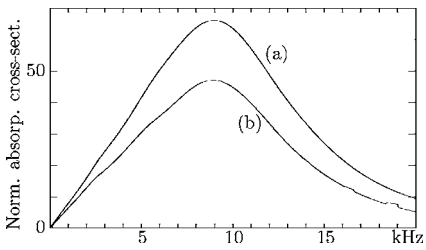


FIG. 4. The effective normalized absorption cross section per cavity is shown by curve (a), for the example case. The contribution from anelastic absorption within 2 mm of the spherical cavity is indicated by curve (b).

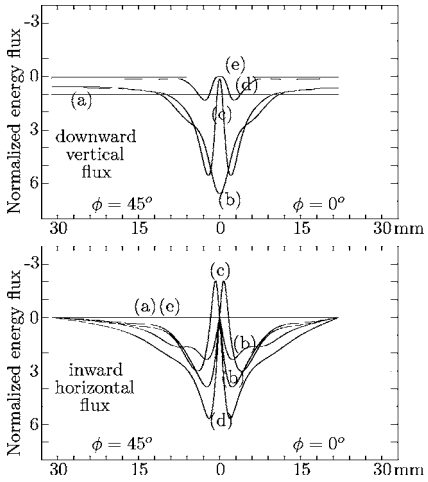


FIG. 5. Spatial distribution of time-averaged energy flux at 9 kHz, for the example case: vertical flux in the downward direction (upper panel), and horizontal flux in the inward direction (lower panel). The horizontal axis indicates horizontal distance, from the center of a particular cavity, along two different azimuthal directions: axial direction ϕ=0° (right part) and diagonal ϕ=45° direction (left part). The five curves (a)–(e) correspond to successively increasing depths z: (a) 1 dm above, (b) at, (c) 2 mm below the water/rubber interface, and (d) 2 mm above, (e) at the rubber/steel interface. The horizontal flux is not continuous across the water/rubber interface, however, and there are two (b) curves in the lower panel, with the lower one (indicating the larger flux) for the rubber side. The vertical axes show dimensionless flux values, normalized with the flux (in W/m^2) of the incident plane wave.

C. Energy flux

The results of Figs. 2 and 4 are now complemented by a study of the spatial distribution of the energy flux at 9 kHz. The upper panel of Fig. 5 shows the vertical component

$$\mathcal{F}_z = \frac{\omega}{2}\,\text{Im}(\tau_{xz}u_x^* + \tau_{yz}u_y^* + \tau_{zz}u_z^*), \tag{26}$$

of the time-averaged energy flux vector $\mathcal{F}$,[16] in terms of the indicated components of the displacement vector **u** and stress tensor τ. The complex conjugate is denoted by an asterix. The horizontal axis indicates the horizontal distance from the cavity at the origin. Results are given along two horizontal directions with different azimuths ϕ: the x axis direction ϕ =0° up to the point of symmetry at $x=d/2$=22 mm (right part), and the diagonal direction ϕ=45° up to the point of symmetry at $x=y=d/\sqrt{2}$=31.1 mm (left part). Five curves are shown, corresponding to five different vertical z levels: (a) 1 dm above, (b) at, (c) 2 mm below the water/rubber interface, and (d) 2 mm above, (e) at the rubber/steel interface.

It is the laterally constant energy flux of the incident plane wave that is seen in curve (a), since the normally reflected wave vanishes at 9 kHz and the evanescent reflected waves are no longer noticeable 1 dm above the water/rubber interface. Already when the interface is reached, curve (b), the vertical flux has been redistributed and focused above the spherical cavities. The total flux is unchanged, since no loss is suffered in the water, as is also realized by an integration over the unit cell with the correct area weighting.

The three remaining curves, (c)–(e), in the upper panel of Fig. 5 show a successively decreasing flux, caused by the

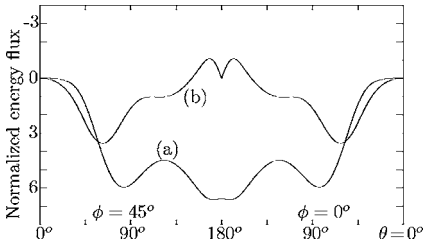


FIG. 6. Angular distribution of time-averaged energy flux at 9 kHz around a spherical cavity, 5 mm from its center, for the example case. The horizontal axis indicates the polar angle θ, as measured from the downward (increasing z) direction. Results are given for two different azimuths: $\phi=0°$ (right part) and $\phi=45°$ (left part). The curves (a) and (b) show the inward radial flux and the flux in the direction of decreasing θ, respectively. As in Fig. 5, the vertical axis shows dimensionless flux values, normalized with the flux (in W/m^2) of the incident plane wave.

loss suffered in the rubber layer. At $(x,y)=(0,0)$, $\mathcal{F}_z$ should vanish at the spherical surface because of the boundary condition and the symmetry. Indeed, curves (c) and (d) exhibit very small values at this point, although zeroes are not reached because of truncation to a finite number of plane and spherical waves. Curve (e) represents the small amount of energy that is transmitted into the steel half-space.

For points in the axial and diagonal directions, the transverse horizontal component of $\mathcal{F}$ vanishes by symmetry. The inward horizontal component is shown in the lower panel of Fig. 5. Vanishing values are obtained at $(x,y)=(0,0)$, as expected, and also at $x=d/2=22$ mm for the axial direction ($\phi=0°$) and at $x=y=d/\sqrt{2}=31.1$ mm for the diagonal direction ($\phi=45°$). Energy is pushed by an inward horizontal flux toward the spherical cavities, close to which it is absorbed, cf. Fig. 4. An outward horizontal flux is only noted in curve (c), close to the cavity top. Apparently, only part of the incoming energy at the top of a spherical cavity can be absorbed there. Combining both (c) curves, the remaining part appears to be transported downwards, close to the cavity surface, for absorption later.

Figure 6 shows the flux variation along a spherical surface with a radius of 5 mm centered at one of the cavities, with $\mathcal{F}$ in spherical basis vectors. The horizontal axis indicates the polar angle θ, as measured from the downward (increasing z) direction. Results for the azimuths $\phi=0°$ and $\phi=45°$ appear to the right and to the left, respectively. Curve (a) shows large inward radial flux at $r=5$ mm. This flux is fairly constant over the upper hemisphere ($90° < \theta < 180°$). Directly below a cavity (at $\theta=0°$), the flux is very small. The flux values, in relation to the flux of the incident plane wave, are consistent with an average inward flux of about 4.2, over the whole spherical surface at $r=5$ mm, needed to produce the absorption losses at 9 kHz from curve (b) of Fig. 4. The inward radial flux is accompanied by a flux in the orthogonal direction that appears to be smaller in general, by curve (b) in Fig. 6.

The differences between results for the axial and diagonal directions in Figs. 5 and 6 are small, indicating a similar behavior in different horizontal directions around each cavity. The only clear exceptions are the right and left end parts in the lower panel of Fig. 5, caused by the different distances to lattice points of symmetry in the axial and diagonal directions.

D. Different field representations

Two different representations of the field in a particular layer, with periodically distributed scatterers, have been introduced and combined in Sec. II. The first is a linear combination of plane waves of the type (3) with **g** from the reciprocal lattice (2). This representation was used for the flux computations for Fig. 5. It is most appropriate away from the scatterer interface where most of the evanescent waves are weak enough to be ignored, which reduces the number of plane waves (**g** values) needed. This is seen by Eq. (20).

The second representation involves spherical waves, separated into incoming and outgoing ones. For the scatterer at the origin, for example, these waves are expressed by Eqs. (17) and (16), respectively. This representation was used for the flux computations for Fig. 6. It is appropriate in the vicinity of the chosen scatterer, where a reasonably small number of spherical waves (lm values) is adequate.

Many plane-wave field components, including evanescent ones, are needed to reproduce the lateral variations in the time traces in the middle panel of Fig. 3. Accurate reflectance results, for the normally reflected beam, can be obtained with very few field components, however. The result of Fig. 2 is virtually unchanged if the computations are truncated to include only the monopole part, $Plm=L00$, for the spherical waves. The changes by a further truncation to normal beams, the $\mathbf{g}=\mathbf{0}$ plane waves, are rather small. Hence, since these truncations allow the numerical computations to be complemented with theoretical analysis, they are used in Sec. IV to analyze the mechanism behind the anechoic effect.

IV. SIMPLIFIED ANALYSIS OF THE ANECHOIC EFFECT

Repeating the winding-number computations of Sec. III to obtain vanishing reflectance at 9 kHz, but with the truncated fields, keeping the $Plm=L00$ and $\mathbf{g}=\mathbf{0}$ terms only, the rubber shear-wave parameter changes are small and the reflectance results are very similar to those in Fig. 2. The echo reduction mechanism, including multiple scattering, is now discussed within the simplified framework, for this example case. Specifically, cf. Sec. III D, a field representation with the monopole term $L00$ of Eqs. (17) and (16) is used around cavities within the rubber layer, and only the normal term $\mathbf{g}=\mathbf{0}$ of Eq. (20) is kept for the boundary-condition match at the interfaces to the normal plane waves in the water and steel half spaces. The monopole index specification $Plm=L00$ is now in general omitted.

A. Reflectance and transmittance

The cavity interface at $z=0$ in Fig. 1 is in rubber with density ρ, compressional-wave velocity α, and shear-wave velocity β. Substitution of Eqs. (18) and (19) into Eq. (17) yields

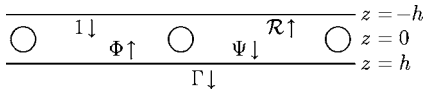


FIG. 7. The rubber layer from Fig. 1 is shown, with thickness $2h$ and the cavity interface at $z=0$ in the middle. Hence, the water and steel half spaces appear in $z<-h$ and $z>h$, respectively. Notation for some normal plane-wave complex displacement amplitudes is also included. The amplitudes 1 (for the incident wave) and $\mathcal{R}$ are evaluated at the water/rubber interface $z=-h$, while Φ,Ψ,Γ are evaluated at the rubber/steel interface $z=h$.

$$(a^0+b')\mathbf{u}_{00}^{0L}(\mathbf{r})=\frac{a^0}{1-\Omega T}\mathbf{u}_{00}^{0L}(\mathbf{r}) \tag{27}$$

as expressions for the spherical wave that is incoming on the cavity at the origin. For the basic case of a normally incident plane compressional wave with complex displacement amplitude unity at $z=0$, $a^0=-\mathrm{i}\sqrt{4\pi}$. Equation (15) for the scattered field then reduces to

$$\mathbf{u}_{\mathrm{sc}}(\mathbf{r})=\Theta\mathrm{e}^{\mathrm{i}k|z|}\mathrm{sgn}(z)\mathbf{e}_z \tag{28}$$

by applying Eqs. (12) and (20), the latter with restriction to the normal plane wave with $\mathbf{g}=\mathbf{0}$. Here, $k=\omega/\alpha$, $\mathbf{e}_z$ is the unit vector in the z direction, and Θ is explicitly given by

$$\Theta=\frac{T}{1-\Omega T}\frac{2\pi}{(kd)^2}. \tag{29}$$

With the sum interpreted as a limit for k with $\mathrm{Im}(k)$ decreasing to zero, Ω can be written[22]

$$\Omega=\sum_{\mathbf{R}\neq 0}\mathrm{h}_0^+(k|\mathbf{R}|). \tag{30}$$

The boundary condition of vanishing normal traction at $r=a$ is fulfilled with

$$T=-\frac{ka\mathrm{j}_0(ka)+4(\beta^2/\alpha^2)\mathrm{j}_0'(ka)}{ka\mathrm{h}_0^+(ka)+4(\beta^2/\alpha^2)(\mathrm{h}_0^+)'(ka)}. \tag{31}$$

The water/rubber and rubber/steel interfaces from Fig. 1 are now incorporated at $z=-h$ and $z=h$, respectively, with $h=5$ mm. This is shown in Fig. 7. The displacement reflection coefficient at the rubber/steel interface is denoted σ. With the restriction to normal beams, the water/rubber interface does not produce any reflections, since the water and rubber are assumed to differ in their shear-wave parameters only. It is still relevant for phase reference, however. With $E=\exp(\mathrm{i}\omega h/\alpha)$, $\mathcal{R}_{B1}=\mathcal{R}_{A1}=E^2\Theta$, $\mathcal{T}_{B1}=\mathcal{T}_{A1}=(1+\Theta)E^2$, $\mathcal{R}_{B2}=\sigma$, and $\mathcal{T}_{B2}=1-\sigma$, Eqs. (6) and (7) now provide the desired reflectance and transmittance relations

$$\mathcal{R}=\mathcal{R}_B=\frac{(1+2E^2\sigma)\Theta+E^2\sigma}{1-E^2\sigma\Theta}E^2, \tag{32}$$

$$\Gamma=\mathcal{T}_B=\frac{(1-\sigma)(1+\Theta)E^2}{1-E^2\sigma\Theta}. \tag{33}$$

Just above the rubber/steel interface in Fig. 7, the wave field is described in terms of complex displacement amplitudes Φ and Ψ of up- and down-going waves, respectively, as evaluated at that interface. Eqs. (10) and (11) yield conveniently, with $\mathbf{D}=1$ and $\mathbf{U}=0$,

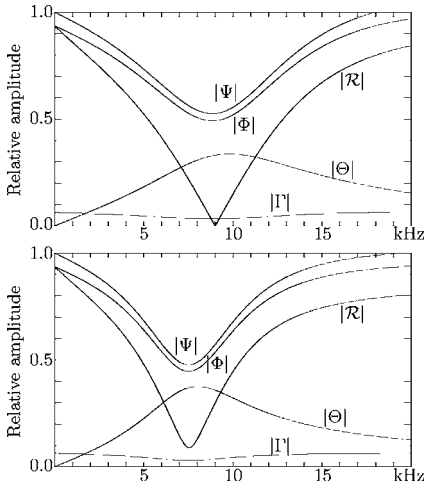


FIG. 8. Amplitudes (relative the incident-wave amplitude) of the field quantities from Fig. 7, and $|\Theta|$ as well, are plotted as functions of frequency, for the example case. The lower panel concerns single-scattering approximation results, with Ω set to zero. Changed to decibels, each $|\mathcal{R}|$ curve is very similar to corresponding 0–20 kHz results from Fig. 2.

$$\Phi=\frac{\sigma(1+\Theta)E^2}{1-E^2\sigma\Theta}, \tag{34}$$

$$\Psi=\frac{(1+\Theta)E^2}{1-E^2\sigma\Theta}. \tag{35}$$

The magnitudes of $\mathcal{R},\Theta,\Phi,\Psi,\Gamma$ are plotted as functions of frequency in the upper panel of Fig. 8, for the example case with $\mathcal{R}=0$ at 9 kHz. It appears that the cavity-interface reflection amplitude Θ should be comparatively large in order to produce a small reflectance value $\mathcal{R}$. Corresponding results for the single-scattering approximation, with Ω set to zero, are shown in the lower panel. In particular, cf. Fig. 2, the minimum of $|\mathcal{R}|$ appears at about 7.5 rather than 9 kHz.

Corresponding results for an elastic rubber material (without absorption) are shown in Fig. 9, with large values of $|\mathcal{R}|$ at all frequencies. Instead of a minimum, a maximum

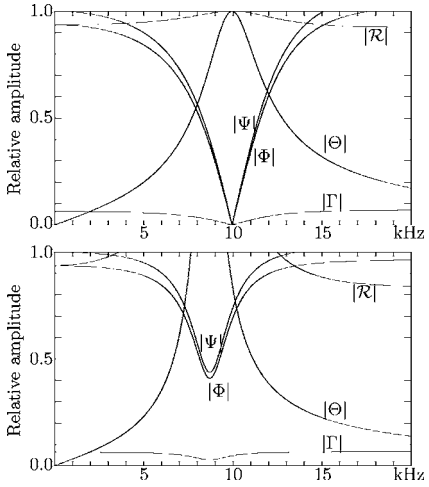


FIG. 9. As Fig. 8 but for an elastic rubber material (the shear-wave absorption is set to zero). Again, the lower panel concerns single-scattering approximation results.

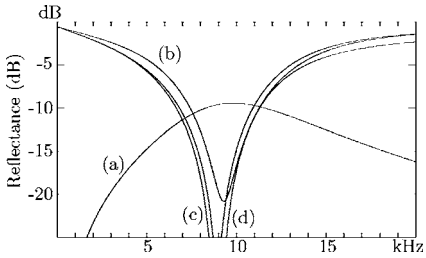


FIG. 10. Reverberation-operator approximations with (a) one, (b) two, (c) three, and (d) four terms according to Eq. (36), for the example case. The basic reflectance curve in Fig. 2 is recovered well.

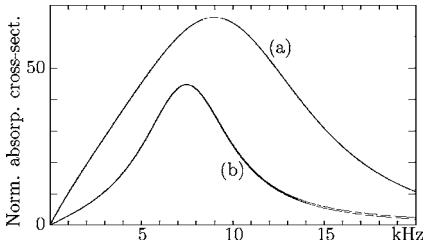


FIG. 11. The effective normalized absorption cross section per cavity is shown by curve (a), for the example case. The normalized absorption cross section for a single spherical cavity of radius 3 mm embedded in the rubber material is shown by (b). Two almost coinciding (b) curves are shown, the lower one includes only the $lm=00$ contribution, see Eq. (24).

with almost total reflection appears at about 10.0 kHz when multiple-scattering effects are included. Indeed, transmission band gaps for phononic crystals are well known.[12] Unphysical single-scattering results appear, with $|\Theta|$ as well as $|\mathcal{R}|$ exceeding unity. As clarified in Sec. IV E, see Eq. (49), multiple-scattering effects are essential to avoid such inconsistencies.

The single-scattering maxima for $|\Theta|$ in Figs. 8 and 9 correspond to maxima for $|T|/\omega^2$, cf. Eq. (29). Related resonance effects are discussed in Sec. IV C.

B. Expanding the Riccati reverberation operator

The reverberation operator $[\mathbf{I}-\mathcal{R}_{A1}\cdot\mathcal{R}_{B2}]^{-1}$ appearing in Eq. (6) can be expanded in a geometric series[14] and individual wave-field components can be isolated. Eq. (6) yields

$$\mathcal{R}_B=\mathcal{R}_{B1}+\mathcal{T}_{A1}\cdot\mathcal{R}_{B2}\cdot\mathcal{T}_{B1}+\mathcal{T}_{A1}\cdot\mathcal{R}_{B2}\cdot\mathcal{R}_{A1}\cdot\mathcal{R}_{B2}\cdot\mathcal{T}_{B1}+\cdots. \tag{36}$$

With $\mathcal{R}=\mathcal{R}_B$ of Eq. (32), $E=\exp(\mathrm{i}\omega h/\alpha)$, $\mathcal{R}_{B1}=\mathcal{R}_{A1}=E^2\Theta$, $\mathcal{T}_{B1}=\mathcal{T}_{A1}=(1+\Theta)E^2$, $\mathcal{R}_{B2}=\sigma$, and $\mathcal{T}_{B2}=1-\sigma$, Fig. 10 shows how the basic reflectance curve of Fig. 2 is recovered. The one-term result is merely a decibel version of the $|\Theta|$ curve from the upper panel of Fig. 8. Already the two-term result, with no more than one wave interaction with the rubber/steel interface, reproduces the anechoic effect rather well. A destructive interference among the waves in the right-hand side of Eq. (36) is apparently essential.

C. Requirements for vanishing reflectance

Together with Eqs. (32) and (33), the expression (25) multiplied with $d^2/\pi a^2$ now provides the effective normalized absorption cross-section per cavity. It is shown by curve (a) of Fig. 11, which is, as expected, very similar to curve (a) of Fig. 4.

At a frequency for which $\mathcal{R}=0$, it follows from Eqs. (32) and (33) that

$$\Theta=-\frac{E^2\sigma}{1+2E^2\sigma}, \tag{37}$$

$$\Gamma=\frac{(1-\sigma)E^2}{1+E^2\sigma}. \tag{38}$$

For the example case, $\sigma=0.9364$ and $E=0.9818+0.1899\mathrm{i}$ at 9 kHz. At this frequency, for which $\mathcal{R}=0$, it follows that $\Theta=-0.3285-0.0437\mathrm{i}$, and that only 3.4% of the incident energy is transmitted through the steel half space while 96.6% is absorbed. Since the coating is thin (10 mm) in relation to the P wavelength at 9 kHz (164 mm), a good approximation is obtained by taking $h=0$ m and $E=1$ in Eq. (37), resulting in $\Theta=-0.3260$.

It remains to understand how scattering from the cavities can produce this particular value of Θ. Effective normalized cross-sectional values as large as 66.0, as seen at 9 kHz in curve (a) of Fig. 11, indicate a resonance effect. A spherical cavity in an elastic solid with a reasonably small shear-wave velocity has a monopole scattering resonance for the radial displacement amplitude at the sphere surface, the classical Meyer *et al.*[4] and Gaunaurd *et al.*[3] resonance, when the circumference equals two shear-wave wavelengths in the solid. With significant absorption in the solid, however, the resonance shows up more clearly in the absorption cross-section.

The corresponding normalized absorption cross section for a single spherical cavity, computed according to Eq. (24), is shown by curve (b) of Fig. 11. There is a peak frequency shift from 7.5 kHz in curve (b) to 9.0 kHz in curve (a). It is caused by multiple scattering among the cavities, as clarified in Sec. IV E.

D. Wave field around a cavity

The wave field in the vicinity of the sphere at the origin can be written

$$\frac{(1+\Phi)E}{1-\Omega T}\mathbf{u}_{\mathrm{ref}}(\mathbf{r}), \tag{39}$$

where $E=\exp(\mathrm{i}\omega h/\alpha)$ and, as follows by inspection of Fig. 7 together with Eqs. (27) and (18),

$$\mathbf{u}_{\mathrm{ref}}(\mathbf{r})=a^0[\mathbf{u}_{00}^{0L}(\mathbf{r})+T\mathbf{u}_{00}^{+L}(\mathbf{r})]=-\mathrm{i}[\mathrm{j}_0'(kr)+T(\mathrm{h}_0^+)'(kr)]\mathbf{e}_r. \tag{40}$$

Here, $a^0=-\mathrm{i}\sqrt{4\pi}$ is the expansion coefficient for a plane compressional wave of displacement amplitude unity and $\mathbf{e}_r$ is the radial unit vector.

Equation (24) shows that the normalized total power loss by anelastic absorption for the single-cavity reference wave field $\mathbf{u}_{\mathrm{ref}}(\mathbf{r})$ equals $4|T|^2[-1-\mathrm{Re}(1/T)]/(ka)^2$, the natural normalization being made with the cross-sectional power for an incident plane wave with displacement amplitude unity. Expressed in another way, $4|T|^2[-1-\mathrm{Re}(1/T)]/(ka)^2$ equals the $lm=00$ contribution to the normalized ab-

sorption cross section for a single cavity which, for the example case, has already been plotted as the lower (b) curve in Fig. 11.

For any spherically symmetric wave field, the time-averaged loss density, or strain power density,[20] at radius r equals $\mathcal{P}'(r)/4\pi r^2$, with $\mathcal{P}$ from Eq. (21). Explicit computations are possible for the single-cavity reference wave field $\mathbf{u}_{\text{ref}}(\mathbf{r})=u(r)\mathbf{e}_r$, with $u(r)$ according to Eq. (40). The radial stress τ_{rr} at the surface $|\mathbf{r}|=r$ can be written

$$\tau_{rr}=\rho\alpha^2\left(\frac{\partial u}{\partial r}+\frac{2u}{r}\right)-4\mu\frac{u}{r}, \tag{41}$$

where $\mu=\rho\beta^2$ is the rubber shear modulus and $-4\mu u/r$ is a restoring stress.[23] It follows that the strain power density includes the factor $\text{Im}(-\mu)$, and that the shear-wave absorption generates large losses close to the cavity even for the spherically symmetric field $\mathbf{u}_{\text{ref}}(\mathbf{r})$.

E. Energy relation

The effective normalized absorption cross section per cavity, for the truncated fields, can be expressed in another way, where the effects of multiple scattering are separated in a modulating factor. Specifically, the following energy relation is valid when $k=\omega/\alpha<2\pi/d$:

$$\frac{d^2}{\pi a^2}\left(1-|\mathcal{R}|^2-\frac{Z_{\text{steel}}}{Z_{\text{rubber}}}|\Gamma|^2\right)=\left|\frac{1+\Phi}{1-\Omega T}\right|^2\frac{4|T|^2[-1-\text{Re}(1/T)]}{(ka)^2}. \tag{42}$$

As a first step in order to prove Eq. (42), substitution of Eqs. (32) and (33) and the relation $Z_{\text{steel}}/Z_{\text{rubber}}=(1+\sigma)/(1-\sigma)$ into the left-hand side leads to

$$\frac{d^2}{\pi a^2|1-E^2\sigma\Theta|^2}[|1-E^2\sigma\Theta|^2-|\Theta+E^2\sigma+2E^2\sigma\Theta|^2-(1-\sigma^2)|1+\Theta|^2]. \tag{43}$$

Application of Eqs. (29) and (34) in the right-hand side produces the expression

$$\frac{|1+E^2\sigma|^2}{|1-E^2\sigma\Theta|^2}\left(\frac{kd^2}{\pi a}\right)^2|\Theta|^2\left[-1-\text{Re}\left(\frac{1}{T}\right)\right]. \tag{44}$$

When k is real with $kd<2\pi$, a wave-theoretic argument[24] provides the explicit expression

$$\text{Re}(\Omega)=\frac{2\pi}{(kd)^2}-1. \tag{45}$$

Using Eq. (45) together with Eq. (29), the expression (44) can be written

$$-\frac{2d^2|1+E^2\sigma|^2[|\Theta|^2+\text{Re}\Theta]}{\pi a^2|1-E^2\sigma\Theta|^2}. \tag{46}$$

It follows that Eq. (42) boils down to the relation

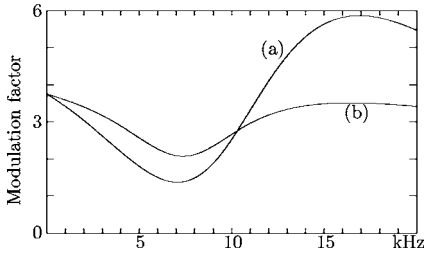


FIG. 12. Curve (a) shows the modulating factor (50) for the effective normalized absorption cross section per cavity in Eq. (42), for the example case. The too-weak modulation that would result by ignoring multiple scattering among the cavities, and setting Ω to zero in Eq. (50), is indicated by curve (b).

$$|1-E^2\sigma\Theta|^2-|\Theta+E^2\sigma+2E^2\sigma\Theta|^2-(1-\sigma^2)|1+\Theta|^2=-2|1+E^2\sigma|^2[|\Theta|^2+\text{Re}\Theta]. \tag{47}$$

When $|E|=1$ and σ is real, it is easy to verify Eq. (47). This concludes the proof.

Since it is assumed that α and ka are real, it follows from Eq. (31) and a Wronskian expression that

$$-1-\text{Re}(1/T)=\frac{4\,\text{Im}(-\beta^2)}{\alpha^2 ka|kaj_0(ka)+4(\beta^2/\alpha^2)j_0'(ka)|^2}\geqslant 0. \tag{48}$$

Hence, both sides of Eq. (42) are non-negative. For the particular case when the steel is replaced by rubber, $\sigma=0$ and it follows from Eqs. (42) and (48) that

$$|\Theta|^2+|1+\Theta|^2\leqslant 1. \tag{49}$$

Multiple-scattering effects are crucial for this bound on Θ, cf. Eq. (29) and Fig. 9.

The factor $4|T|^2[-1-\text{Re}(1/T)]/(ka)^2$ for Eq. (42) has already been identified in Sec. IV D as a normalized absorption cross section for a single cavity.

The modulating factor $|(1+\Phi)/(1-\Omega T)|^2$ for Eq. (42) can also be expressed as

$$\left|\frac{1+E^2\sigma}{1-\Omega T-2\pi E^2\sigma T/(kd)^2}\right|^2, \tag{50}$$

by Eqs. (34) and (29). It is shown by curve (a) in Fig. 12. As already seen by comparing curves (b) and (a) in Fig. 11, the modulation causes loss magnification and the loss peak is shifted from 7.5 to 9.0 kHz. The incorrect modulation that would result from the single-scattering approximation, by setting Ω to zero in Eq. (50), is shown by curve (b) in Fig. 12.

V. CONCLUSIONS

Echo reduction by Alberich anechoic coatings can be studied by semianalytical wave theory, as reviewed in Sec. II. Scatterers in a periodic planar lattice redistribute normally incident plane-wave energy to reflected (and transmitted) beams, not only in the normal direction but in a discrete set of other directions. Almost all reflected energy in the example case appears in the normal direction (Sec. III, Fig. 2).

Moreover, the non-normal beams are evanescent at low frequency, but they are noticeable at the water/rubber interface in Fig. 3.

The example case shows that reflections can be eliminated with a rubber coating with thickness 6% of its compressional-wave wavelength, much less than a quarter of a wavelength, if spherical cavities are introduced and the rubber shear-wave parameters are specified appropriately. Computed absorption cross sections and energy flux (Figs. 4–6) show that the absorption loss is suffered mainly in the vicinity of the cavities. Effects of multiple scattering among the cavities are noted in the reflectance versus frequency curves (Fig. 2).

Allowing lower shear-wave velocity and higher shear-wave absorption for the rubber, vanishing reflectance can be achieved at lower frequencies, for still thinner coatings in relation to the wavelength. The limits are set by the realistically possible material parameters.

As outlined in Sec. IV, the main features of the anechoic effect, and the significance of multiple scattering among the cavities, can be understood by a simplified analysis in terms of normal plane waves and monopole scattering. A single cavity of radius $a=3$ mm has a monopole resonance[3,4] at about 8.8 kHz, if the absorption in the surrounding rubber with shear-wave velocity 83 m/s is ignored. A plane array of cavities in a quadratic lattice with period $d=44$ mm can then produce almost total reflection at about 10.0 kHz for a normally incident plane compressional wave (Fig. 9, $|\Theta|$ in the upper panel). For the coating with shear-wave absorption, most of the resonance scattering energy disappears by viscoelastic loss processes (Fig. 8, $|\Theta|$ in the upper panel), but an appropriate amount remains to enable vanishing reflectance at 9 kHz by destructive interference among upgoing normal compressional waves from the cavity and rubber/steel interfaces (Fig. 10). As stated in Ref. 6, the coating and the underlying structure must be considered together.

The shear-wave properties of the rubber material are important for generating loss from scattered spherically symmetric compressional waves. This loss arises from a restoring stress close to the spherical cavities, that is absent in a fluid. Eq. (42) expresses the absorption cross-sectional contribution by each cavity as its single-cavity absorption cross-section modulated by a factor that includes the effects of multiple scattering among the cavities.

A coating useful in practice should provide significant echo reduction within a broad frequency interval. Global optimization techniques, such as genetic or differential evolution algorithms, are better adapted to design such coatings[25] than the winding-number method of Sec. III. The semianalytical technique of Sec. II allows fast calculations, which is important. Additional computations indicate that inclusion of compressional-wave absorption can be useful to reduce the reflectance at higher frequencies. Coatings with mixed cavity sizes should be of interest, possibly with more than one scatterer interface. Different "local" periods can be allowed for the different scatterer interfaces, as long as there is an overall period d, which is a multiple of each of the local ones.[25] Nonspherical scatterers could be allowed by incorporating existing T-matrix routines for such objects.

Other applications of the theory reviewed in Sec. II include point-source computations by wave-number integration, and validation of homogenization or effective medium approaches.

[1] H. Oberst, "Resonant sound-absorbers," in *Technical Aspects of Sound*, edited by E. G. Richardson (Elsevier, Amsterdam, 1957), Chap. 7, pp. 287–327.

[2] G. Gaunaurd, "One-dimensional model for acoustic absorption in a viscoelastic medium containing short cylindrical cavities," J. Acoust. Soc. Am. **62**, 298–307 (1977).

[3] G. Gaunaurd, K. P. Scharnhorst, and H. Überall, "Giant monopole resonances in the scattering of waves from gas-filled spherical cavities and bubbles," J. Acoust. Soc. Am. **65**, 573–594 (1979).

[4] E. Meyer, K. Brendel, and K. Tamm, "Pulsation oscillations of cavities in rubber," J. Acoust. Soc. Am. **30**, 1116–1120 (1958).

[5] G. Gaunaurd and J. Barlow, "Matrix viscosity and cavity-size distribution effects on the dynamic effective properties of perforated elastomers," J. Acoust. Soc. Am. **75**, 23–34 (1984).

[6] H. Strifors and G. Gaunaurd, "Selective reflectivity of viscoelastically coated plates in water," J. Acoust. Soc. Am. **88**, 901–910 (1990).

[7] A. N. P. Cederholm and I. Karasalo, "Inverse modelling of rubber coated steel plates with a genetic algorithm," Acta. Acust. Acust. **88**, 694–698 (2002).

[8] A.-C. Hladky-Hennion and J.-N. Decarpigny, "Analysis of the scattering of a plane wave by a doubly periodic structure using the finite element method: application to Alberich anechoic coatings," J. Acoust. Soc. Am. **90**, 3356–3367 (1991).

[9] V. Easwaran and M. L. Munjal, "Analysis of reflection characteristics of a normal incidence plane wave on resonant sound absorbers: a finite element approach," J. Acoust. Soc. Am. **93**, 1308–1318 (1993).

[10] A.-C. Hladky-Hennion and J.-N. Decarpigny, "Note on the validity of using plane-wave type relations to characterize Alberich anechoic coatings," J. Acoust. Soc. Am. **92**, 2878–2882 (1992).

[11] Z. Liu, C. T. Chan, P. Sheng, A. L. Goertzen, and J. H. Page, "Elastic wave scattering by periodic structures of spherical objects: theory and experiment," Phys. Rev. B **62**, 2446–2457 (2000).

[12] I. E. Psarobas, N. Stefanou, and A. Modinos, "Scattering of elastic waves by periodic arrays of spherical bodies," Phys. Rev. B **62**, 278–291 (2000).

[13] F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).

[14] B. L. N. Kennett, *Seismic Wave Propagation in Stratified Media* (Cambridge Univ. Press, Cambridge, 1983).

[15] J. B. Pendry, *Low Energy Electron Diffraction* (Academic Press, London, 1974).

[16] J. A. Hudson, *The Excitation and Propagation of Elastic Waves* (Cambridge Univ. Press, Cambridge, 1980).

[17] K. Aki and P. Richards, *Quantitative Seismology* (Univ. Science Books, Sausalito, 2002).

[18] N. Stefanou, V. Yannopapas, and A. Modinos, "Multem2: A new version of the program for transmission and band-structure calculations of photonic crystals," Comput. Phys. Commun. **132**, 189–196 (2000).

[19] S. Ivansson and I. Karasalo, "Computation of modal wavenumbers using an adaptive winding-number integral method with error control," J. Sound Vib. **161**, 173–180 (1993).

[20] S. Ivansson, "Shear-wave induced transmission loss in a fluid-solid medium," J. Acoust. Soc. Am. **96**, 2870–2875 (1994). Correction: Replace 1/4 by 1/2 in Eqs. (18) and (21).

[21] A. D. Pierce, *Acoustics* (Acoustical Society of America, New York, 1991).

[22] K. Maslov and V. K. Kinra, "Acoustic response of a periodic layer of nearly rigid spherical inclusions in an elastic solid," J. Acoust. Soc. Am. **106**, 3081–3088 (1999).

[23] F. G. Blake, "Spherical wave propagation in solid media," J. Acoust. Soc. Am. **24**, 211–215 (1952).

[24] I. Tolstoy and A. Tolstoy, "Line and plane arrays of monopole scatterers," J. Acoust. Soc. Am. **87**, 1038–1043 (1990).

[25] S. Ivansson, "Riccati-method modelling of echo reduction by Alberich anechoic coatings," on the CD-ROM: *Proc. UDT Europe 2003 (Undersea Defence Technology).*

Approximations for the scattered field potential from higher mode transmission in rectangular apertures

J. L. Horner
Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, LE11 3TU, United Kingdom

K. S. Peat
Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough, LE11 3TU, United Kingdom

(Received 11 August 2005; revised 24 March 2006; accepted 27 March 2006)

This paper considers the wave fields that result when a plane wave impinges at an arbitrary angle on a rectangular aperture in a rigid, thick wall. A nondimensional form of a prior Fourier transform solution of this problem is derived, from which it is more easy to appreciate the relationship between the physical attributes of the aperture and incoming wave and the resultant acoustic fields. The scattered field from the aperture is examined in detail, in particular the modal contributions to the driving function for the amplitude of the velocity potential. Although the full scattered field contains both modal sum and modal coupling effects, it is shown that neglect of the modal coupling effects introduces minimal error to the solution in certain situations. An approximate analytical solution to the uncoupled analysis is then developed, which is accurate provided that the aperture is acoustically large, such that there are several cut-on modes within the aperture. The full nature of the scattered field can be easily interpreted from the nondimensional, analytical solution. In particular, an error analysis is given from which one can determine the required number of modes in a solution for given accuracy. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197792]

PACS number(s): 43.20.Fn, 43.20.El [LLT] Pages: 3568–3576

I. INTRODUCTION

A common problem in many engineering applications is the transfer of acoustic energy from one domain to another via an opening or aperture. Many such apertures are rectangular in cross section and have a finite depth. At low frequencies, or small Helmholtz numbers, when the wavelength of the incident wave is much greater than the aperture dimensions, well established approximate methods, such as that developed by Sauter and Soroka,[1] may be used to estimate the fields resulting from plane wave impingement on a rectangular aperture. When the wavelength of the incident field approaches the dimensions of the aperture, higher-order duct modes will propagate in the aperture and solutions must include coupling between these higher-order modes.[2] This is a numerically intensive activity that does not simply yield a solution. Although there is much previous work that evaluates the fully coupled problem over a wide frequency range, see, for example, Refs. 3 and 4, there is none that provides simple approximate solutions which include a contribution from propagating higher-order modes. The objective of this investigation is to determine a simple approximate solution for the scattered field from a rectangular aperture in a rigid baffle separating two semi-infinite half spaces when the frequency of excitation is such that higher-order modes freely propagate in the aperture.

Park and Eom[5] used an approach utilizing the Fourier transform and mode matching to obtain a rigorous solution to the fully coupled problem with improved numerical efficiency as compared with earlier methods. The approach of mode matching restricts the solution to aperture cross sections of separable geometries, such as rectangles. An effective numerical scheme was also introduced for the integral which determines the modal coupling coefficient. Davies[6] also formulated approximations for the modal coupling coefficients that utilized orthogonal properties of the integral. A similar approach for the coupling coefficient was used in work to establish simple approximations to properties of the scattered field for circular apertures in rigid baffles.[7] This previous work is now extended to consider plane wave impingement on rectangular apertures. As with the previous work, this investigation concentrates on determining the scattered field at the cut-on wave number of a higher-order mode. At these wave numbers the scattered wave amplitude will be greatest[8] for the mode of interest and thus render conservative predictions.

In the first instance, Sec. II, the earlier modal solution for the case of a rectangular aperture[5] is recast into a nondimensional format such that it is easier to understand which physical parameters affect the solution and the nature of that influence. The key parameters are found to be the aspect ratio and depth ratio of the aperture, together with the Helmholtz number of the incident plane wave. The nature of the solution is also explored, in particular to show that only a relatively small number of modes will make a significant contribution to a given solution. The next step, Sec. III, is to compare the differences between fully coupled calculations and the same calculations with the cross-coupling terms suppressed in the modal coupling coefficient matrix, the uncoupled solution. When this error is small, it is also possible

 0001-4966/2006/119(6)/3568/9/$22.50

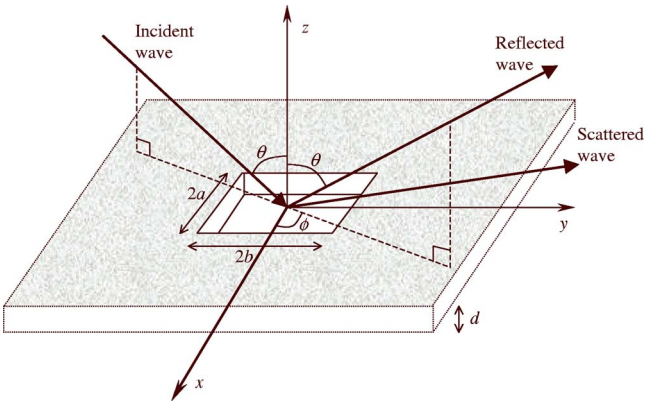


FIG. 1. (Color online) Sketch of aperture.

to treat the higher-order modes as approximate lumped parameter systems for use in extending the upper wave number limit of transmission line theory solutions.[9] It is shown that when the incident plane wave excites an aperture mode at its cut-on frequency, the error in using an uncoupled solution is insignificant.

From Sec. IV onward attention is focused on an approximate uncoupled solution. This is initiated by an analytical approximation to the self-coupling coefficients and results in an approximate analytical expression for all of the field potentials, although attention here is restricted to the scattered field. In Sec. V, various corollaries regarding the physical nature of the scattered field solution are drawn from the approximate solution. In particular, an error estimate for the solution with respect to the number of modes retained in the solution is derived. Finally, in Sec. VI, the overall accuracy of the approximate uncoupled solution with reference to the exact fully coupled solution is investigated for various example cases. Errors are shown to be consistently below 1%.

II. DESCRIPTION OF ACOUSTIC FIELDS

Consider a rectangular aperture of size $2a$ by $2b$ in a rigid wall of thickness d which is of infinite extent (Fig. 1). Let the aspect ratio be $A=a/b$ and the depth ratio be $D=d/a$. For simplicity the following analysis is undertaken in terms of velocity potentials, from which quantities such as impedance may be obtained. Following Ref. 5 the velocity potential for an incident wave of Helmholtz number ka and arbitrary impingement angles θ and ϕ, may be given by

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

where $k_x = k\cos\phi\sin\theta$, $k_y = k\sin\phi\sin\theta$, and $k_z = k\cos\theta$.

A time base of $e^{-i\omega t}$ is suppressed throughout for brevity. The resulting reflected and scattered velocity potentials are given by

$$\Phi^r(x,y,z) = e^{i(k_x x + k_y y + k_z z)}, \tag{2}$$

$$\Phi^s(x,y,z) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{\Phi}^s(\zeta a,\eta b)e^{-i(\zeta x+\eta y-\kappa z)}d(\zeta a)d(\eta b), \tag{3}$$

where the scattered Helmholtz number $(\kappa a) = \sqrt{(ka)^2-(\zeta a)^2-(\eta bA)^2}$ and $z>0$. ζ and η are the recoil wave numbers. Inside the duct the field is given over the depth d by

$$\Phi^d(x,y,z) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\left[c_{mn}\cos\xi_{mn}d\left(\frac{z}{d}+1\right) + d_{mn}\sin\xi_{mn}d\left(\frac{z}{d}+1\right)\right] \times\cos\frac{m\pi}{2}\left(\frac{x}{a}+1\right)\cos\frac{n\pi}{2}\left(\frac{y}{b}+1\right), \tag{4}$$

where $(\xi_{mn}d)=D(\xi_{mn}a)=D\sqrt{(ka)^2-(m\pi/2)^2-(n\pi A/2)^2}$ and $\xi_{mn}d$ is an in-aperture axial Helmholtz number.

Assuming a rigid wall of infinite extent surrounding the aperture, the velocity potential gradients may be summed for the aperture area and regions outside. Applying a Fourier transform to the two expressions determined from the summed velocity potential gradients, yields the following expression:[5]

$$\tilde{\Phi}^s(\zeta a,\eta b) = -i\frac{1}{(\kappa a)}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}(\xi_{mn}a)[c_{mn}\sin(\xi_{mn}d) - d_{mn}\cos(\xi_{mn}d)]G_m(\zeta a)G_n(\eta b), \tag{5}$$

where

$$G_m(u) = \frac{u[(-1)^m e^{iu} - e^{-iu}]}{u^2-(m\pi/2)^2}. \tag{6}$$

Equation (5) describes the amplitude of the scattered field, which depends on the recoil wave numbers in the x and y directions, ζ and η, respectively. As c_{mn} and d_{mn} are the amplitudes of the waves in the aperture, the term $G_m(\zeta a)G_n(\eta b)$ in Eq. (5) may be considered to propagate the scattered field. As it depends on both m and n, each mode will give a different value of this product.

It should be noted that c_{mn} and d_{mn} are determined by considering the transmitting side of the aperture and solving at all boundaries for continuity of velocity potentials and particle velocities. This will result in a set of simultaneous equations whose solution yields values for c_{mn} and d_{mn} that include modal contributions from all modes considered in the calculation. If attention is restricted to a finite number of significant modes, these coefficients are determined by the solution of a matrix equation of the general form[5]

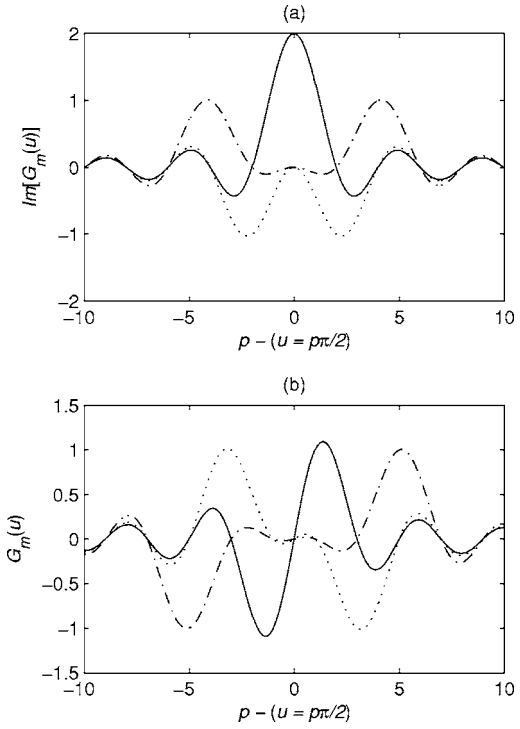


FIG. 2. Function $G_m(u)$. (a) —— $m=0$; ····· $m=2$; –·–·–·–·– $m=4$. (b) –––––– $m=1$; ····· $m=3$; –·–·–·–·– $m=5$.

$$\begin{bmatrix} \Psi_1 & \Psi_2 \\ \Psi_3 & \Psi_4 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \Gamma \\ 0 \end{bmatrix}, \tag{7}$$

where

$$\psi_{1,mnpq} = \varepsilon_p \varepsilon_q \cos(\xi_{pq} d)\, \delta_{mp} \delta_{nq} - \frac{i}{4\pi^2} I_{mnpq} \sin(\xi_{mn} d),$$

$$\psi_{2,mnpq} = \varepsilon_p \varepsilon_q \sin(\xi_{pq} d)\, \delta_{mp} \delta_{nq} + \frac{i}{4\pi^2} I_{mnpq} \cos(\xi_{mn} d),$$

$$\psi_{3,mnpq} = -\varepsilon_p \varepsilon_q \delta_{mp} \delta_{nq},$$

$$\psi_{4,mnpq} = \frac{i}{4\pi^2} I_{mnpq},$$

$$\gamma_{pq} = -2 G_p(k_x a) G_q(k_y b). \tag{8}$$

$\varepsilon_0=2$, $\varepsilon_1=\varepsilon_2=\cdots=1$, δ_{mp} is the Kronecker delta and the modal coupling coefficient is given by

$$I_{mnpq} = (\xi_{mn} a) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(\kappa a)} G_m(\zeta a) G_p(-\zeta a) G_n(\eta b) \times G_q(-\eta b) d(\zeta a) d(\eta b). \tag{9}$$

The column vector Γ of γ_{pq} terms on the right-hand side of Eq. (7) provides the driving terms for the wave amplitudes c_{mn} and d_{mn}. The behavior of $G_m(u)$ is shown in Fig. 2, where the peak value of $G_m(u)$ is seen to occur approximately at $G_m(m\pi/2)$. Note that $G_m(u)$ is wholly imaginary for even values of m and wholly real for odd values of m. It follows from Fig. 2 and Eq. (8) that it is possible to select values of k_x and k_y for a particular mode p,q (or m,n) that will result in a large amplitude of γ_{pq} for a given aperture. The largest magnitude of γ_{pq} and hence the wave amplitudes occurs approximately at cut on of the p,q mode,[8] or when $k_x a \approx p\pi/2$, $k_y b \approx q\pi/2$. It is therefore possible to "tune in" the wave number and direction of the incident planar wave to a cut-on wave number of one particular mode. It may also be observed from Fig. 2 that, for modes other than the cut-on mode, the amplitude of the driving term for those modes rapidly diminishes the further away they are from the cut-on wave number of interest. Thus relatively few modes make a significant contribution to the scattered field potential such that the infinite summations in Eq. (5) can be truncated drastically. Also, it may be observed that if a mode is excited at its cut-on wave number, the zeros of $G_m(u)$ occur at other modal wave numbers, further reducing the number of modes contributing to the overall summed field. In this context, it may be noted that the modal energy is proportional to the square of the product $G_m(u)G_n(u)$.

Solution of Eq. (7) will yield the complete solution to the overall problem in which there is coupling between all the modes considered. A much simpler solution may be obtained if coupling between modes is ignored, such that the off-diagonal terms in ψ_1 to ψ_4 are neglected, i.e., all modal cross-coupling coefficients are put to zero.

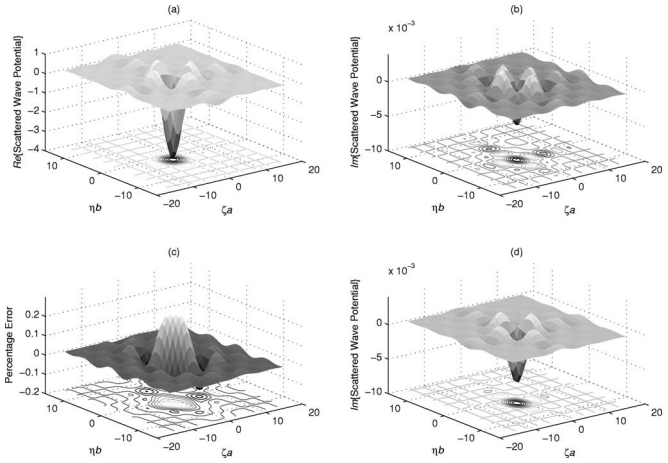


FIG. 3. Scattered field potential. (a) Coupled solution, real part. (b) Coupled solution, imaginary part. (c) Error in the uncoupled solution, real part. (d) Uncoupled solution, imaginary part.

III. COMPARISON OF COUPLED AND UNCOUPLED RESULTS

Figure 3 shows an example of the scattered field potential, Eq. (5), from the fully coupled solution. An aperture of aspect ratio $A=1.11$ and depth ratio $D=0.1$ was used. The Helmholtz number of the incident wave was $ka=50$ and 25 modes were used in the calculations. All 25 modes were cut on in the duct. Justification for the use of this number of modes is given in Sec. V. The incident plane wave was "tuned" to the cut-on wave number of the (1,1) mode. Similar figures were obtained by "tuning" into other modal wave numbers.

Figure 3(a) shows the real part of the scattered wave potential from the full, coupled solution. The corresponding imaginary part is shown in Fig. 3(b). In both cases the domi-

TABLE I. The maximum value of the scattered field potential for various cases and the corresponding error values from uncoupled and approximate solutions.

Driven mode (m,n)	Aspect ratio A	Coupled Re $\{\Phi^s\}$	Uncoupled error percent	Approximate error percent	Coupled Im $\{\Phi^s\}$
(1,0)	1.11	4.0046	0.2469	0.4322	0.0050
(1,1)	1.11	3.9711	0.2662	0.3695	0.0050
(2,0)	1.11	3.9710	0.2378	0.4115	0.0045
(2,1)	1.11	3.9379	0.3106	0.3520	0.0047
(2,2)	1.11	3.8618	0.3283	0.4209	0.0046
(1,0)	2	4.0119	0.2864	0.5207	0.0059
(1,1)	2	4.0123	0.3714	0.5870	0.0094
(2,0)	2	3.9782	0.2867	0.4982	0.0056
(2,1)	2	3.9789	0.3961	0.5725	0.0095
(2,2)	2	3.9436	0.4965	0.5093	0.0064
(1,0)	5.8	3.8398	0.5978	0.7602	0.0333
(1,1)	5.8	3.8249	0.5163	0.4347	0.0325
(2,0)	5.8	3.3225	0.5854	0.7170	0.0289
(2,1)	5.8	3.3097	0.5689	0.3922	0.0286
(2,2)	5.8	3.2878	0.8517	0.8142	0.0283

nant response in Helmholtz number space is centerd around $\zeta a=-\pi/2$, $\eta b=-\pi/2$, corresponding to the incident wave of $k_x a=\pi/2$, $k_y b=\pi/2$. The peak response actually occurs at the ζa, ηb values that correspond to the u value of the negative peak of the $G_1(u)$ curve, see Fig. 2(b). It may also be seen that the imaginary part is almost negligible in comparison to the real part, the peak response of the former being only 0.125% of the value of the latter. The small imaginary part represents a slight phase shift between the various wave components in the scattered field.

In an uncoupled solution, i.e., where $p=m$ and $q=n$, the c_{mn} and d_{mn} coefficients of Eq. (5) are determined from a form of Eq. (7) in which only the diagonal terms are nonzero, namely, the equations

$$\left[\varepsilon_m\varepsilon_n\cos(\xi_{mn}d)-\frac{i}{4\pi^2}I_{mnmn}\sin(\xi_{mn}d)\right]c_{mn}+\left[\varepsilon_m\varepsilon_n\sin(\xi_{mn}d)+\frac{i}{4\pi^2}I_{mnmn}\cos(\xi_{mn}d)\right]d_{mn}=-2G_m(k_x a)G_n(k_y b) \tag{10}$$

and

$$\varepsilon_m\varepsilon_n c_{mn}=\frac{i}{4\pi^2}I_{mnmn}d_{mn}. \tag{11}$$

Note that the modal coupling coefficient I_{mnmn} now only contains self-coupling contributions. Thus it is possible to approximate the uncoupled modal in-aperture amplitudes using Eqs. (10) and (11). These may be substituted into Eq. (5) to give the uncoupled scattered wave amplitude. It is then possible to determine the uncoupled field using the same parameters as for Fig. 3. The real part of the scattered wave potential from the uncoupled solution is so similar to that from the coupled solution that it appears exactly as the function in Fig. 3(a). To highlight the differences, the relative percentage error between the two sets of results is shown in Fig. 3(c). The relative error is defined here as the potential from the coupled solution minus that from the uncoupled solution divided by the maximum modulus value of the field potential from the coupled solution. Note the clear lines of maximum error. It may be seen that the difference between the coupled and uncoupled solutions is greatest at the Helmholtz number that corresponds to the peak response but is nevertheless very small, with a maximum error of 0.27%. Similar results may be obtained for other aperture dimensions and modal wave numbers provided the cut-on wave numbers are below the incident wave number, as discussed further below.

There is visible discrepancy in the values of the imaginary part of the scattered wave potential from the coupled and uncoupled solutions. The latter is shown in Fig. 3(d) and should be compared with Fig. 3(b). However, as noted earlier, the imaginary part of the potential is negligible as compared to the real part. For the uncoupled solution, the peak response of the former is only 0.174% of the value of the latter, thus the visible discrepancies between the imaginary part of the coupled and uncoupled solutions in this free-field situation are of no practical concern and may be considered as trivial.

Further examples of comparative results from the coupled and uncoupled solutions for various aperture aspect ratios and driven modes are shown in Table I. All these results relate to an aperture depth ratio of $D=0.1$ and a total of 25 modes, all of which are cut on in the aperture. The Helmholtz number of the incoming wave is $ka=50$ for the aspect ratios of 1.1 and 2, $ka=150$ for aspect ratio 5.8. The increase in the latter case was required to ensure all 25 modes are cut on. For each aspect ratio and driven mode, Table I shows the modulus of the maximum values of the real and imaginary parts of the scattered wave potential from the coupled solution and the maximum percentage error between the real values from coupled and uncoupled solutions. It should be noted that the maximum value quoted is only the maximum over the set of grid points ζa, ηb at which the function was evaluated and is unlikely to be the absolute maximum value

of the function. Increasing the grid point resolution increases the computational effort required to determine a solution but of course it does not affect the accuracy of solution at a given grid point. The column of values for error from the approximate solution will be commented upon later.

The results imply that it is possible to evaluate the scattered field potential from an uncoupled solution with very good accuracy, less than 1% error in all cases. With the use of only 25 modes, the fully coupled analysis demands solution of a 50×50 set of simultaneous complex equations. This requirement is eliminated in the uncoupled solution. However the major computational effort in the coupled solution is in fact due to the numerical evaluation of the modal coupling coefficient I_{mnpq} integrals. The major reduction in computational effort between the coupled and uncoupled solutions is due to the fact that the number of such integral evaluations is reduced from 25^2 to 25 for a 25 mode calculation, i.e., 4% of the fully coupled solution evaluations.

IV. APPROXIMATION TO THE UNCOUPLED SOLUTION

Although an uncoupled solution clearly represents an enormous saving in terms of computational time and effort in comparison to a fully coupled solution, the calculation of even the diagonal coefficients is still a task that requires significant computational effort.[5] As noted above, this is principally due to the evaluation of the I_{mnmn} integrals, see Eq. (9), for the diagonal coefficients of Eq. (7). However, an approximate analytical solution to the complete problem can be found, as is shown below. This results in further large savings in terms of computational time and effort as well as giving much more insight into the nature of the solution and the effect of parameters.

Central to the approximation are the assumptions that the aperture is acoustically large, such that there are many cut-on modes, and that the (p,q) mode being driven at cut on by the incident wave is of low order. As noted earlier, only modes m,n that are close to the driven mode p,q will contribute significantly to the scattered field potential. It is assumed that all of the modes that make a significant contribution to the potential, and so must be included in the solution, are cut on. A sufficient condition for the approximation to be valid is therefore $(ka)^2>(m\pi/2)^2+(n\pi A/2)^2$ for all modes m,n included in the solution. The number of modes required for a given accuracy of solution will be discussed in Sec. V.

Assume for the moment that the Helmholtz number of the scattered wave κa is essentially constant, then Eq. (9) for the uncoupled modes can be simplified to

$$I_{mnmn}\approx\frac{(\xi_{mn}a)}{(\kappa a)}\int_{-\infty}^{\infty}G_m(\zeta a)G_m(-\zeta a)d(\zeta a)\times\int_{-\infty}^{\infty}G_n(\eta b)G_n(-\eta b)d(\eta b). \tag{12}$$

The integrands $G_m(u)G_m(-u)$ for various m are shown in Fig. 4, where it is seen that the curves are dominated by the peak product terms at $u=\pm m\pi/2$. This is a similar observation to Davies.[6] Since the principal contribution to the integral of

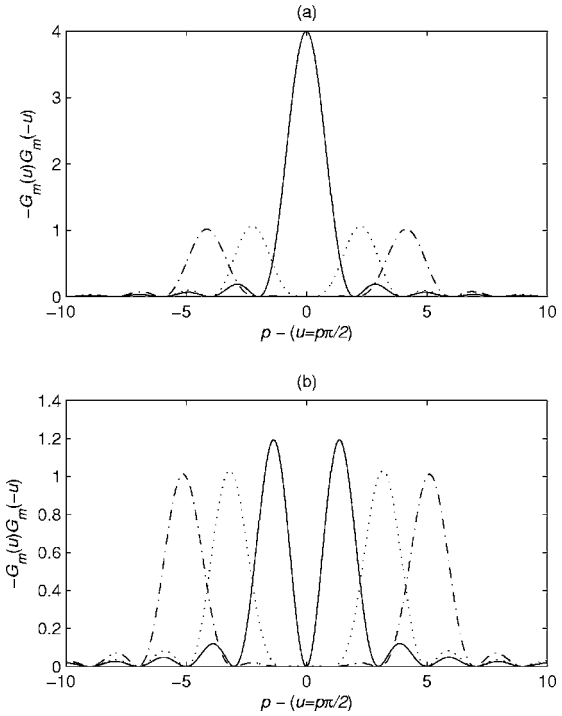


FIG. 4. Function $-G_m(u)\ G_m(-u)$. (a) —— $m=0$; ····· $m=2$; –·–·– $m=4$. (b) —— $m=1$; ····· $m=3$; –·–·–·–·– $m=5$.

Eq. (9) comes from a small element of the total range of integration, the initial assumption that κa is constant is reasonable. If one uses for κa the value relevant to the center of the principal peaks of $G_m(\zeta a)$ and $G_n(\eta b)$, then $\kappa=\xi_{mn}$. Thus[10]

$$I_{0000}\approx\left(\int_{-\infty}^{\infty}G_0(u)G_0(-u)du\right)^2=\left(8\int_0^{\infty}\frac{\sin^2 u}{u^2}du\right)^2=16\pi^2. \tag{13}$$

Since I_{0000} is the self-coupling coefficient of the plane wave, it is the dominant coupling coefficient. Now the peaks of the function $G_p(u)G_p(-u)$ have essentially the same form for all p, see Fig. 4. When $p=0$ there is a single peak of magnitude 4 and base width $\Delta u=2\pi$. For all $p>1$, there are two symmetrical peaks each of magnitude 1 and base width $\Delta u=2\pi$. Since, relative to the $p=0$ case, the magnitude has decreased by a factor of 4 but there are now two peaks instead of one of the same shape, the net integral contribution has decreased by a factor of 2. For the case $p=1$, there are two symmetrical peaks each of magnitude ≈1.2 and base width $\Delta u=3\pi/2$. Since, relative to the $p=0$ case, the magnitude has decreased by a factor $\approx4/1.2$ and the base width by a factor of 4/3, but there are now two peaks instead of one of the same shape, the net integral contribution has decreased by a factor ≈2. These observations, together with Eq. (13), imply that the integral approximation and self-coupling coefficient for any mode m,n is

$$I_{mnmn}\approx4\pi^2\varepsilon_m\varepsilon_n. \tag{14}$$

This may be interpreted that the plane wave coefficient is halved to obtain an axial mode coefficient and quartered to

TABLE II. Percentage error between Re(I_{0000}) and approximation to I_{0000}, for apertures of various aspect ratio, $ka=50$.

Aspect ratio A	Re$\{I_{0000}\}$	Im$\{I_{0000}\}$	Percent error in approx Re$\{I_{0000}\}$
1	157.78	−2.19	0.087
2	157.73	−2.57	0.12
4	157.35	−4.10	0.36
8	156.25	−6.76	1.07
16	153.60	−10.62	2.81

obtain a tangential mode coefficient. Table II gives exact values of Re$\{I_{0000}\}$ and Im$\{I_{0000}\}$, together with the error in the approximate values of I_{0000} as given by Eq. (14) relative to the exact Re$\{I_{0000}\}$, for apertures of various aspect ratio. The errors are found to be less than 1% for apertures of moderate aspect ratio. As the aperture starts to geometrically resemble a slit, the errors in approximating the modal coupling coefficients increase. The approximate expression for the integral contribution results in a zero imaginary part, whereas the exact integral evaluation is complex. It follows from Eq. (9) that the imaginary contribution to the integral arises from integration over the acoustic shadow region of the ζ-η plane, where κa is imaginary. The magnitude of the imaginary part is seen to be much less than that of the real part, as a consequence of the restriction to cases for which $(ka)^2>(m\pi/2)^2+(n\pi A/2)^2$. However, it may be observed that the relative error for the approximate integral expression due to neglect of the imaginary component is slightly higher than that due to error in the real component.

Table III shows results of the error in the approximate evaluation of I_{mnmn}, relative to the exact Re$\{I_{\mathrm{mnmn}}\}$, for various m and n and apertures of aspect ratio 2 and 8, respectively. For the smaller aspect ratio, the errors are consistently below 0.5%. In both cases, there is no observable pattern to the variation in the error and no consistent tendency for the error to increase with increasing mode number. In Table III(b), values are not given for $n>3$ since these violate the fundamental assumption that $(ka)^2>(m\pi/2)^2+(n\pi A/2)^2$ when $ka=50$, as used here. In this large aspect ratio case, for $n=3$, the error in the real part is sometimes more than 4% and once again the error due to neglect of the imaginary part is greater still. It may be inferred that, once again, the error increases as the aspect ratio of the aperture increases, but in reality the error seen here is connected with the necessary condition that $(ka)^2>(m\pi/2)^2+(n\pi A/2)^2$. The closer to zero this condition becomes for a given mode, the greater the possible error.

TABLE III. Percentage error between Re(I_{mnmn}) and approximation to I_{mnmn}. (a) Aspect ratio $A=2$, $ka=50$; (b) aspect ratio $A=8$, $ka=50$.

				(a)		
m/n	0	1	2	3	4	5
0	0.12	0.10	0.10	0.06	0.04	0.02
1	0.09	0.18	0.21	0.03	0.08	0.14
2	0.16	0.28	0.06	0.20	0.07	0.04
3	0.17	0.13	0.37	0.03	0.24	0.21
4	0.22	0.21	0.15	0.12	0.01	0.05
5	0.06	0.30	0.16	0.12	0.01	0.07
				(b)		
0	1.07	1.12	1.03	1.05	0.92	0.90
1	0.67	0.66	0.42	0.09	0.74	1.84
2	3.12	3.09	3.03	2.88	2.69	2.39
3	4.28	4.23	3.37	2.28	0.52	4.07

Now substitution from Eq. (14) into Eq. (11) gives

$$c_{mn} \approx id_{mn} \tag{15}$$

and then it follows from Eq. (10) that

$$c_{mn} \approx -\frac{G_m(p\pi/2)G_n(q\pi/2)}{\varepsilon_m\varepsilon_n[\cos(\xi_{mn}d) - i\sin(\xi_{mn}d)]}, \tag{16}$$

where $k_x a=p\pi/2$, $k_y b=q\pi/2$. For integer values of p,q this implies that the incident plane wave is "tuned" to the cut-on wave number of the (p,q) mode. Analytical solutions for the in-duct wave potential Eq. (4) and the transmitted field potential then follow. Attention here is restricted to the scattered wave potential, Eq. (5), which can now be written as

$$\tilde{\Phi}^s(\zeta a,\eta b) \approx -\frac{1}{(\kappa a)}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}(\xi_{mn}a)G_m(p\pi/2)G_m(\zeta a) \times G_n(q\pi/2)G_n(\eta b). \tag{17}$$

V. OBSERVATIONS FROM THE APPROXIMATE SOLUTION

In the first instance, a relationship is derived between the number of modes used in the analysis and the accuracy of the solution. Now the values of $G_m(p\pi/2)$, i.e., the function evaluated at modal wave numbers, follow from Eq. (6) and may be summarized as follows:

At coincidence, $m=p$

$$G_p\left(\frac{p\pi}{2}\right) = \varepsilon_p i(-i)^p, \tag{18}$$

otherwise, $m\neq p$,

$$G_m\left(\frac{p\pi}{2}\right) = \begin{cases} 0, & m \text{ and } p \text{ even or } m \text{ and } p \text{ odd} \\ \dfrac{-(-i)^p 4p}{\pi(p^2-m^2)}, & \text{otherwise.} \end{cases} \tag{19}$$

The most significant modal contribution to the summations in Eq. (17) is that from the coincidence mode, namely, $m=p$, $n=q$. Furthermore the maximum response in the scattered field is generated at Helmholtz numbers near $\zeta a=\pm p\pi/2$, $\eta b=\pm q\pi/2$. At these positions, the relative contribution to the solution from general mode (m,n), relative to that from the coincidence mode (p,q) is

$$\left|\frac{G_m(p\pi/2)G_m(\pm p\pi/2)}{G_p(p\pi/2)G_p(\pm p\pi/2)}\right|\left|\frac{G_n(q\pi/2)G_n(\pm q\pi/2)}{G_q(q\pi/2)G_q(\pm q\pi/2)}\right| < = \left(\frac{4p}{\varepsilon_p\pi(p^2-m^2)}\right)^2\left(\frac{4q}{\varepsilon_q\pi(q^2-n^2)}\right)^2, \quad m\neq p,\ n\neq q. \tag{20}$$

For the example given in Sec. III, $p=q=1$, thus since 25 modes were used, error due to modal truncation is principally due to neglect of modes with $m=6$ or $n=6$. Thus the relative error is of magnitude $\{4/[\pi(1-36)]\}^2=0.0013$, or 0.13%. Equation (20) can, of course, be used to determine the number of modes (m,n) required to achieve a solution to a specified error level. The lower limit on the Helmholtz number of incoming waves for which the solution is valid then follows from the sufficient condition of Sec. III, namely, $(ka)^2>(m\pi/2)^2+(n\pi A/2)^2$.

An explanation is now given for the pattern of the peak response of the scattered field potential in the recoil wave number plane. It was observed from Fig. 3, for which the incident wave is the (1,1) mode $k_xa=\pi/2$, $k_yb=\pi/2$, that the maximum response in the scattered field is near $\zeta a=-\pi/2$, $\eta b=-\pi/2$. Attention to the signs of the wave numbers in Eqs. (2) and (3) then implies that the part of the scattered wave that has maximum amplitude is that which scatters along the direction of the reflected wave, as might be expected from physical intuition.[2] There is no such similar peak of response in other Helmholtz number pairs from the set $\zeta a=\pm\pi/2$, $\eta b=\pm\pi/2$, as might have been expected from Eq. (17). It is shown next that is always the case, whatever mode (p,q) the incoming wave is tuned to. It follows from Eq. (17) that

$$\tilde{\Phi}^s(\pm p\pi/2,\pm q\pi/2)\approx -\sum_{m=0}^{\infty}G_m(p\pi/2)G_m(\pm p\pi/2)\times\sum_{n=0}^{\infty}G_n(q\pi/2)G_n(\pm q\pi/2). \tag{21}$$

Now from Eqs. (18) and (19)

$$\sum_{m=0}^{\infty}G_m(p\pi/2)G_m(\pm p\pi/2)=-\varepsilon_p^2\pm\sum_{m=0}^{\infty}\left[\frac{4p}{\pi(p^2-m^2)}\right]^2, \tag{22}$$

where the summation is taken only over odd or even values of m according to whether p is even or odd. To be precise, when $\zeta a=+p\pi/2$ the entire expression on the right-hand side of Eq. (22) is multiplied by $(-1)^p$, but the key issue is the $\pm$ sign within the expression. This sign determines whether contributions from the noncoincidence modes are in phase or out of phase with the coincidence mode, i.e., whether they interfere constructively or destructively. It is seen that constructive interference occurs only at $\zeta a=-\pi/2$ and, by similarity, $\eta b=-\pi/2$.

Finally, it was observed from Fig. 2 that, for $m>0$, the $G_m(u)$ functions have peak magnitudes of approximately 1 close to, but not precisely at, the locations $u=\pm m\pi/2$. Similarly, the peak response in the scattered field for an incoming wave $k_xa=p\pi/2$, $k_yb=q\pi/2$, for example, Fig. 3, is found to be close to $\zeta a=-p\pi/2$, $\eta b=-q\pi/2$, but actually occurs at the location of the corresponding coincidence peaks of the $G_p(u)$, $G_q(u)$ functions. The peak values and locations of these functions may be found by differentiation of Eq. (6). To a first-order approximation

$$|G_m(u)|_{\max}\approx 1+\frac{3}{2m^2\pi^2}\quad \text{at } u\approx\pm\left(\frac{m\pi}{2}+\frac{3}{m\pi}\right). \tag{23}$$

For $m\geqslant 2$, the error in value is less than 0.6% and the error in location is less than 2.4%. However, these values rise to 5.5 and 17.4%, respectively, when $m=1$, but a second-order approximation for $m=1$ gives

$$|G_1(u)|_{\max}\approx 1+\frac{\varepsilon}{\pi}-\varepsilon^2\left(\frac{1}{6}+\frac{1}{\pi^2}\right)\quad \text{at } u\approx\pm\left(\frac{\pi}{2}+\varepsilon\right),\quad \text{where } \varepsilon=\frac{\pi}{7}\left[\sqrt{\frac{42}{\pi^2}+1}-1\right], \tag{24}$$

for which the errors in value and location are 0.22 and 0.05 %, respectively. Thus the precise location of the peak value of the scattered potential in the recoil wave number can be predicted quite accurately. Furthermore, since the driving terms for the wave amplitudes c_{mn} and d_{mn} are of the form $G_p(k_xa)G_q(k_yb)$, see Eqs. (7) and (8), it follows that the maximum response in the scattered, in-duct, and transmitted fields will actually occur when the incoming wave is tuned such that $k_xa=(p\pi/2)+(3/p\pi)$, $k_yb=(q\pi/2)+(3/q\pi)$, rather than precisely to cut on of the in-duct mode, as was done for Fig. 3.

VI. RESULTS FROM THE APPROXIMATE SOLUTION

The approximate solution for the scattered field potential follows from Eq. (17) for any incident plane wave where $k_xa=p\pi/2$, $k_yb=q\pi/2$. As noted earlier, integer values of p,q would imply that the incident plane wave is "tuned" to the cut-on wave number of the (p,q) mode. Figure 5(a) shows the approximate solution from Eq. (17) for exactly the same aperture, incident wave, and mode numbers as used in Fig. 3, i.e., $p=q=1$. No observable difference is apparent. Figure 5(b) shows the percentage error between this approximate solution and the exact fully coupled solution. The maximum error is less than 0.37%.

Figure 6(a) shows the scattered field potential from the fully coupled solution, Eq. (5), for an aperture of aspect ratio $A=2$ and depth ratio $D=0.1$. The incident wave has Helmholtz $ka=50$ and was selected to drive the (2,1) mode, i.e., $p=2$, $q=1$. A plot of the corresponding field from the approximate solution, Eq. (17), shows no observable difference. Thus the error between the approximate uncoupled and full coupled solutions is shown in Fig. 6(b). Similar results are shown in Fig. 7 for an aperture of aspect ratio $A=5.8$, incoming wave $ka=150$ driving the (2,0) mode. In both cases definite lines of error may be observed, but of a magnitude that is inconsequential.

The column labeled "approximate error" in Table I gives the maximum error in value of potential over the scattered field between the fully coupled solution and the approximate

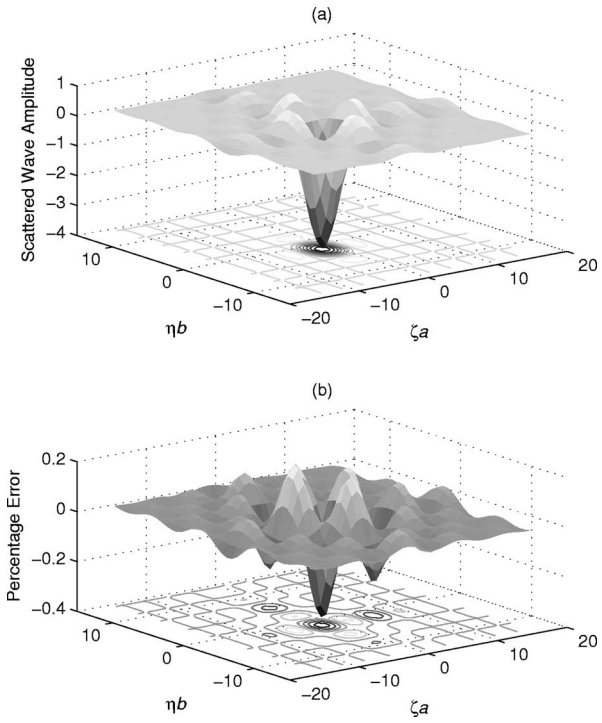


FIG. 5. Scattered field potential for an aperture of $A=1.11$, driving (1,1) mode. (a) Fully coupled solution. (b) Error using the approximate uncoupled solution.

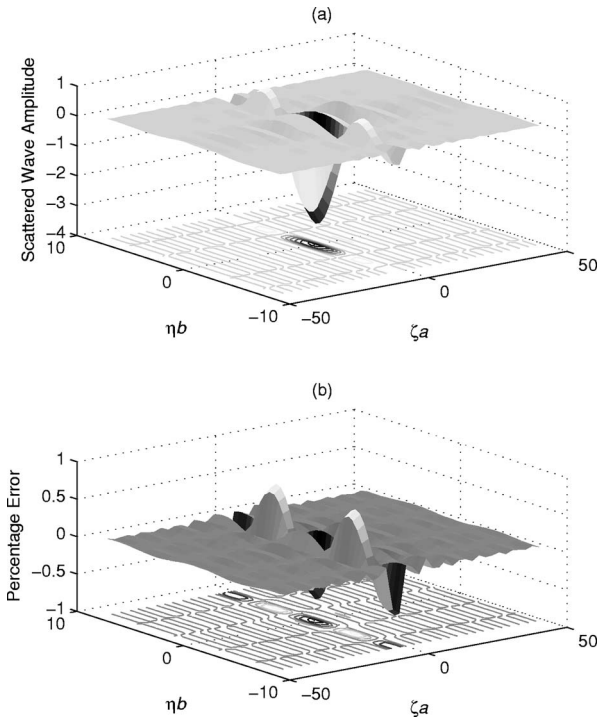


FIG. 7. Scattered field potential for an aperture of $A=5$, driving (2,0) mode. (a) Fully coupled solution. (b) Error using the approximate uncoupled solution.

uncoupled solution as given by Eq. (17). The errors are still below 1% and are little different to the errors due to use of an uncoupled solution. The error due to uncoupling is seen to increase slightly with increasing aspect ratio, but the error due to the approximation shows no such trend. In the cases shown in this table for which the aspect ratio has an integer value, the Helmholtz number of the incoming wave is such that it drives two different modes. In other cases there is a second mode very close to the driven mode. In neither case is the error due to uncoupling or the approximate solution found to be affected.

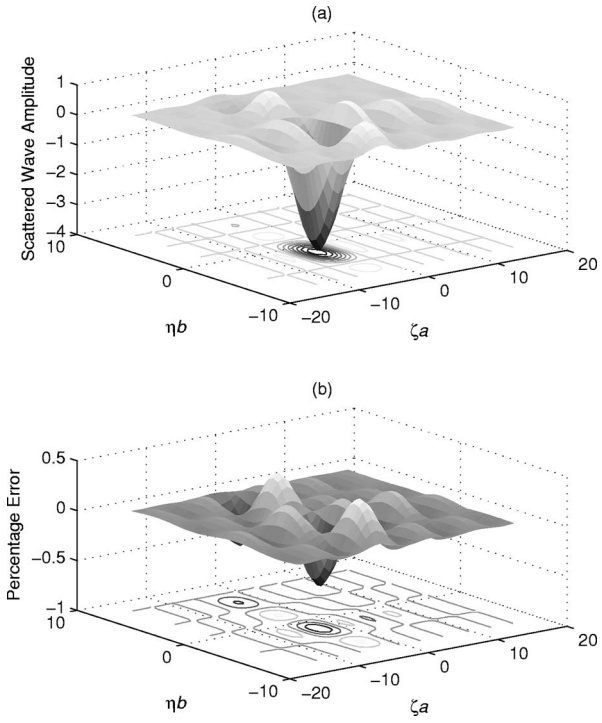


FIG. 6. Scattered field potential for an aperture of $A=2$, driving (2,1) mode. (a) Fully coupled solution. (b) Error using the approximate uncoupled solution.

For results of comparable accuracy, the approximate solution requires 2% of the computing effort for a fully coupled solution. However, the fully coupled results given here were evaluated to an accuracy at which their numerical error was insignificant as compared to the differences between the approximate and the fully coupled results, in order to assess the error of the former. In this context, the approximate solution requires only 0.003% of the computing effort of the coupled solution.

A viable alternative form of solution to the problem addressed in this paper is to use a boundary integral formulation, see, for example, Refs. 11 and 12. The computing effort required for given accuracy from this method is of the same order as that for the fully coupled modal solution used here. However the modal solution affords a level of insight into the physical characteristics of the problem that is lost in the boundary integral method. The approximate modal solution given in this paper increases this advantage much further since the physical attributes of the problem are readily visible in the solution and scaling laws can then be readily derived.

VII. CONCLUSIONS

Investigations have been carried out to compare various solutions to the acoustic potential in the scattered field from an impinging plane wave upon a rectangular aperture with

the objective of developing simple approximate solutions. A nondimensional form of the analysis has indicated that only the aspect ratio and depth ratio of the aperture enter the solution, and for the scattered field the effect of the latter is minor. It has been shown that, although the full solution includes both modal summation and cross-coupling effects, the neglect of coupling results in minimal error, less than 1%, provided the wave number of the incident wave exceeds that of the mode of primary interest. An approximate analytical solution to the uncoupled equations has been developed which is valid for acoustically large apertures, in which there are several cut-on modes. The extra error due to this approximation, over and above that due to neglect of coupling, is minimal. Thus the maximum error in the scattered field potential from the approximate analytical solution is also less than 1%. The nondimensional, analytical solution allows great insight into the nature of the scattered field and the factors that influence it, as well as providing an extremely simple form of solution. Formulas are developed to give the required number of modes for a given accuracy of solution and to determine the location in the scattered field of the maximum response as well as form of incoming wave that will lead to maximum response amplitudes.

[1]A. Sauter, Jr. and W. W. Soroka, "Sound transmission through rectangular slots of finite depth between reverberent rooms," J. Acoust. Soc. Am. **47**(1), 5–11 (1970).

[2]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986), Chap 9.

[3]K. Honga and H. Serizawa, "Diffraction of an acoustic plane wave by a rectangular hole in a infinitely large rigid screen," J. Acoust. Soc. Am. **106**, 29–35 (1999).

[4]A. D. Pierce, R. O. Cleveland, and M. Zampoli, "Radiation impedance matrices for rectangular interfaces within rigid baffles: Calculation methodology and applications, "J. Acoust. Soc. Am. **111**(2), 672–684 (2002).

[5]H. H. Park and H. J. Eom, "Acoustic scattering from a rectangular aperture in a thick hard screen," J. Acoust. Soc. Am. **101**(1), 595–598 (1997).

[6]H. G. Davies, "Low frequency random excitation of water-loaded rectangular plates," J. Sound Vib. **15**(1), 107–126 (1971).

[7]J. L. Horner, R. Lyons, and B. A. T. Petersson, "An approximate method for determining the maximum amplitude of higher-order duct modes," J. Acoust. Soc. Am. **110**(1), 80–85 (2001).

[8]R. T Muehleisen and D. C Swanson, "Modal coupling in acoustic waveguides: Planner discontinuities," Appl. Acoust. **63**, 375–1392 (2002).

[9]R. Lyons, P. C. Macey, and J. L. Horner, "Acoustic performance of finite length apertures using finite element analysis," in *Collected Papers Proc. 137th Meeting of the Acoustical Society of America* (Technische Universitat Berlin, 1999), CD-ROM ISBN 3-9804568-5-4.

[10]I. S. Gradshteyn and I. M. Ryzhik, in *Table of Integrals, Series and Products*, 5th ed., edited by A. Jeffrey (Academic Press, London, 1994).

[11]A. F. Seybert, C. Y. R. Cheng, and H. C. Wang, "The solution of coupled interior/exterior acoustic problems using the boundary element method," J. Acoust. Soc. Am. **88**(3), 1612–1618 (1990).

[12]G. W. Emms and C. Fox, "Control of sound transmission through an aperture using active sound absorption techniques: a theoretical investigation," Appl. Acoust. **62**, 735–747 (2001).

Stochastic inversion in acoustic scattering

Beatrice Faverjon and Roger Ghanem[a)]
Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, California 90089

(Received 19 July 2004; revised 4 April 2006; accepted 5 April 2006)

A novel approach is presented for the probabilistic characterization of the solution in inverse acoustic scattering problems. The approach consists of representing the unknown parameters of the scatterer as stochastic processes with spatially random fluctuations. The stochastic processes are represented by their polynomial chaos decomposition. The deterministic coefficients in this parametrization are then computed by solving a deterministic optimization algorithm with an appropriate objective function that captures the probabilistic significance of these coefficients. The general procedure is demonstrated by its application to a two-dimensional problem with the hidden object, a hard obstacle, being immersed in a heterogeneous medium. The formulation provides, as well, a measure of validity of a particular predictive model at matching experimental evidence.
 [DOI: 10.1121/1.2200149]

PACS number(s): 43.20.Fn, 43.30.Hw, 43.60.Pt, 43.30.Pc [PEB] Pages: 3577–3588

I. INTRODUCTION

Inverse wave scattering deals with the deconvolution of a target signature from a measured data. The signature can be used to reconstruct the shape, elasticity, or some other property of a hidden object. The measured data, generally, carries with it, in addition to the target signature, measurement noise and the signature of all it has interacted with along its path. These interactions include the exact shape of the object, its exact elasticity, other unknown hidden objects, as well as the real propagation medium including its naturally fluctuating properties and any hidden structure it may possess. These fluctuations and structure are often associated with interactions of the wave motion with other physical processes such as thermal processes and boundary interfaces. Moreover, whether the signal is a far field or a near field signal, it is usually sampled over a limited aperture, further challenging the information extraction task. Uncertainty in the inverse problem can thus be associated with a number of sources ranging from the selection of relevant physics and associated mathematical models, to measurement errors, or the design of experiments for collecting the data used in the inversion. Clearly, while some of these uncertainties can be, in the limit, reduced through analysis or data collection, many of them are intrinsic to the problem and are tied to an underlying natural variability in the propagation medium.

In an attempt to mitigate the effect of this ubiquitous uncertainty, methods of inverse analysis and optimization are typically relied upon to produce estimates of parameters or characterizations of hidden objects that are most consistent, in some meaningful sense, with observations. Although this approach is quite useful, it does not utilize all the information about the physical environment, and in particular, the possibility of a heterogeneous environment with spatially varying random fluctuations. Nor does the standard optimization approach provide the means of managing the uncertainty by assessing its sensitivity to various physical elements or its impact on the confidence in various predictive statements about related problems. It should be noted here that significant effort has been expanded into developing probabilistic models for this uncertainty that could be of great use when coupled with inversion algorithms. Many of these probabilistic models are specifically adapted to problems of wave propagation and scattering.[1–5] The present article describes a novel methodology that integrates probabilistic models, specifically those based on polynomial chaos representations,[6–8] with optimization methods, thus permitting the probabilistic characterization of signatures carried with the scattered wave. This can be used for the calibration of probabilistic predictive models or for the probabilistic characterization of hidden obstacles and inclusions in an elastic or acoustic medium. The method assumes the availability of a deterministic predictive model which represents and numerically resolves the assumed prevalent physics.

The proposed method is applied, in this article, to the problem of inverse acoustic scattering. A number of approaches have been proposed and pursued over the past several decades for representing and solving the associated equations of mathematical physics. Due attention has also been devoted to the task of inverting the acoustic signatures to identify and locate underwater obstacles.[9–16] Most of these models rely on integral formulations of the scattering problem, thus requiring significant computational resources, both in terms of computing time as well as in terms of specialized algorithms, for their numerical resolution. In this article, and in order to focus attention on the new developments, the scattered acoustic field is estimated using an analytical method, which is not directly based on integral representations but which, rather, is based on Rayleigh's hypothesis[17–24] and invokes the shape differentiation of the scattered field. It was recently used as the forward solution engine at the core of an optimization algorithm for treating the inverse acoustic scattering problem.[25–29]

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

The general method introduced in this article relies on the representation of stochastic processes and variables in terms of an orthogonal basis set of random variables. This so-called *polynomial chaos* basis permits the representation of random variables in a mean-square convergent expansion in terms of multidimensional Hermite polynomials of normalized Gaussian variables. The inverse problem consists then of estimating the value of the deterministic coefficients in this expansion. The resulting representation then provides a complete probabilistic characterization of the target random variables or stochastic processes. The measurements used in the inverse analysis are obtained through computer simulations. The effect of heterogeneities and other unmodeled physical phenomena on these simulated measurements is represented as a multiplicative and additive noise to the analytical field.

In this article, the probabilistic characterization of the scatterer reflects the uncertainty in its estimated value. In the present context, this uncertainty is associated with the inability of the chosen predictive model to reproduce the far field signals, even in the limit with ever increasing number of observation points. This would typically be due to unmodeled physics and dynamics, which in the present case is lumped into multiplicative and additive noise to the measured signal. Clearly, a more realistic measurement noise could be simulated by solving the wave propagation problem in a heterogeneous medium. The present choice of measurement model, however, serves well to highlight the features of the proposed methodology. Thus, although the original problem does not have any probabilistic component associated with it, the lack of knowledge and information is represented in a probabilistic framework. It can be expected that, as the complexity of the predictive model approaches that of the actual physical phenomenon, the magnitude of the probabilistic scatter in the estimated shape decreases, eventually reaching its deterministic limit. It should be noted that in some situations, enough information on the propagation medium may be available extraneously so as to permit a probabilistic model of the medium itself. The present approach can be readily generalized to address those situations, with the main difficulty being the availability of a predictive tool for the propagation of waves in a spatially varying medium.

Section II presents the acoustic scattering problem for an obstacle in an infinite medium. The forward problem is solved for the particular case of a two-dimensional scatterer with the analytical solution being detailed for the scattered acoustic field associated with Dirichlet boundary conditions. Section III presents the stochastic parameterization of the shape of the scatterer and the resulting stochastic model of the scattered field. From a stochastic model of the scattered field and from the measured field, the optimization problem is solved in a statistical sense in Sec. IV. Section V presents, for several forms and magnitudes of noise, numerical results for the scattered acoustic field and for the random shape.

II. SOLUTION OF THE SCATTERING PROBLEM

A. The forward problem

Consider an obstacle scattering in an infinite medium. Without any loss of generality, we will restrict our attention to scattering in two space dimensions. Denote the domain occupied by the obstacle by Ω and its boundary by $\partial\Omega$. Then the infinite exterior domain is denoted by $\mathbb{R}^2\backslash\bar{\Omega}$, and $\bar{\Omega}=\Omega\cup\partial\Omega$. It will be assumed that $\partial\Omega$ can be well approximated with $\partial\Omega_\lambda$ which is obtained by slightly deforming a body with a simple geometry having a surface $\partial\Omega_0$ such that for all $\mathbf{x}\in\partial\Omega$, $\mathbf{x}=\mathbf{x}_0+\delta\Omega_\lambda(\mathbf{x})$, where $\mathbf{x}_0\in\partial\Omega_0$ and $\delta\Omega_\lambda(\mathbf{x})$ is an infinitesimal deformation specified in the form $\delta\Omega_\lambda(\mathbf{x})=\lambda f(\theta)$.[25–29] Here, $f(\theta)$ denotes a functional form describing the variation of shape away from $\partial\Omega_0$, θ the angular coordinates of point $\mathbf{x}$ and λ a real parameter determining the magnitude of the deformation. Helmholtz equation for the velocity potential, $\psi(\mathbf{x},\omega)$, is given by

$$(\Delta+k_0^2)\psi(\mathbf{x},\omega)=0,\quad \mathbf{x}\in\mathbb{R}^2\backslash\bar{\Omega}, \tag{1}$$

with the impedance boundary condition

$$\frac{\partial}{\partial n}\psi(\mathbf{x},\omega)+\mathrm{i}\gamma(\mathbf{x})k_0\psi(\mathbf{x},\omega)=0,\quad \mathbf{x}\in\partial\Omega, \tag{2}$$

where $\mathbf{x}$ is the coordinate vector in a cartesian coordinate system. Here n denotes the outward unit normal to $\partial\Omega$ and γ is a real-valued nonnegative function relating the normal velocity on the boundary to the excess pressure. Zero Neumann boundary conditions correspond to a value of $\gamma=0$ in which case the obstacle is said to be sound-hard, whereas for $\gamma\to\infty$, zero Dirichlet boundary conditions give rise to the so-called sound-soft obstacle.

Let $\psi=\psi^{\mathrm{inc}}+\psi^{\mathrm{sc}}$ be the total acoustic field with ψ^{sc} the scattered field, $\psi^{\mathrm{inc}}=\mathrm{e}^{-i\mathbf{k}_0\mathbf{x}}$ denotes an incident acoustic field modeled as a plane wave where $\mathbf{k}_0=k_0\mathbf{d}$ is the wave number vector with $\mathbf{d}$ the direction of wave propagation, $k_0=\omega/c_0$ the wave number where ω is the frequency and c_0 the speed of sound. The harmonic time dependence of the plane wave is assumed to be of the form $\mathrm{e}^{-i\omega t}$ where t denotes time. The solution of Helmholtz equation yields an expression of the far field, $\psi_\infty^{\mathrm{sc}}=\mathcal{F}(r_\lambda)$ as function of the shape r_λ of the scatterer. The remainder of this section describes a particular approach for approximating $\mathcal{F}(r_\lambda)$ that is well-adapted to a certain class of shapes and which can be easily implemented in a computational context.

B. Approximation of the scattered field

As an example application, we consider the case of a two-dimensional scatterer whose shape is slightly deformed from a circle and whose surface is $\partial\Omega_0$. Thus, let the shape r_λ of the two-dimensional non circular scatterer, depending on a small parameter λ, be defined by the dependence of its radius on θ as[25–29]

$$r_\lambda=r_0+\lambda f(\theta), \tag{3}$$

in which r_0 is the radius of a circle providing the initial approximation to the shape and $f(\theta)$ is the functional form describing the variation of the shape away from this circle. The variation $f(\theta)$ is represented by its finite Fourier series in the form,

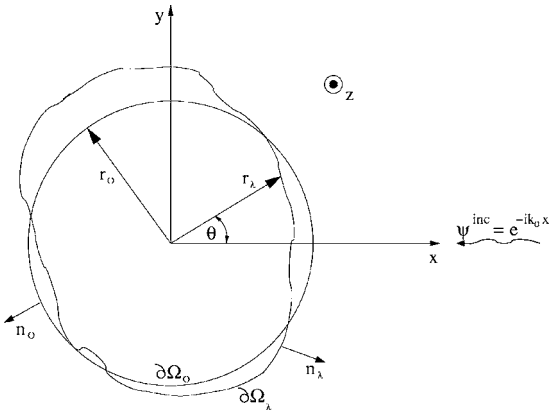


FIG. 1. Geometric configuration of the two-dimensional shape of the scatterer.

$$f(\theta) = \sum_{l=-L}^{L} \alpha_l e^{il\theta}. \tag{4}$$

Figure 1 shows the two-dimensional shape of the scatterer constructed around a circle of radius r_0 with boundary $\partial\Omega_0$ for a given value of the scalar λ determining r_λ and associated with boundary $\partial\Omega_\lambda$. The outward unit normals to $\partial\Omega_0$ and $\partial\Omega_\lambda$ are denoted by n_0 and n_λ, respectively.

According to Rayleigh's hypothesis[17,28,19–21,25,29] for this simple shape, the scattered field can be expanded in terms of outgoing wavefunctions of $\partial\Omega_0$ on and outside $\partial\Omega_0$ and the Rayleigh expansion is an analytical function.[17–19] Moreover, it can be shown[20,21] that for small deformations of simple shapes, that is for small λ, the scattered acoustic field ψ^{sc} is analytical in λ, and can thus be expanded in a Taylor series in λ[20,21,25] according to the following expression:

$$\psi^{sc}(r,\theta,\lambda) = \sum_{m=0}^{\infty} \frac{1}{m!} \psi^{sc(m)}(r,\theta)\lambda^m, \tag{5}$$

where the superscript (m) denotes the mth-order derivative evaluated at the origin. Parameters r and θ are the polar coordinates. Noting that, $\psi = \psi^{sc} + \psi^{i}$, where ψ^{i} denotes the incident potential, and differentiating equations (1) and (2) m times with respect to λ, equations are obtained which can be solved for the functions $\psi^{sc(m)}$. The solution $\psi^{sc}(r,\theta,\lambda)$ can then be synthesized using its Taylor expansion. Each $\psi^{sc(m)}$ can be expanded as[21,29]

$$\frac{1}{m!}\psi^{sc(m)}(r,\theta) = \sum_{n=-\infty}^{\infty} (-i)^n \beta_{m,n} H_n^{(1)}(k_0 r) e^{in\theta}, \tag{6}$$

where $H_n^{(1)}$ denotes the cylindrical Hankel function of the first kind of integer order n, $\beta_{m,n}$ are coefficients obtained recursively by substituting Eq. (6) into the equations obtained by differentiating equations (1) and (2) in succession m times with respect to λ. For the particular case of zero Dirichlet boundary conditions, the coefficients $\beta_{m,n}$ are given by[21,29]

$$\beta_{m,n} = -(T_1 + T_2)/H_n^{(1)}(k_0 r_0), \tag{7}$$

$$T_1 = (k_0^m/m!) \sum_{r=n-mL}^{n+mL} (-i)^{r-n} \alpha_{m,n-r} J_r^{(m)}(k_0 r_0), \tag{8}$$

$$T_2 = \sum_{l=0}^{m-1} (k_0^{m-l}/(m-l)!) \times \sum_{r=n-(m-l)L}^{n+(m-l)L} (-i)^{r-n} \alpha_{m-l,n-r} \beta_{l,r} (H_r^{(1)})^{(m-l)}(k_0 r_0), \tag{9}$$

where $J_r^{(m)}$ denotes the mth derivative of the Bessel function of integer order r where the initial guess $\beta_{0,n}$, given as

$$\beta_{0,n} = -J_n(k_0 r_0)/(H_n^{(1)})(k_0 r_0), \tag{10}$$

is associated with the Rayleigh expansion of the acoustic field $\psi^{(0)}$ scattered from the undeformed circular shape. In the previous equation, $\alpha_{m,n-r}$ denotes the coefficients in the Fourier expansion of the mth derivative, $f^{(m)}(\theta)$, of $f(\theta)$. The far field predicted from the aforementioned formulation requires truncation of the summations in Eqs. (4)–(6). The resulting approximation is denoted by $\mathcal{F}_{L,M,N}$ where the subscripts refer to the respective bounds on the above summations. Convergence of the approximation $\mathcal{F}_{L,M,N}$ as function of M, N, and L has already been established.[29]

The measured scattered acoustic field ψ_m^{sc} will be assumed to correspond to the far-field and is constructed from the predicted scattered field ψ^{sc} by multiplying it with a shaping function $w(\theta)$, then adding a spatially uncorrelated random noise, ϵ, having uniform probability density,

$$\psi_m^{sc} = w(\theta)\psi^{sc} + \epsilon. \tag{11}$$

The function $w(\theta)$ is meant to represent the effect of nonuniformity in the propagation medium which selectively affects the scattered field as a function of θ, while the additive noise is meant to represent other sources of error that do not exhibit spatial correlation. The following specific form is adopted for $w(\theta)$ in the rest of this article:

$$w(\theta) = A \cos\theta, \tag{12}$$

where A is a parameter that modulates the amplitude of the spatial distorsion. It should be noted that the measured field is obtained from Eq. (5) using Eqs. (6)–(9) in which Fourier coefficients α_l, $l=\{-L,\ldots,L\}$ in Eq. (4) are deterministic and the predicted stochastic model of the scattered field is similarly constructed but by using a random parameterization of the shape. The stochastic approach is explained in the next section.

III. STOCHASTIC PARAMETERIZATION OF THE SHAPE

The randomness is introduced in the modeling of the perturbation of the shape of the scatterer away from a circular geometry. In particular, a representation in terms of the polynomial chaos basis functions is used.[6] As can be seen in Eq. (3), the shape of the scatterer is completely defined by the radius r_0 of its nominal circular approximation and by the perturbation, $f(\theta)$, of the shape away from this circle. Moreover, $f(\theta)$ is specified through its Fourier coefficients as in-

dicated in Eq. (4). The measured field ψ_m^{sc}, which is constructed from ψ^{sc} with additive noise ϵ and multiplicative noise w, depends on r_0 and on the Fourier coefficients α_l, $l=-L$ to L.

The random parameterization of the shape is achieved by modeling the parameter r_0 as a random variable and $f(\theta)$ as a stochastic process, and denoting them by $\hat{r}_0(\tau)$ and $\hat{a}(\theta,\tau)$, respectively, where the argument τ indicates the random character of the quantity involved. This argument will also be used to index particular samples from a statistical population. The scattered field ψ_m^{sc} is approximated by the random field ψ_{pred} obtained from the prediction,

$$\psi_{\text{pred}}(\theta,\tau) = \mathcal{F}_{L,M,N}(\hat{s}_\lambda(\theta,\tau)), \tag{13}$$

where $\hat{s}_\lambda$ is given as

$$\hat{s}_\lambda(\theta,\tau) = \hat{r}_0(\tau) + \lambda\hat{a}(\theta,\tau). \tag{14}$$

A second-order random variable $\mu(\tau)$ can be approximated, in a mean-square convergent manner, using the polynomial chaos representation of the form,

$$\mu(\tau) = \sum_{p\geqslant 0} \sum_{n_1+\cdots+n_r=p} \sum_{\rho_1,\ldots,\rho_r} a_{\rho_1\cdots\rho_r}^{n_1\cdots n_r} \Gamma_p(\xi_{\rho_1}(\tau),\ldots,\xi_{\rho_r}(\tau)) = \sum_{i=0}^{\infty} \mu_i \phi_i(\tau). \tag{15}$$

The polynomial chaoses $\Gamma_p(\cdot)$ of any order p consist of all p-order finite dimensional orthogonal polynomials, with respect to the Gaussian function involving any combination of the random variables $\{\xi_{\rho_i}(\tau)\}_{i=1}^n$. The subscript n_i refers to the number of occurrences of $\xi_{\rho_i}(\tau)$ in the argument list for $\Gamma_p(\cdot)$. Moreover, the symbol $\phi_i(\tau)$ refers to a rearrangement of the polynomials Γ_p for ease of mathematical manipulations. In the case of one-dimensional polynomial chaoses, the vector $\{\xi_i\}$ reduces to a scalar ξ. The polynomial chaos of order greater than zero have mean zero and they are constructed either by a method based on a sequence of orthogonalization procedures or by using the generating function of the Hermite polynomials which gives,[6]

$$\Gamma_n(\xi_{i_1}(\tau),\ldots,\xi_{i_n}(\tau)) = (-1)^n \frac{\partial^n}{\partial\xi_{i_1}\cdots\partial\xi_{i_n}} e^{-\boldsymbol{\xi}^T\boldsymbol{\xi}/2}, \tag{16}$$

in which $\boldsymbol{\xi}$ denotes the vector of n random variables $\{\xi_{i_k}\}_{k=1}^n$ and superscript T denotes the transpose operator. The polynomial chaos decomposition is a representation in terms of a vector random variable $\boldsymbol{\xi}$ that essentially provides a discretization of Brownian motion. The parameters $\hat{r}_0(\tau)$ and $\hat{\alpha}(\theta,\tau)$ can then be expanded in their own polynomial chaos expansions as,

$$\hat{r}_0(\tau) = \sum_{i=0}^{\infty} r_{0i}\phi_i(\boldsymbol{\xi}(\tau)), \quad \hat{a}(\theta,\tau) = \sum_{i=0}^{\infty} a_i(\theta)\phi_i(\boldsymbol{\xi}(\tau)), \tag{17}$$

where the coefficients in this expansion can be expressed as generalized Fourier coefficients in the form,

$$r_{0,i} = \frac{\langle \hat{r}_0\phi_i\rangle}{\langle\phi_i^2\rangle}, \quad a_i(\theta) = \frac{\langle \hat{a}\phi_i\rangle}{\langle\phi_i^2\rangle}, \tag{18}$$

where $\langle\cdot\rangle$ refers to the operator of mathematical expectation. Thus each of the coefficients $a_i(\theta)$ can itself be represented in a finite Fourier expansion as

$$a_i(\theta) = \sum_{l=-L}^{L} a_{i,l} e^{il\theta}. \tag{19}$$

For the one-dimensional chaos case, associated with a scalar ξ, and which will be treated in the rest of this article, Eq. (17) reduces to

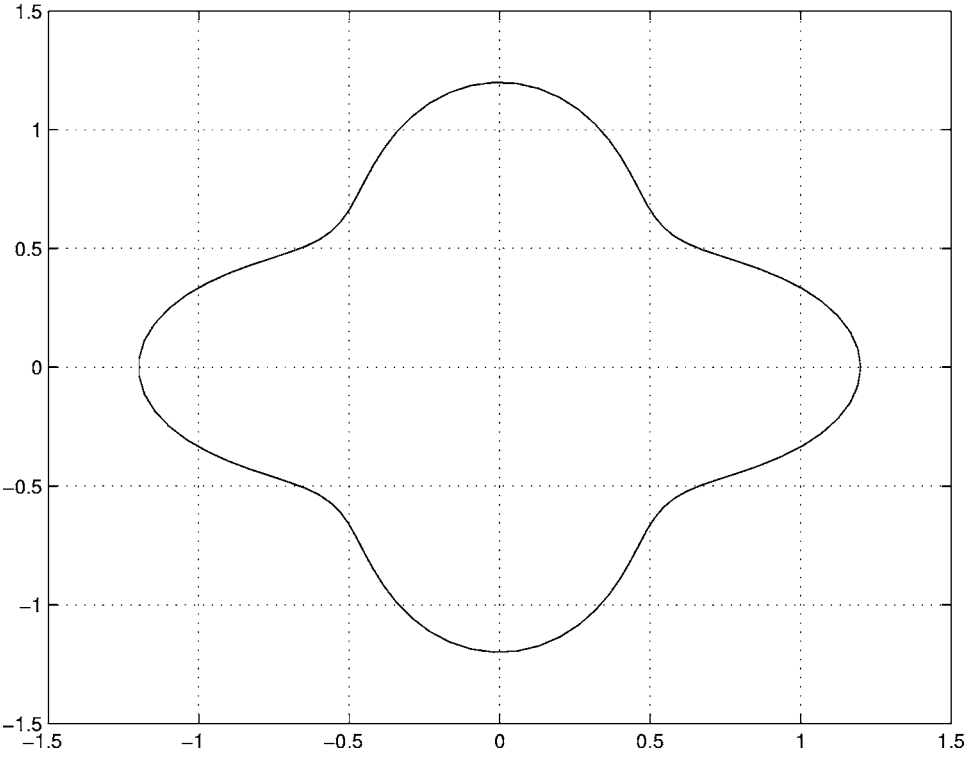


FIG. 2. Shape of the obstacle defined by $r_\lambda = r_0 + \lambda f(\theta)$ with $f(\theta)=\cos(4\theta)$ and $\lambda = 0.2$.

TABLE I. One-dimensional polynomial chaoses, $n=1$.

Order p	i	$\phi_i(\xi)$
0	0	1
1	1	ξ
2	2	ξ^2-1
3	3	$\xi^3-3\xi$
4	4	$\xi^4-6\xi^2+3$

$$\hat{r}_0(\tau)=\sum_{i=0}^{\infty} r_{0i}\phi_i(\xi(\tau)), \quad \hat{a}(\theta,\tau)=\sum_{i=0}^{\infty} a_i(\theta)\phi_i(\xi(\tau)), \quad (20)$$

with the functional forms of the first few $\phi_i(\xi)$ given in Table I. It should be noted that mean-square convergence is associated with the infinite-dimensional chaos representation. The one-dimensional representation adopted in this paper can be construed as either providing a representation with milder convergence behavior (namely, convergence in distribution), or as a first order approximation of the infinite-dimensional representation commensurate with a stochastic variation exhibiting a strong correlation structure. Either way, the present formulation yields a useful and meaningful representation of the shape of the scatterer. Substituting Eq. (17) into Eq. (13) yields

$$\psi_{\text{pred}}(\theta,\tau)=\mathcal{F}_{L,M,N}\left(\sum_{i=0}^{\infty}(r_{0i}+\lambda a_i(\theta))\phi_i(\tau)\right), \quad (21)$$

which is a parametrization of the predicted far field in terms of the chaos coefficients of the shape. It is evident that the form of the far-field operator $\mathcal{F}$ reflects the particular predictive tool used for the forward problem, and can therefore accommodate more accurate analysis methods of the Helmholtz equation such as integral and finite element formulations. Truncating the chaos representation of the shape after the Pth term, the predicted stochastic far field can then be represented as $\psi_{\text{pred}}=\mathcal{F}_{L,M,N,P}$.

It should also be noted that the coefficients in the polynomial chaos decomposition of a random quantity are in fact statistics of that quantity which can be estimated in a number of ways. If these coefficients are estimated directly from observations, the accuracy of each one of them will depend only on the observation sample used, and not on the accuracy of the other coefficients. If, however, as is done in this paper, the coefficients are estimated through a model-based inversion, then they are all coupled. In this case, the accuracy in each one of them depends on the number of polynomial chaos terms used in constructing the stochastic forward model, as well as on the physical and mathematical suitability of that model. Clearly, either way, the accuracy of the stochastic representation of the random quantity of interest depends both on the accuracy of each coefficient as well as on the number of terms included in the polynomial chaos expansion.

IV. STOCHASTIC INVERSE PROBLEM

A. Formulation

With the above preliminaries, the objective of the inverse analysis can be stated as computing the chaos characterization of the shape so as to minimize some measure of discrepancy between the observed and the predicted acoustic fields. The following two norms will prove useful in formulating and analyzing a solution to this problem;

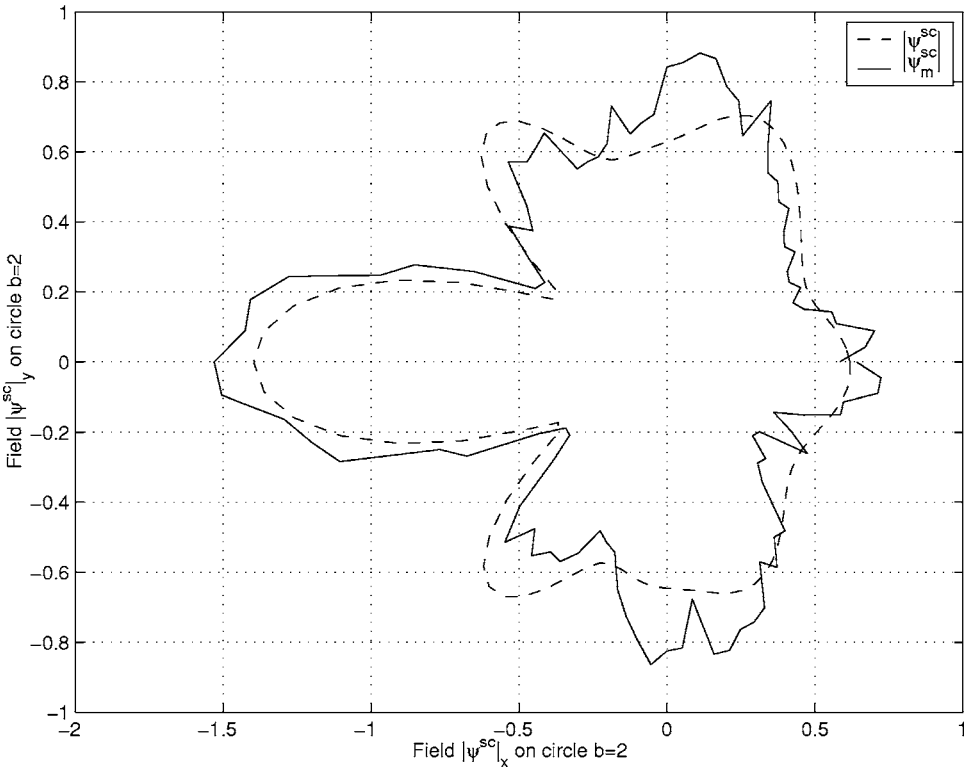


FIG. 3. The modulus of measured scattered field ψ^{sc}_m (solid line), compared to the modulus of predicted scattered field ψ^{sc} (dashed lines); $w(\theta)=1$; uniform random noise of 20%.

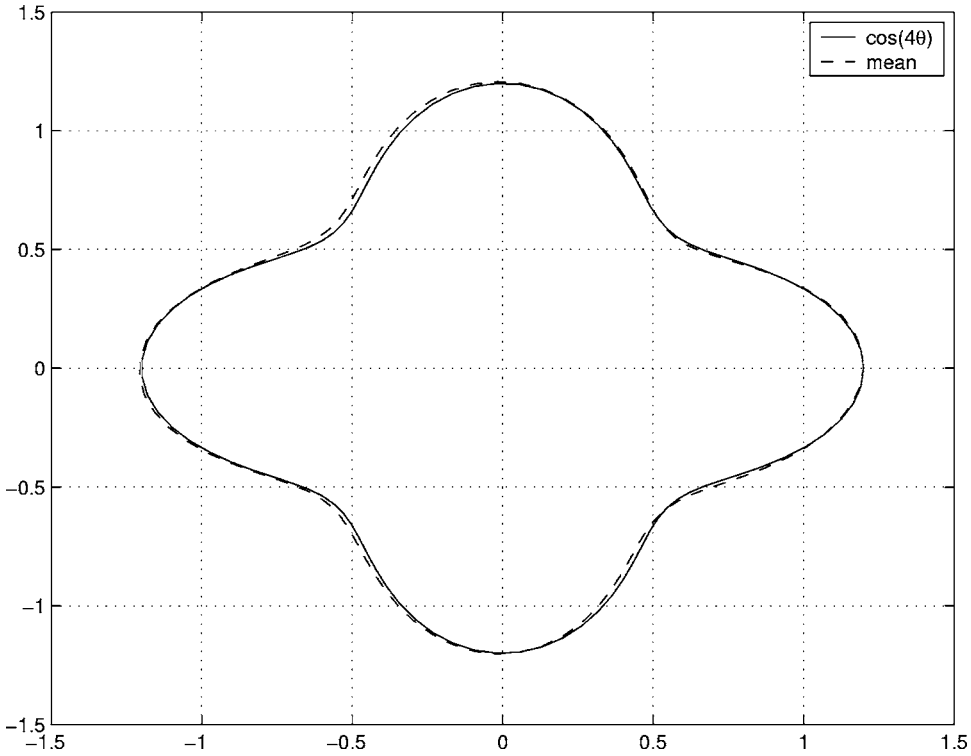


FIG. 4. Shape of the obstacle (dashed line) calculated from a deterministic model and compared to the theoretical shape $\cos(4\theta)$ (solid line); without noise.

$$\|\psi_{\text{pred}} - \psi_m^{\text{sc}}\|^2 = \int_0^{2\pi} |\psi_{\text{pred}}(\theta, \tau_r) - \psi_m^{\text{sc}}(\theta)|^2 d\theta \tag{22}$$

and

$$\|\|\psi_{\text{pred}} - \psi_m^{\text{sc}}\|\|^2 = \frac{1}{N}\sum_{r=1}^{N} \int_0^{2\pi} |\psi_{\text{pred}}(\theta, \tau_r) - \psi_m^{\text{sc}}(\theta)|^2 d\theta, \tag{23}$$

where the second norm involves sampling N variables ξ, constructing the N associated polynomial chaos, and averaging over the N corresponding predicted fields. The polynomial chaos coefficients in Eq. (18) are then estimated from a statistical sample of size N, which is taken to be large enough for the first few coefficients to be well approximated. As noted previously, these chaos coefficients are in fact statistics of the underlying stochastic process, and can be estimated using a wide range of possible methods. The present reliance on standard Monte Carlo sampling for their estimation omits efficiency consideration in favor of clarity of presentation so as not to detract from the objective of this paper which is to present a general methodology.

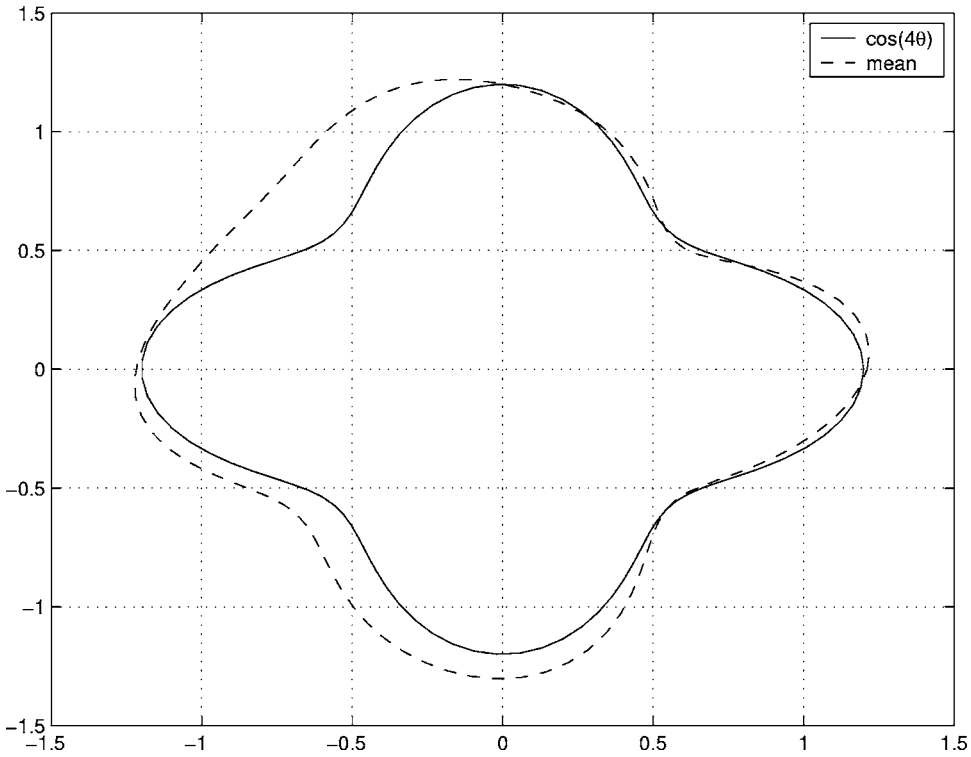


FIG. 5. Shape of the obstacle (dashed line) calculated from a deterministic model and compared to the theoretical shape $\cos(4\theta)$ (solid line); uniform random noise of 20%.

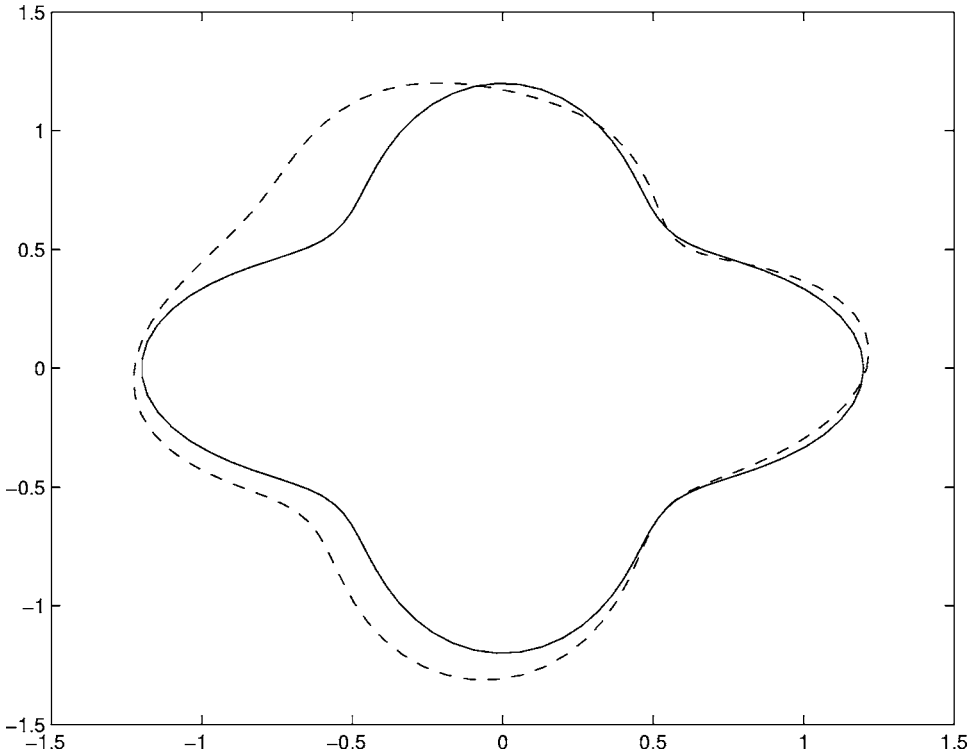


FIG. 6. Mean of the random shape of the obstacle (dashed line) calculated from a stochastic model with 200 simulations and compared to the theoretical shape $\cos(4\theta)$ (solid line); uniform random noise of 20%.

The optimization problem associated with the inverse analysis is stated as

$$\min_{r_{0i},a_{i,l}} |||\psi_{\text{pred}} - \psi_m^{sc}|||, \tag{24}$$

where the minimization is taken over the polynomial chaos coefficients r_{0i} and $a_{i,l}$. In the previous formalism, the observed field, ψ_m^{sc} is assumed to be deterministic, with the uniform noise in Eq. (11) reflecting the effect of a deterministic, unmodeled, medium. Once the coefficients r_{0i} and $a_{i,l}$ have been estimated, an optimal characterization of the random shape is then obtained and realizations, as well as statistics, of this shape can be easily recovered. In particular, the mean shape is given by the 0th order terms, the first-order term reflecting Gaussian perturbations

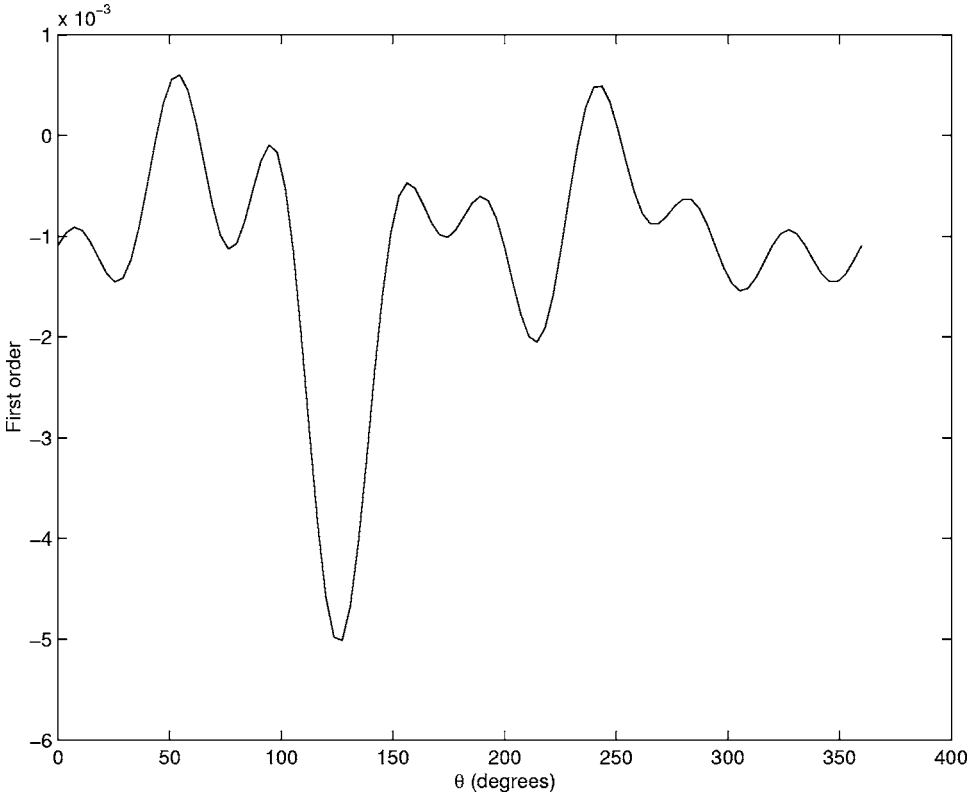


FIG. 7. First order coefficient in the polynomial chaos expansion calculated from a stochastic model with 200 simulations as a function of θ uniform random noise of 20%.

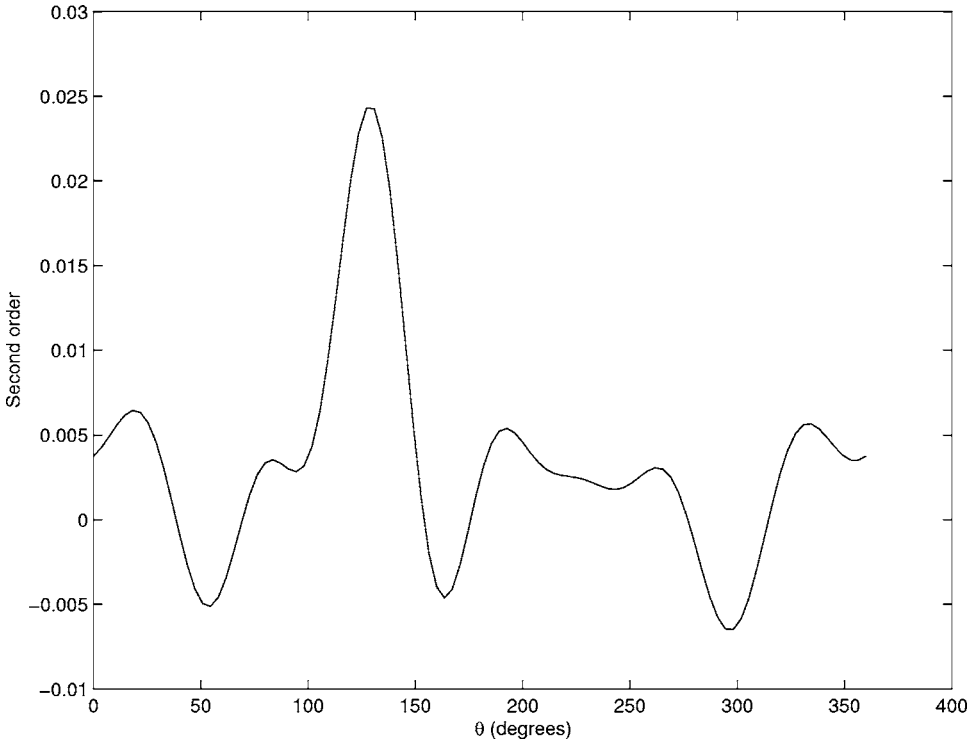


FIG. 8. Second order coefficient in the polynomial chaos expansion calculated from a stochastic model with 200 simulations as a function of θ; uniform random noise of 20%.

around this mean, and the higher-order terms representing non-Gaussian perturbations. The functional form of the random shape, $\hat{s}_\lambda(\theta,\tau)$, defined in Eq. (14), can be rewritten as

$$\hat{s}_\lambda(\theta,\tau) = s_0(\theta) + \sum_{i=1}^{\infty} s_i(\theta)\phi_i(\tau), \tag{25}$$

in which $s_0(\theta)$ and $s_i(\theta)$ are defined by

$$s_0(\theta) = r_{00} + \lambda a_{0,l} e^{il\theta}, \quad s_i(\theta) = r_{0i} + \lambda \sum_{l=-L}^{L} a_{i,l} e^{il\theta}. \tag{26}$$

The random shape is defined by $\hat{s}_\lambda(\theta,\tau)$, the mean shape by $s_0(\theta)$ and the ith order of polynomial chaos expansion by $s_i(\theta)$. The effect of these perturbations is quantified by calculating the variance $\sigma^2(\theta)$ and the coefficient of variation $\sigma_r(\theta)$ defined respectively by

$$\sigma^2(\theta) = \langle (\hat{s}_\lambda(\theta,\tau) - s_0(\theta))^2 \rangle,$$
$$\sigma_r(\theta) = \sqrt{\langle (\hat{s}_\lambda(\theta,\tau) - s_0(\theta))^2 \rangle}/s_0(\theta). \tag{27}$$

Given the orthogonality of the polynomial chaoses, Eq. (27) can be rewritten as

$$\sigma^2(\theta) = \sum_{i=1}^{\infty} s_i^2(\theta)\langle \phi_i^2 \rangle,$$
$$\sigma_r(\theta) = \sqrt{\sum_{i=1}^{\infty} s_i^2(\theta)\langle \phi_i^2 \rangle}/s_0(\theta). \tag{28}$$

Clearly, in a computational context, the summations in the previous expressions are truncated after some term, p.

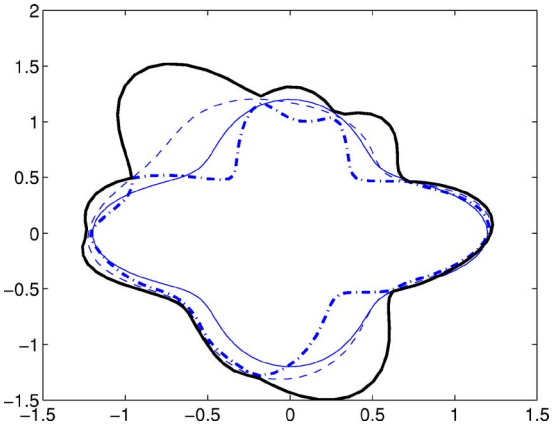


FIG. 9. (Color online) 95% confidence region for estimated shape; uniform random noise of 20% (solid and dotted-dashed lines); true shape (thin solid line); and mean predicted shape (dashed line).

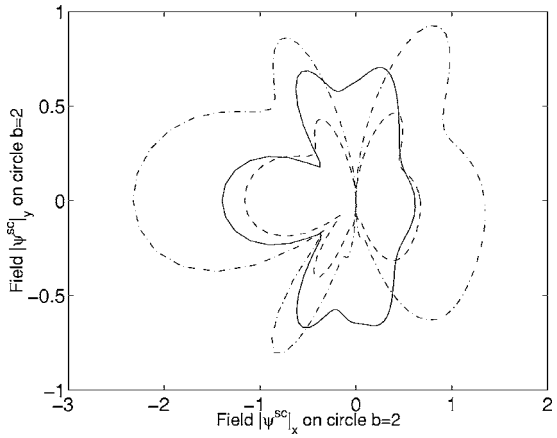


FIG. 10. Scattered field with multiplicative noise: $A=1$ (dashed line) and $A=2$ (dotted dashed line).

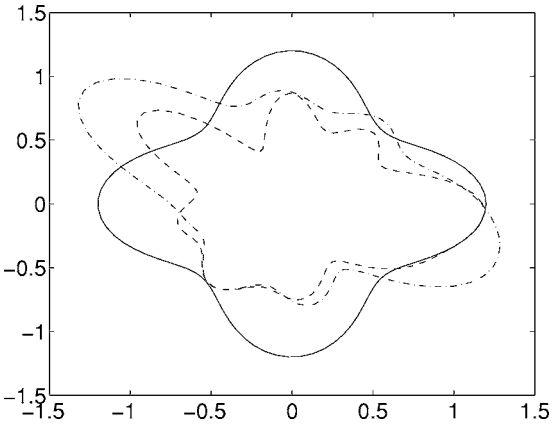


FIG. 11. Mean shape with multiplicative noise: true shape (solid line); $A=1$ (dashed line) and $A=2$ (dotted dashed line).

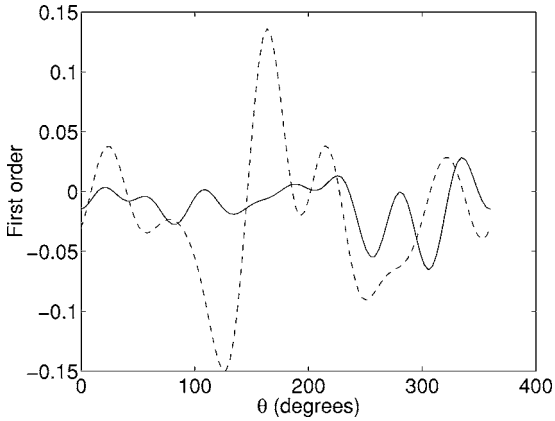


FIG. 12. First order chaos component of predicted field with multiplicative noise: $A=1$ (solid line) and $A=2$ (dashed line).

B. Discussion

By seeking a solution for the inverse problem in the form of a polynomial chaos expansion, the class of functions from which a solution is sought is now provided with a richer structure. Specifically, the solution is now defined by its projection on an infinite hierarchy of spaces corresponding to an increasing order of approximation. A description of the solution process in a finite dimensional representation provides a restriction of the solution to a proper subset of the original space, thus providing some level of regularization to the ill-posed problem. This statistical regularization is in addition to the regularization achieved by truncating the Fourier series representation of the spatial description of the shape.

Further, the norm in Eq. (24) can be expressed as

$$\|\|\psi_{\text{pred}} - \psi_m^{\text{sc}}\|\|^2 = \|\psi_{0,\text{pred}} - \psi_{0,m}^{\text{sc}}\|^2 + \langle\|\psi_{f,\text{pred}}\|^2\rangle + \langle\|\psi_{f,m}^{\text{sc}}\|^2\rangle - 2\langle\|\psi_{f,\text{pred}} - \psi_{f,m}^{\text{sc}}\|^2\rangle, \tag{29}$$

where a 0 subscript refers to the mean, whereas an f subscript refers to a zero-mean fluctuation. Since, in the present case, it is assumed that the measured scattered field is deteministic, and given that it is not a function of $a_{i,l}$, Eq. (29) simplifies to

$$\min_{r_{0i},a_{i,l}} \|\|\psi_{\text{pred}} - \psi_m^{\text{sc}}\|\|^2 = \min_{r_{0i},a_{i,l}} \|\psi_{0,\text{pred}} - \psi_{0,m}^{\text{sc}}\|^2 + \langle\|\psi_{f,\text{pred}}\|^2\rangle. \tag{30}$$

It is clear from this equation that in the case where the first norm on the right-hand is equal to zero, the optimal solution can be achieved when the second norm is equal to zero. Thus, if the mean measurements are in the range of the predictive model, the above formulation yields a deterministic shape as the optimal solution. In these cases, the variance of the estimated shape does not depend on its mean, and the deterministic shape is optimal. For those situations, however, where the measurements do not lie in the range of the model, a stochastic model will yield a more consistent match with the measurements. In these cases, the mean of the solution does not coincide with the solution of the mean problem, and the estimated mean shape depends, to a greater extent, on the estimsated variance of the shape. It is for these cases, where significant modeling errors exist, that the present stochastic formulation seems to be the most adapted. Based on this observation, it can be expected that whenever an additive noise is introduced on the scattered field, a stochastic component to the model will not be justified. Such a model for noise is often used in numerical investigations aimed at demonstrating the robustness of various inversion algorithms. If, on the other hand, a more structured noise is introduced, it can be expected that the average scattered field will no longer reside in the range of the model. A stochastic model would be more appropriate in this case. This observation is further explored in the following section.

The quantity $\langle\|\psi_{f,\text{pred}}|^2\rangle$ can be viewed as a measure of the lack-of-validity of the model, given the measurements. Thus, even in a deterministic reality, the modeling error can be somewhat mitigated by relying on a stochastic predictive model.

For situations where the measured scattered field is itself random, it can be expected that the probabilistic model parameters should depend, statistically, on these measurements. The correlation between model and measurements, appearing as the last term in Eq. (29) must then be obtained through appropriate statistical data analysis.

V. NUMERICAL EXAMPLE

Consider an obstacle whose cloverleaf shape is defined by Eq. (3) with $r_0=1$ and $f(\theta)=\cos(4\theta)$. Since the expansion used in Eq. (5) is a Taylor development, its accuracy degrades rapidly with increasing λ. A limiting value of $\lambda=0.25$ has been suggested[25] and $\lambda=0.2$ is used in the present example. This value of λ was kept fixed throughout the Monte Carlo simulation. It should be noted that Padé type approximants have been successfully used to extend the range of validity of these developments,[25] but they are not used in the present analysis.

Two cases will be investigated separately. In the first case, only additive uniform noise will be included in the representation of the measured signal ($w(\theta)=1$), whereas in the second case, only a multiplicative noise is included of the form given by Eq. (12). It will be observed that the strength

of the random component of the operator required to be consistent with the measurements depends greatly on the structure of the noise.

Figure 2 depicts the shape of the scatterer. In a computational setting the integral in Eq. (24) is discretized over θ resulting in the following optimization statement:

$$\min_{r_{0i},a_{i,l}} \sum_{r=1}^{N} \sum_{s=1}^{T} |\psi_{\text{pred}}(\theta_s,\tau_r) - \psi_m^{\text{sc}}(\theta_s)|^2. \tag{31}$$

Figure 3 presents the modulus of the measured scattered field ψ_m^{sc} (solid line), for an additive noise level of 20% as given by Eq. (11), compared to the deterministic predicted scattered field ψ^{sc} (dashed lines) defined by Eq. (5) with Eqs. (6)–(9).

Although numerical experiments carried out with values of $N=10$, 50, 100, and 200 indicated that a value of 100 was sufficient for the present problem, only results associated with $N=200$ are shown here. Also, 17 terms were included in the Fourier representation of $f(\theta)$ ($L=8$), a fifth-order Taylor expansion used in Eq. (5), and the summation index, n, in Eq. (6) varied from -10 to $+10$. Moreover, a value of $T=100$ observation points sampled uniformly around the circle was used in all the computations. The initial values for the optimization algorithm are, $a_{i,l}=0$, $\forall i$, l and $r_{0i}=1.2\delta_{i1}$, where δ_{i1} denotes the Kronecker delta. The resulting optimization problem is a large scale problem and is solved using the optimization software DAKOTA.[30] The results associated with the reconstruction of the random shape consist of the mean shape $s_0(\theta)$, the first and the second order terms in the polynomial chaos expansion. The variance and the coefficient of variation which, for a second order polynomial chaos expansion, are given by

$$\sigma^2 = s_1^2(\theta) + 2s_2^2(\theta), \quad \sigma_r = \sqrt{s_1^2(\theta) + 2s_2^2(\theta)}/s_0(\theta), \tag{32}$$

are also evaluated and displayed for all the cases considered here. Figures 4 and 5 show the reconstruction of the shape (dashed lines) from deterministic calculations for two noise levels: 0%, and 20%. It is clear that the added noise causes the reconstructed shape to depart from the theoretical shape $\cos(4\theta)$ (solid line). Similar observations are made for the mean shape deduced from stochastic calculations (dashed lines in Fig. 6). The mean of the reconstructed shape, its first and second-order chaos components are plotted in Figs. 6–8. It is clear from Figs. 6–8 that the magnitude of the random fluctuations, as witnessed by the chaos components, is almost negligible. It is also noted that the second order chaos components are larger than the first-order ones, indicating a strongly non-Gaussian character of the estimated shape. Figure 9 shows bands of 95% confidence for the estimated shape. While these bands do not necessarily bound the true shape at the 95% confidence level, they are expected to provide a close delineation of the true shape, once a converged stochastic representation has been achieved. It is clear that in the present case these bands provide a close, yet imperfect, band around the true shape.

Figure 10 shows the scattered field used to represent the measurements under conditions of multiplicative noise. The solid line depicts the scattered field under condition of $w(\theta)=1$ whereas the dashed lines show the scattered field associated with values of $A=1$ and $A=2$. The estimated mean shape is compared in Fig. 11 against the true shape. A significant discrepancy in the overall geometry is noticed for both values of the parameter A. Figure 12 and 13 show the first and second order chaos components. Notice the significant stochastic component, of the optimal shape as indicated by its first and second chaos components. Notice, here again, the nongaussian character of the estimated stochastic shape, as indicated by the relatively larger second order chaos. Figure 14 shows the 95% confidence zone for the reconstructed scatterer. As noted earlier, it is expected that as the order of polynomial chaos approximation increases, the quality of the bounds should improve. For comparison, the confidence bands associated with a first order, one-dimensional chaos representation is shown in Fig. 15. It is clear that the low resolution of the stochastic representation in this case yields a stochastic characterization of the optimal shape that is not consistent with observations. Rather than comparing with experimental data, which is not always practical, the validity of the stochastic characterization can also be verified by tracking the convergence of the chaos representation of the optimal shape, as the number of terms in the chaos expansion is

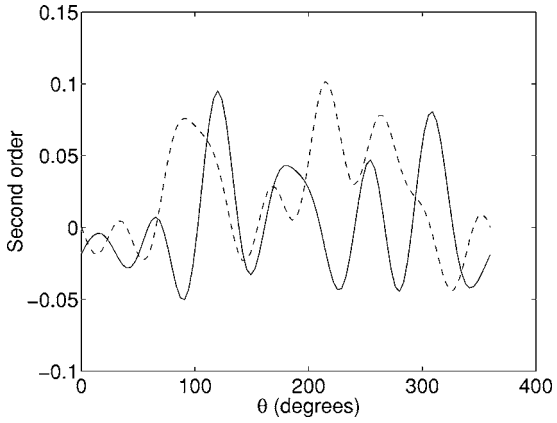


FIG. 13. Second order chaos component of predicted field with multiplicative noise: $A=1$ (solid line) and $A=2$ (dashed line).

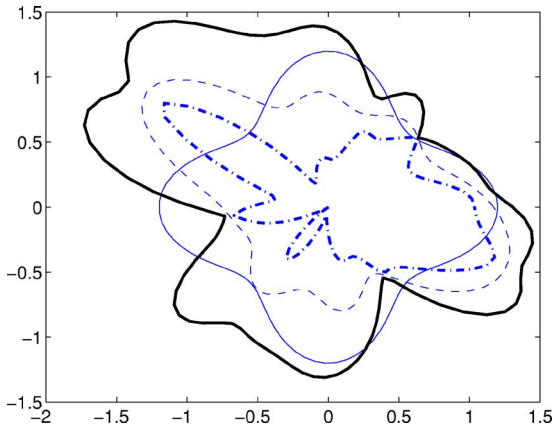


FIG. 14. (Color online) 95% confidence region for estimated shape for multiplicative noise: $A=2$ (solid and dotted-dashed lines); true shape (thin solid line); and mean predicted shape (dashed line); second order one-dimensional polynomial chaos approximation.

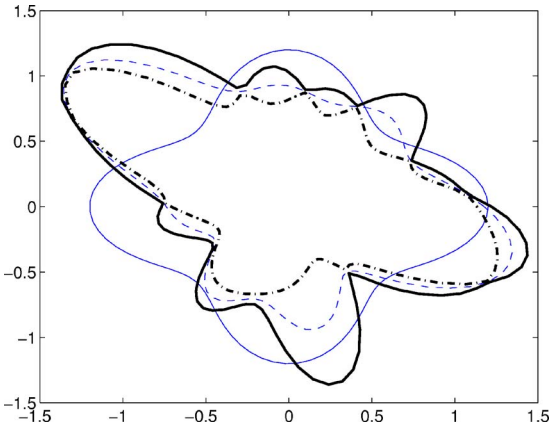


FIG. 15. (Color online) 95% confidence region for estimated shape for multiplicative noise: $A=2$ (solid and dotted-dashed lines); true shape (thin solid line); and mean predicted shape (dashed line); first order one-dimensional polynomial chaos approximation.

augmented. Although advances are underway to mitigate the computational cost associated with stochastic model-based predictions, this task remains currently computationally intensive.

VI. CONCLUSIONS

Clearly, the ability to extract probabilistic information from measured signals is of great significance in the effort toward assessing the worth of information and the risk associated with predictions based on limited information. This article has presented a novel formulation to characterize unobserved parameters in a probabilistic fashion, with application to the problem of stochastic inverse acoustic scattering. The shape of the scatterer is modeled as a random process and parameterized using the polynomial chaos expansion. From this random parametrization, a stochastic model of the scattered acoustic field is deduced. The optimization problem then consists of evaluating the coefficients in the polynomial chaos parametrization so as to minimize a norm of the discrepancy between the measured and predicted acoustic fields. This norm involves averaging over the probabilistic measure associated with the random shape.

It has been noted that in the presence of purely additive errors, such as measurement errors, the optimal scatterer shape is nearly deterministic, with a negligible random component. This is consistent with the observation that Helmholtz equation, representing waves in a homogeneous medium, is an appropriate model for this situation. When the data, on the other hand, are synthesized through a nonlinear distortion of Helmholtz-consistent data, the stochastic component of the optimal scatterer shape gained in significance indicating, as expected, the reduced confidence associated with using the predictive instrument. It should be emphasized that the prediction instrument consists of the Helmholtz equation, together with the particular choice of its numerical resolution.

The proposed method can be adapted to account for additional uncertainties in the system. For instance a polynomial chaos decomposition of the measured field itself would permit the modeling of random spatial fluctuations or missing observations in the observed far field.

Of particular interest is the observation that even deterministic problems can be better approximated by stochastic models, depending on the sufficiency of the physical phenomena retained in the predictive model. Indeed, a probabilistic approach, as proposed in this article, can be employed to quantify the validity of such models.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Contract No. N00014-99-1-0900.

[1]H. Ogura, N. Takahashi, and M. Kuwahara, "Scattering of waves from a random cylindrical surface," Wave Motion **14**, 273–295 (1991).
[2]J. A. Ogilvy, *Theory of Wave Scattering from Random Rough Surfaces* (IoP, Bristol, 1991).
[3]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (IEEE Press and Oxford University Press, New York, 1997).
[4]R. Ohayon and C. Soize, *Structural Acoustics and Vibration* (Academic, London, 1998).
[5]G. Bal, V. Freilikher, G. Papanicolaou, and L. Ryzhik, "Wave transport along surfaces with random impedance," Phys. Rev. B **62**, 6228–6240 (2000).
[6]R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, New York, 1991).
[7]R. Ghanem and A. Sarkar, "Reduced models for the medium-frequency dynamics of stochastic systems," J. Acoust. Soc. Am. **113**, 834–846 (2003).
[8]R. Ghanem and A. Sarkar, "Mid-frequency structural dynamics with parameter uncertainty," Comput. Methods Appl. Mech. Eng. **191**, 5499–5513 (2003).
[9]D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory* (Wiley, New York, 1983).
[10]D. Colton, *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer, Berlin, 1992).
[11]D. Colton, J. Coyle, and P. Monk, "Recent developments in inverse acoustic scattering theory," SIAM Rev. **42**, 369–414 (2000).
[12]A. Zinn, "The numerical solution of an inverse scattering problem for time-harmonic acoustic waves in a penetrable medium," in *Inverse Problems and Imaging*, edited by G. F. Roach (Longmans Green, London, 1991), pp. 242–263.
[13]A. Kirsch and R. Kress, "An optimisation method in inverse acoustic scattering," in *Boundary Elements*, edited by C. A. Brebbia, W. L. Wendland, and G. Kuhn (Springer, Heidelberg, 1987), pp. 3–18.
[14]C. Farhat, R. Tezaur, and R. Djellouli, "On the solution of three-dimensional inverse obstacle acoustic scattering problems by a regularized Newton method," Inverse Probl. **5**, 1229–1246 (2000).
[15]R. Feijóo, M. Malhotra, A. Oberai, and M. Pinsky, "Shape sensitivity calculations for exterior acoustics problems," Eng. Comput. **18**, 376–391 (2001).
[16]R. Guillermin, P. Ladaygues, J. P. Sessarego, and A. Wirgin, "Imaging an object buried in the sediment bottom of a deep sea by linearized inversion of synthetic and experimental scattered acoustic wavefields," Inverse Probl. **16**, 1777–1797 (2000).
[17]M. L. Burrows, "Equivalence of the Rayleigh solution and the extended boundary-condition solution for scattering problems," Electron. Lett. **5**, 277–278 (1969).
[18]R. F. Millar, "Rayleigh hypothesis in scattering problems," Electron. Lett. **5**, 416–418 (1969).
[19]P. M. Van den Berg and J. T. Fokkema, "The Rayleigh hypothesis in the theory of diffraction by a cylindrical obstacle," IEEE Trans. Antennas Propag. **AP-27**, 577–583 (1979).
[20]O. P. Bruno and F. Reitich, "Solution of a boundary value problem for the Helmholtz equation via variation of the boundary into the complex domain," Proc. - R. Soc. Edinburgh, Sect. A: Math. **122A**, 317–340 (1992).
[21]O. P. Bruno and F. Reitich, "A new approach to the solution of problems of scattering by bounded obstacles," *Ultra-Wideband, Short-Pulse Electromagnetics 2* (Plenum, New York, 1995).

[22]T. S. Angell, R. E. Kleinmann, B. Kok, and G. F. Roach, "A constructive method for identification of an impenetrable scatterer," Wave Motion **II**, 185–200 (1989).
[23]T. Scotti and A. Wirgin, "Shape reconstruction of an impenetrable scattering body via Rayleigh hypothesis," Inverse Probl. **12**, 1027–1055 (1996).
[24]D. S. Jones and X. Q. Mao, "The inverse problem in hard acoustic scattering," Inverse Probl. **5**, 731–748 (1989).
[25]D. N. Ghosh Roy, L. Couchman, and J. Warner, "Scattering and inverse scattering of sound-hard obstacles via shape deformation," Inverse Probl. **13**, 585–606 (1997).
[26]D. N. Ghosh Roy, J. Warner, L. Couchman, and J. Shirron, "Inverse obstacle transmission problem in acoustics," Inverse Probl. **14**, 903–929 (1998).
[27]L. Couchman, D. N. Ghosh Roy, and J. Warner, "Inverse Neumann obstacle problem," J. Acoust. Soc. Am. **104**, 2615–2621 (1998).
[28]J. Warner, D. N. Ghosh Roy, J. Bucaro, and L. Couchman, "Inversion of penetrable obstacles from far-field data on narrow angular apertures," J. Acoust. Soc. Am. **107**, 1111–1120 (2000).
[29]D. N. Ghosh Roy and L. Couchman, *Inverse Problems and Inverse Scattering of Plane Waves* (Academic, New York, 2001).
[30]M. S. Eldred, A. A. Giunta, B. G. Van Bloemen Waanders, S. F. Wojtkiewicz, W. E. Hart, and M. P. Alleva, "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis," Sandia National Laboratories, Albuquerque, 2002.

Natural resonance frequencies, wave blocking, and energy localization in an elastic half-space and waveguide with a crack

Evgeny Glushkov,[a] Natalia Glushkova, and Michael Golub
Department of Applied Mathematics, Kuban State University, Krasnodar, 350040 Russia

Anders Boström
Department of Applied Mechanics, Chalmers University of Technology, SE-412 96 Göteborg, Sweden

(Received 6 October 2005; revised 13 March 2006; accepted 19 March 2006)

Sharp stopping of time-harmonic wave transmission in elastic structures with defects is considered as a manifestation of the well-known trapped mode effect. It is associated with natural resonance poles lying close to the real axis in the complex frequency plane. Nonresonant wave blocking may also occur due to antiphase combination of the incident and scattered waves. The present paper is aimed to give an insight into such phenomena using an analytically based computer model which strictly takes into account all wave interactions in a cracked structure. Numerical examples are restricted to the case of a line horizontal crack in a half-plane or in a layer (2D in-plane motion), that is, nevertheless, quite enough to demonstrate two kinds of the Rayleigh wave stopping mechanisms (resonant and nonresonant) as well as a possibility of pure real natural resonance frequencies and of a full blocking effect with energy localization. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2195269]

PACS number(s): 43.20.Gp, 43.20.Ks, 43.20.Mv, 43.20.Bi [LLT] Pages: 3589–3598

I. INTRODUCTION

The diffraction of traveling waves by a subsurface defect sometimes entails a sharp growth of the surface displacement above the obstacle. Conventionally such a phenomenon is explained by the trapped mode effect,[1,2] featured by the wave energy localization near the obstacle. Mathematically, the trapped mode occurs when a spectral point ω_n of the boundary value problem modeling a time-harmonic wave process $\mathbf{u}e^{-i\omega t}$ lies close to the real axis $\mathrm{Im}\,\omega=0$ in the complex frequency plane. Excitation of the structure at frequencies $\omega \approx \mathrm{Re}\,\omega_n$ with $|\mathrm{Im}\,\omega_n|/\mathrm{Re}\,\omega_n \ll 1$ gives rise to resonance effects such as wave trapping and screening. The rigorous mathematical foundation of resonance classification for inhomogeneous elastic waveguides is grounded by Vorovich.[3,4] Besides the resonance effects, an antiphase combination of incident and diffracted waves can also result in nonresonant stopping of the wave transmission.

With regard to practical applications, there exists a strong interest in studying the elastodynamic scattering by cracklike obstacles. It is of importance both to the development of crack detection methods and to the assessment of a possible failure due to resonance effects.

Investigations of the location of the resonance poles in the complex ω plane for a crack in an unbounded elastic space[5–7] has shown that ω_n lies rather far from the real axis. Obviously, in the presence of boundaries, providing additional wave fields reflected recurrently from the crack and boundary surfaces, the location of ω_n changes so that in some situation $|\mathrm{Im}\,\omega_n|$ may become very small. Therefore, the first goal of our work is to trace the poles in the complex plane as the crack changes its size or approaches the surface. The prime question here is: is it even possible to have a purely real spectral point ω_n? Accompanying questions are: what happens at resonant frequencies regarding to wave transmission by and energy localization near the crack; what is the mechanism of wave blocking; is it even possible to have a full stopping of traveling wave transmission by a subsurface obstacle?

Traditionally the mathematical modeling in nondestructive evaluation relies on the ray methods of the geometrical theory of diffraction (GTD) which operates with physically evident wave expressions. Unfortunately, these methods are not suitable to clear up these questions due to the limits of their asymptotic nature. On the other hand, the direct numerical methods based on the mesh discretization of a solid volume are too time consuming and cannot provide an insight into the wave structure of the solution.

To address the questions stated above we rely on the analytically based computer model developed earlier for the purposes of ultrasonic crack detection.[8] Within this model the incident and scattered fields are expressed in terms of path Fourier integrals and their asymptotics via Green's matrices of the structures considered, given source characteristics (loads) and unknown crack opening displacements (c.o.d.). The latter are found from the boundary integral equations (BIEs) implied by the boundary conditions at the crack sides.

The use of the integral equation approach for the problem considered is justified, therefore, by the fact that it holds an intermediate position between the ray and direct numerical methods. First, it provides the same quantitative solution like FEM (finite element method) or finite differences, but with much lower costs. Then, it is easily applicable to infinite domains (open waveguides), and asymptotics derived from the integral representations give the same physically signifi-

[a]Electronic mail: evg@math.kubsu.ru

cant expressions as ray-based approaches, while keeping in contrast all the comprehensive information about sources, structures, and scatters. These asymptotics are quite appropriate for the GTD as well.

Though the developed model[8] provides rigorous 3D elastodynamic solutions for arbitrarily shaped and inclined cracks in layered structures, in the present work we limit ourselves to 2D in-plane diffraction problems for horizontal striplike (line) cracks. It is quite sufficient for examining the wave effects of interest.

Earlier we already considered blocking effects associated with the diffraction by obstacles of a different nature. The investigation into the fine structure of energy fluxes in an elastic layer with a surface rigid obstacle helped us, for example, in understanding the role of energy vortices in these processes.[9,10] The first results for the blocking by a crack in a half-space are given in a recent short communication,[11] then the investigation for a crack in a layer is started.[12] The present work summarizes and develops those previous results. The main distinctive attainment is the discovery of pure real spectral points ω_n and the analysis of the eigensolutions yielding wave localization.

The description of the mathematical technique used is available in the papers mentioned above. Therefore, in Sec. II we give only a brief description of the model sufficient for understanding the results. Nevertheless, the explicit analytical representations of Green's matrices for the structures considered, which may be of independent interest, are attached in the Appendix.

Then, in Sec. III, we focus on the Rayleigh wave diffraction by a horizontal crack in a half-plane. With this simplest model we consider typical situations of wave transmission stopping and energy localization. The most remarkable fact discovered here is a possibility of a full Rayleigh wave blocking without any energy trapping (nonresonant antiphase annihilation), whereas the resonance poles ω_n yield strong energy localization. They are not purely real and do not assure, therefore, the complete wave blocking in a half-plane.

For a layer the dependence of the transmission coefficient on geometry (crack's width and depth) and frequency is much more complicated. The examples of Sec. IV show a crucial role of the resonance poles for the transmission control in layer waveguides. In contrast to the half-plane structure, examples of purely real poles ω_n as well as examples of a full resonant blocking have been found here. As is to be expected, the calculation of displacement eigenforms associated with the real spectral points ω_n shows strong wave localization near the crack. Hence, a sharp resonant blocking at $\omega \approx \mathrm{Re}\omega_n$ is easily explained by the fact that, in such a case, most of the wave energy is distributed to the dominating eigenform featured by a sharp localization around the crack.

II. MATHEMATICAL MODEL

Let us consider an in-plane steady-state harmonic oscillation $\mathbf{u}(\mathbf{x})e^{-i\omega t}[\mathbf{u}=(u_x,u_z)]$ of an elastic isotropic layer of thickness H, governed by the Lamé equations

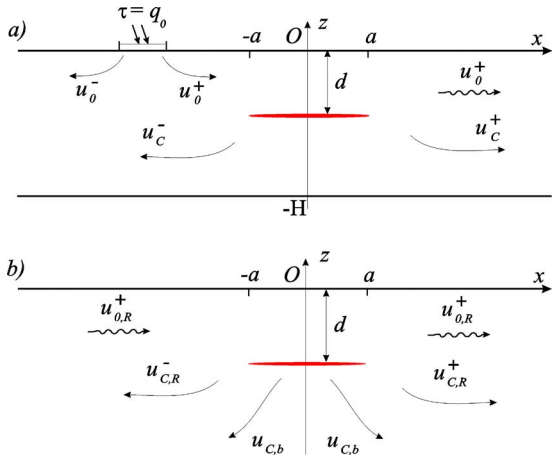


FIG. 1. (Color online) Wave fields' geometry in a layer (a) and a half-plane (b).

$$(\lambda + \mu)\,\nabla\ \mathrm{div}\,\mathbf{u} + \mu\Delta\mathbf{u} + \rho\omega^2\mathbf{u} = 0.$$

Here and below, due to linearity of the governing relations, the harmonic multiplier $e^{-i\omega t}$ is conventionally omitted. The material properties are specified by Lamé constants λ, μ, and density ρ.

The layer contains a subsurface horizontal line crack of width $2a$ at the depth d from the upper surface [Fig 1(a)]. The thickness H may be infinite ($H=\infty$); in such a case we deal with a half-plane [Fig. 1(b)].

In the Cartesian coordinate system $\mathbf{x}=(x,z)$ the layer occupies the domain $|x| \leqslant \infty$, $-H \leqslant z \leqslant 0$. The crack is modeled by an infinitesimally thin cut along a line segment $|x| \leqslant a$, $z=-d$ with stress-free sides:

$$\boldsymbol{\tau}|_{z=-d} = 0, \quad |x| < a, \tag{1}$$

here $\boldsymbol{\tau}=(\tau_{xz},\sigma_{zz})$ is the stress vector at a horizontal surface element. The displacement vector is discontinuous on the crack with an unknown jump (c.o.d.):

$$\mathbf{v}(x) = \mathbf{u}(\mathbf{x})|_{z=-d-0} - \mathbf{u}(\mathbf{x})|_{z=-d+0}. \tag{2}$$

It is assumed that the crack sides do not touch each other.

For definiteness let us consider a free waveguide excited by a surface load $\mathbf{q}_0$ or by an incident wave $\mathbf{u}_0^+$ coming from infinity. Thus, the sides $z=0$ and $z=-H$ are stress-free except, possibly, at the local source zone $|x-x_0| \leqslant b$ to which the load $\mathbf{q}_0$ is applied:

$$\boldsymbol{\tau}|_{z=0} = \begin{cases} \mathbf{q}_0(x), & |x-x_0| \leqslant b, \\ \mathbf{0}, & |x-x_0| > b, \end{cases} \tag{3}$$

$$\boldsymbol{\tau}|_{z=-H} = 0. \tag{4}$$

For a half-plane condition (4) is unnecessary.

Finally, for ideally elastic material properties boundary conditions (1)–(4) are to be supplemented by certain radiation conditions assuring the uniqueness. As the radiation condition we use the principle of limiting absorption,[13] which means that the solution for an ideally elastic medium is the limit of the unique solution of the corresponding problem for a medium with attenuation ε as $\varepsilon \to 0$.

In the presence of a crack the total wave field is made up of the incident field $\mathbf{u}_0$ and the field diffracted by the crack $\mathbf{u}_c$: $\mathbf{u}=\mathbf{u}_0+\mathbf{u}_c$. Let us mark the part of a wave field propagating in the waveguide from the left to the right by a superscript "plus" ($\mathbf{u}_0^+, \mathbf{u}_c^+$), while $\mathbf{u}_0^-$, $\mathbf{u}_c^-$ are traveling waves going in the opposite direction. For definiteness, let the incident wave come to the crack from the left, i.e., it is $\mathbf{u}_0^+$. Consequently, the transmitted field is $\mathbf{u}_0^+ + \mathbf{u}_c^+$ and the reflected one is $\mathbf{u}_c^-$ (Fig. 1). In a half-plane both $\mathbf{u}_0$ and $\mathbf{u}_c$ contain in addition the body-wave constituents $\mathbf{u}_{0,b}$ and $\mathbf{u}_{c,b}$ radiating downward.

Using the Fourier transform technique with respect to the horizontal coordinate x, $\mathbf{u}_0$ and $\mathbf{u}_c$ are derived in the form of inverse Fourier integrals:

$$\mathbf{u}_0 = \frac{1}{2\pi}\int_\Gamma \mathbf{U}_0(\alpha,z)e^{-i\alpha x}\,d\alpha \equiv \mathcal{F}^{-1}[\mathbf{U}_0],$$

$$\mathbf{U}_0 = \int_{-\infty}^{\infty} \mathbf{u}_0(x,z)e^{i\alpha x}\,dx \equiv \mathcal{F}[\mathbf{u}_0]. \tag{5}$$

Similarly,

$$\mathbf{u}_c = \mathcal{F}^{-1}[\mathbf{U}_c], \quad \mathbf{U}_c = \mathcal{F}[\mathbf{u}_c].$$

By $\mathcal{F}$ and $\mathcal{F}^{-1}$ we denote direct and inverse transforms. The integration path Γ goes in the complex plane α along the real axis $\mathrm{Im}\,\alpha=0$, deviating from it for bypassing real poles and branch points of the integrands $\mathbf{U}_0$ and $\mathbf{U}_c$ in accordance with the principle of limiting absorption.[4] The solution in the Fourier transform domain is obtained in the form

$$\mathbf{U}_0(\alpha,z) = K(\alpha,z)\mathbf{Q}_0(\alpha),$$

$$\mathbf{U}_c(\alpha,z) = M(\alpha,z)\mathbf{V}_0(\alpha), \tag{6}$$

where $\mathbf{Q}_0=\mathcal{F}[\mathbf{q}_0]$ and $\mathbf{V}=\mathcal{F}[\mathbf{v}]$ are the Fourier transforms for $\mathbf{q}_0$ and $\mathbf{v}$ while K and M are transforms of Green's matrices (2×2), whose columns are fields exited by horizontal and vertical surface point loads applied to the surface $z=0$ (for K) and fields related to horizontal cracklike point sources with shear and normal crack opening displacements (for M). Elements of these matrices are derived in a closed analytical form (see the Appendix). In the analytical calculations we use the technique developed in Ref. 4; more details of its specific implementation may be also found in Refs. 8, 10, 11, and 14.

With a layer waveguide $K(\alpha,z)$ and $M(\alpha,z)$ are meromorphic in the complex α-plane matrix functions. They have a finite number of real poles $\pm\zeta_k, k=1,2,\ldots,N$, and a countable set of complex ones $(k=N+1,N+2,\ldots)$. The wave fields $\mathbf{u}_0^\pm$ and $\mathbf{u}_c^\pm$ can be expressed as a sum of residues from these poles:

$$\mathbf{u}_0^\pm(\mathbf{x}) = \sum_{k=1}^{\infty} \mathbf{a}_k^\pm(z)e^{\pm i\zeta_k x}, \quad |x-x_0|>b, \tag{7}$$

$$\mathbf{u}_c^\pm(\mathbf{x}) = \sum_{k=1}^{\infty} t_k^\pm \mathbf{a}_k^\pm(z)e^{\pm i\zeta_k x}, \quad |x|>a, \tag{8}$$

$$\mathbf{a}_k^\pm = \mp i\,\mathrm{res}K(\alpha,z)|_{\alpha=\mp\zeta_k}\mathbf{Q}_0(\mp\zeta_k),$$

$$t_k^\pm \mathbf{a}_k^\pm = \mp i\,\mathrm{res}M(\alpha,z)|_{\alpha=\mp\zeta_k}\mathbf{V}(\mp\zeta_k).$$

The first N terms of these series are related to real poles describing traveling waves (normal modes) propagating along the layer to infinity with the phase and group velocities $c_{p,k}=\omega/\zeta_k$ and $c_{g,k}=d\omega/d\zeta_k$. The rest of the terms decay at infinity as $e^{-|\mathrm{Im}\,\zeta_k||x|}$, $|x|\to\infty$. The appearance of dispersion curves $c_{p,k}(\omega)$ and $c_{g,k}(\omega)$ for the layer under consideration is well studied (e.g., see Refs. 15 and 16). In the numerical examples considered in Sec. IV, only the two first fundamental modes a_0 and s_0 exist in the dimensionless frequency range $0<\omega<\pi$, and three or more traveling waves for $\omega>\pi$.

Whereas the normal modes $\mathbf{a}_k^\pm e^{\pm i\zeta_k x}$ in the expansion (7) are known, the related normal modes for $\mathbf{u}_c$ contain the unknown coefficients $t_k^\pm$, which are expressed via the unknown c.o.d. $\mathbf{v}$ [more precisely, via $\mathbf{V}(\mp\zeta_k)$ components].

For a half-plane ($H=\infty$) the components of the matrices K and M have two branch points and only one Rayleigh pole ζ, so that in this case

$$\mathbf{u}_0 = \mathbf{u}_{0,R} + \mathbf{u}_{0,b}, \quad \mathbf{u}_c = \mathbf{u}_{c,R} + \mathbf{u}_{c,b},$$

where

$$\mathbf{u}_{0,R}^\pm = \mathbf{a}_1^\pm(z)e^{\pm i\zeta x}, \quad |x-x_0|>b;$$

$$\mathbf{u}_{c,R}^\pm = t_1^\pm \mathbf{a}_1^\pm(z)e^{\pm i\zeta x}, \quad |x|>a, \tag{9}$$

are Rayleigh waves radiated to infinity from the source and the crack, respectively. The body waves $\mathbf{u}_{0,b}$, $\mathbf{u}_{c,b}$ are represented in terms of integrals over the sides of cuts drawn in the α plane from the branch points to infinity.

The transmission and reflection coefficients κ^+ and κ^- are introduced as the ratios $\kappa^\pm=E^\pm/E_0^+$, in which E_0^+ is the absolute value of the time-averaged energy of the waves $\mathbf{u}_0^+$ incident on the crack, while $E^\pm$ are the energy of the transmitted and reflected traveling waves. These energy characteristics are calculated by integrating the horizontal component e_x of the time-averaged power density vector (Umov-Poynting vector) $\mathbf{e}=(e_x,e_z)$ over a vertical cross section: $E=\left|\int_{-H}^0 e_x(x,z)\,dz\right|$, $x=\mathrm{const}$. Here $e_x=-(\omega/2)\,\mathrm{Im}\,(\mathbf{u}\cdot\boldsymbol{\sigma})$, where $\boldsymbol{\sigma}=(\sigma_{xx},\tau_{xz})$ is a stress vector at a vertical surface element; as $\mathbf{u}$ we take here $\mathbf{u}_0^+$, $\mathbf{u}_0^+ + \mathbf{u}_c^+$, and $\mathbf{u}_c^-$ for E_0^+, E^+, and E^-, respectively. Often only one normal mode of the sum (7) or a single Rayleigh wave (9) is taken as an incident wave $\mathbf{u}_0^+$ to study its transmission and reflection. In such a case the source is insignificant; we can consider $\mathbf{u}_0^+$ as a wave coming from minus infinity.

The source characteristics are accounted for in the energy distribution coefficients $\mu^\pm=E^\pm/E_0$, introduced similarly to $\kappa^\pm$, but with E_0 to be the source power (total energy radiated into the structure from the load $\mathbf{q}_0$: $E_0=\left|\int_{-\infty}^{\infty} e_z(x,0)\,dx\right|$, $e_z=-(\omega/2)\,\mathrm{Im}\,(\mathbf{u}\cdot\mathbf{q}_0)$, $\mathbf{u}=\mathbf{u}_0+\mathbf{u}_c$); $E^\pm$ are expressed as above but via $\mathbf{u}_0^\pm+\mathbf{u}_c^\pm$.

For a half-space the surface-to-body transformation coefficients $\kappa_b=E_b/E_0^+$, $\mu_b=E_b/E_0$ should be introduced in addition (here E_b is the total energy of body waves integrated over the lower semi-circle as its radius $R\to\infty$). The power

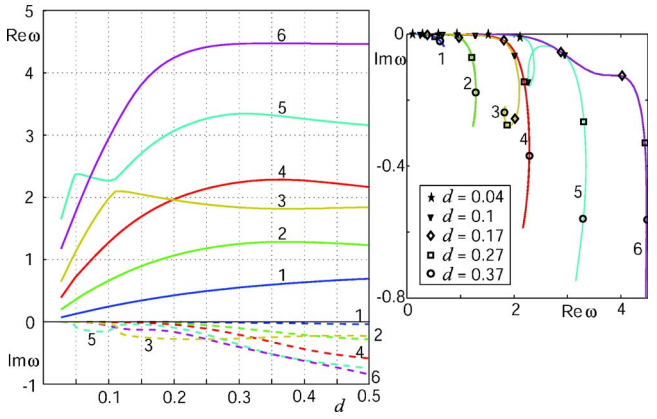


FIG. 2. (Color online) The six complex natural frequencies ω_n closest to the real axis for a half-plane as functions of crack depth d (left) and their trajectories in the ω plane (right).

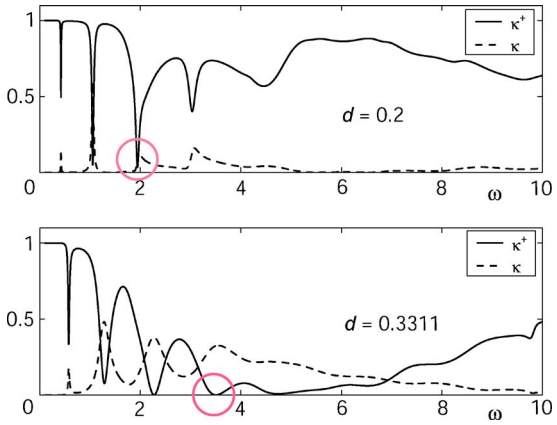


FIG. 3. (Color online) Frequency dependence of the Rayleigh wave transmission and reflection coefficients $\kappa^{\pm}$ at different depths d.

conservation law entails the balance equalities $\kappa^+ + \kappa^- + \kappa_b = 1$ and $\mu^+ + \mu^- + \mu_b = 1$. In a layer $E_b, \kappa_b, \mu_b \equiv 0$. The difference between $\kappa^{\pm}$ and $\mu^{\pm}$ becomes apparent taking into consideration that for a structure without defects $\kappa^+ = 1$ and $\kappa^- = 0$ while $\mu^+ = \mu^-$. With a full blocking $E^+ \equiv 0$, so both κ^+ and μ^+ are equal to zero.

As soon as the Umov-Poynting vector $\mathbf{e}$ is determined at each point $\mathbf{x}$ of the domain, the continuous vector field $\mathbf{e}(\mathbf{x})$ determines in the $\mathbf{x}$ plane noncrossing curves $\mathbf{x} = \mathbf{x}(s)$ to which $\mathbf{e}(\mathbf{x})$ is tangential at any point:

$$\frac{d\mathbf{x}}{ds} = \mathbf{e}(\mathbf{x})/|\mathbf{e}(\mathbf{x})|. \tag{10}$$

Here $\mathbf{x} = \mathbf{x}(s)$ is a parametric equation of a curve and s is a natural parameter. Every specific curve passing through a given point x_0,

$$\mathbf{x}(0) = x_0, \tag{11}$$

is determined uniquely from the Cauchy problem in (10) and (11). These curves, referred to as energy streamlines (similarly to streamlines in fluid dynamics), are used for visualization of time-averaged energy transfer in a harmonic wave field,[14,10] while energy concentration is illustrated by the space distribution of the power density amplitude $|\mathbf{e}(\mathbf{x})|$.

Whereas the incident field $\mathbf{u}_0$ is known, $\mathbf{u}_c$ is expressed in terms of the unknown c.o.d. $\mathbf{v}(\mathbf{x})$, which is determined by a Wiener-Hopf integral equation

$$\mathcal{L}\mathbf{v} \equiv \int_{-a}^{a} l(x-\xi)\mathbf{v}(\xi)d\xi = \mathbf{f}(x), \quad |x| < a, \tag{12}$$

derived by substituting $\mathbf{u} = \mathbf{u}_0 + \mathbf{u}_c$ in the form of (5) and (6) into the homogeneous boundary condition (1) on the crack sides. The hypersingular kernel of this equation $l(x)$ is derived in terms of its Fourier symbol $L(\alpha)$: $l = \mathcal{F}^{-1}[L]$; a specific analytical representation for L is given in the Appendix [see Eq. (A5)]. The right-hand side $\mathbf{f}(x) = -\boldsymbol{\tau}_0(x, -d)$, where $\boldsymbol{\tau}_0$ is the stress vector at a horizontal area element associated with the incident field $\mathbf{u}_0$.

The resonance poles ω_n are spectral points of the integral operator $\mathcal{L}$. Numerically they are approximated by the roots of the characteristic equation $\det A(\omega) = 0$, where $A(\omega)$ is the matrix of a linear algebraic system $A\mathbf{t} = \mathbf{f}$ into which Eq. (12) is discretized. There have been developed and implemented a variety of discretization methods such as expansion in terms of splines, orthogonal polynomials, or normal modes, reduction to infinite algebraic systems in terms of residuals with their stable truncation (for layered structures), the Wiener-Hopf factorization technique, and so on. Each of them is most effective within some specific range of input parameters. Due to their overlapping, these approaches provide reasonable results for both small and large frequencies and crack size.

The reliability of the numerical results are controlled by numerically checking the boundary conditions and the power balance, as well as by comparing with known independent results, e.g., with Ref. 17 for arbitrarily inclined subsurface cracks in a half-plane and with Ref. 18 for cracks in a layer.

III. THE HALF-PLANE

Let us consider the diffraction of a classical Rayleigh wave $\mathbf{u}_{0,R}$ (9) by a subsurface crack. Within this section we use dimensionless parameters set by the equalities $a = 1$, $v_s = 1$, $\rho = 1$, and Poisson's ratio $\nu = 1/3$. That is, the unit of length is fixed by the crack's semi-width a and the dimensionless angular frequency is $\omega = 2\pi f a / v_s$, where f, a, and v_s are dimensional frequency, semi-width, and S wave velocity.

First of all, let us consider the location of the poles ω_n in the complex ω plane. Figure 2 gives the six natural frequencies closest to the real axis as functions of the depth d [plots

TABLE I. The six poles ω_n closest to the real axis for the same depths d as in Fig. 3.

d	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
0.2	$0.43 - i0.004$	$1.07 - i0.02$	$1.96 - i0.27$	$1.97 - i0.04$	$3.07 - i0.09$	$4.24 - i0.16$
0.3311	$0.58 - i0.02$	$1.27 - i0.13$	$1.83 - i0.25$	$2.26 - i0.27$	$3.32 - i0.46$	$4.45 - i0.47$

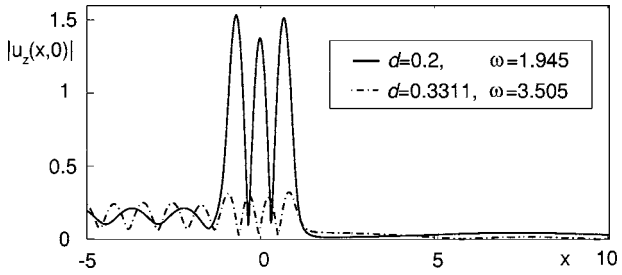


FIG. 4. Surface displacement with resonance and antiphase screening.

Re $\omega_n(d)$, Im $\omega_n(d)$ in the left part and the related trajectories $\omega_n(d)$ in the right part of the figure]. The plots show that ω_n approach very close to the real axis as $d \to 0$, but, nevertheless, they do not touch the axis. Therefore, we believe that no real poles ω_n are possible for a half-plane with a crack.

To illustrate their influence on the Rayleigh wave transmission, Fig. 3 presents examples of typical frequency behavior of the coefficients $\kappa^{\pm}$ with two different crack depths d. At $d=0.2$ one can see narrow dips in the $\kappa^{+}(\omega)$ plot indicating sharp (resonant) screening of the Rayleigh wave propagation at $\omega=0.43$, 1.07, and 1.97. It would appear reasonable that this occurs at $\omega=\mathrm{Re}\,\omega_n$ for the poles ω_n close to the real axis.

Indeed, Table I, displaying values ω_n for the depths under consideration, shows the same real and very small imaginary parts for $\omega_n, n=1,2,4$ ($d=0.2$). The poles ω_5 and ω_6 also cause dips in the curve $\kappa^{+}(\omega)$, which, however, are not so sharp as these points are further away from the real axis.

At $d=0.331$, wide and deep depressions with minima at $\omega=3.5$ and 4.8 are added to the dips. There are no complex poles ω_n close to these frequencies, therefore these blocking frequencies are hardly caused by a resonance effect. To illustrate this, we compare surface displacements and energy concentration near the crack at such frequencies with the resonance ones. Examples of the comparison for the two blocking situations marked in Fig. 3 by circles are given in Figs. 4–6.

Figure 4 shows the amplitude of the vertical surface displacement $|u_z|$ at a sharp screening frequency ($\omega=1.945$, $d=0.2$, solid line) and at the minimum of a wide depression

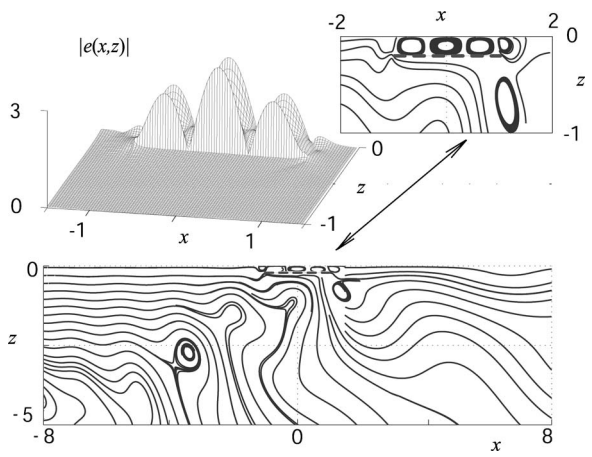


FIG. 5. Energy streamlines and power density distribution with a resonant Rayleigh wave trapping ($d=0.2$, $\omega=1.945$).

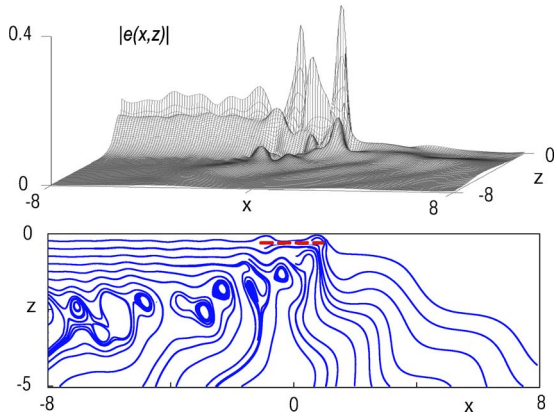


FIG. 6. (Color online) The same with a non-resonant blocking ($d=0.331$, $\omega=3.505$).

($\omega=3.505$, $d=0.331$, dashed line). In both cases the coefficient κ^{+} is very small, therefore, at the right of the crack we see a shadow zone with a minimal surface oscillation decreasing as $x \to \infty$. Above the crack, in contrast, the surface oscillation exhibits sharp growth in the first case, while it remains at the same level as in the incoming zone for the second kind of wave blocking. The growth indicates resonance trapping, which is confirmed by the picture of energy localization in Fig. 5.

The energy streamlines in these cases shown in Figs. 5 and 6, respectively, are plotted for the energy flow in the total field $\mathbf{u}=\mathbf{u}^{+}_{0,R}+\mathbf{u}_c$, where $\mathbf{u}^{+}_{0,R}$ is the Rayleigh wave (9) coming from infinity. We should remark that the density of streamlines has no meaning; they only show the trajectories of an averaged energy transfer in a time-harmonic field. The power is specified by the energy density $|\mathbf{e}(\mathbf{x})|$ given in the upper subplots.

In both cases the presence of the crack turns the flow downward, indicating considerable transformation into body waves. As the main difference, there are several energy vortices between the crack and the surface in the first case (see enlarged upper right subplot in Fig. 5). The upper left subplot shows that much more energy density is accumulated in these vortices than in the surrounding field. Hence, it is safe to say that this is a classical trapping with energy localization. The vortices block the way for energy transmission, similarly to the situation in a waveguide with a surface obstacle.[10] They leave a space only for a weak energy flow going to infinity along the surface, which results in very small but nonzero κ^{+} in this case. It might be worthwhile to remark that such powerful vortices, nevertheless, do not upset the total power balance, since they do not contribute to the energy outflow to infinity.

In the second case (Fig. 6) no vortices accumulating energy appear near the crack, but the flow is turned downward so that within numerical accuracy $\kappa^{+}=0$. In a previous paper[11] we explain this nonresonant stopping by the antiphase annihilation of the incoming and diffracted Rayleigh waves $\mathbf{u}^{+}_{0,R}$ and $\mathbf{u}^{+}_{c,R}$ [coefficient t^{+}_{1} in (9) becomes equal to -1 in this case].

Searching for other possible full blocking configurations, we plot the surface κ^{+} as a function of ω and d. Figure

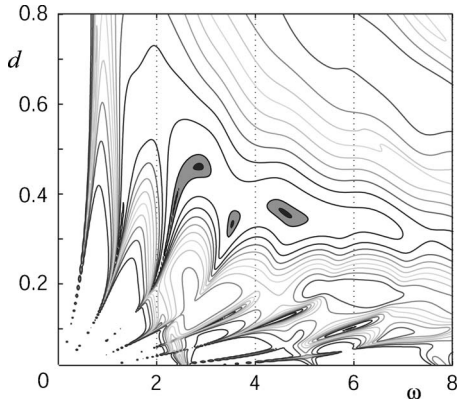


FIG. 7. Level lines of the surface $\kappa^+(\omega,d)$ for the incoming Rayleigh wave.

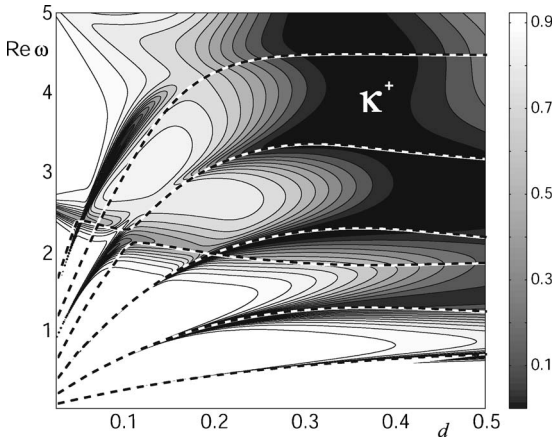


FIG. 8. Curves Re ω_n(d) as compared with the canyon structure of the surface $\kappa^+(d,\omega)$ (half-plane).

7 shows the level lines of this surface (its general appearance is also seen in the left subplot of Fig. 9). A characteristic feature of the relief is a deep and wide valley extending from northwest to southeast, approximately from $(\omega,d)=(2,0.6)$ to (7, 0.3). Gray and black colors indicate zones in which $\kappa^+\leqslant 0.01$ and $\kappa^+\leqslant 0.001$, respectively.

The global minima of the function κ^+ are sought in these zones, since a deviation from the valley leads only to an increase in κ^+. This increase, with κ^+ tending to unity as $\omega\to\infty$ or $d\to\infty$, is obviously associated with the fact that with increasing depth-to-wavelength the wave interaction with the crack ceases. A monotonic increase in κ^+ is violated only close to the axis $d=0$, where the relief becomes very irregular due to almost real ω_n. Therefore, the presence of global minimum points in the band of small d also cannot be totally excluded.

Using a numerical minimization of the function $\kappa^+(\omega,d)$ we find several discrete points in the (ω,d) plane at which the transmission is completely blocked ($\kappa^+\equiv 0$!). The coordinates of such points for different Poisson's ratios ν are listed in Table II. The assertion about theoretically complete blocking relies on the numerical results which stably converge to these points with different discretization accuracy. At these points they yield $\kappa^+\approx 10^{-12}-10^{-14}$ while in the vicinity $\kappa^+\approx 10^{-2}-10^{-4}$.

As for the blocking due to resonance poles, it appears in narrow canyons, or even slits, in the $\kappa^+(\omega,d)$ relief. This is illustrated in Fig. 8 where the curves Re $\omega_n(d)$ from Fig. 2 (dashed lines) are put on the surface $\kappa^+(d,\omega)$ (rotated and mirrored part of Fig. 7). We should remark that the canyons are narrower the closer ω_n come to the real axis. That is why they are practically not seen under the curves Re ω_n as $d\to 0$, but they are present, becoming narrower and deeper the smaller Im ω_n is. They may be seen as sharp dips on the section profiles $\kappa^+(\omega)$, $d=$const (see, e.g., Fig. 3). A certain mismatch of the higher curves with the canyons is apparently due to the presence of the next neighbor poles not shown in this figure.

Thus, there exist two mechanisms of a Rayleigh wave screening by a horizontal subsurface crack. The first one is related to a resonance in the crack-surface system, but it does not ensure complete blocking. The second mechanism results from nonresonant antiphase annihilation of the incident wave and the wave arising due to diffraction. It does not entail any energy trapping and localization, providing, however, complete transmission stopping for certain discrete combinations of the crack depth and frequency.

TABLE II. The (ω,d) points of the Rayleigh wave complete nonresonant blocking for different Poisson's ratios ν.

ν	1	2	3	4
0.1	2.025, 0.433	⋯	⋯	⋯
0.2	2.146, 0.400	3.009, 0.441	3.202,0.362	⋯
0.3	2.285, 0.374	2.895, 0.465	3.427, 0.338	⋯
0.3333	2.337, 0.366	2.849, 0.462	3.509, 0.333	⋯
0.4	2.450, 0.353	2.739, 0.431	3.689, 0.325	5.036, 0.336

IV. THE ELASTIC LAYER

Unlike the half-plane, where the surface $\kappa^+(\omega,d)$ provides exhaustive information for the Rayleigh wave transmission, it is hardly possible to give a complete description even for a homogeneous layer. First of all, this is due to a wider range of input parameters (additional thickness H and a richer variety of traveling waves). In addition, the patterns themselves become much more complicated due to multiple recurrent reflections among the crack and the two sides. As an example, the surfaces $\kappa^+(\omega,d)$ for the half-plane considered above ($H=\infty$) and for the layer of thickness $H=1$ with a crack of semi-width $a=1$, excited by the first antisymmetric mode a_0: $\mathbf{u}_0^+=\mathbf{a}_1^+e^{i\zeta_1 x}$, are confronted in Fig. 9. Instead of one

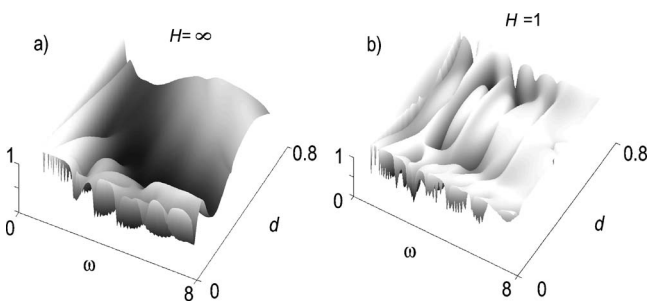


FIG. 9. Typical appearance of the surface $\kappa^+(\omega,d)$ for a half-plane (a) and a layer (b).

wide valley we see in the second case a wrinkled surface with a set of slits throughout the entire domain of input parameters.

Considering a layer, it is natural to take its thickness H as a unit of length. Henceforth, within this section $H=1$, $\omega=2\pi fH/v_s$, and a may differ from unity; Poisson's ratio $\nu=1/3$ as above. To avoid a separate consideration for different incoming modes, in some examples we take a multimode incident field $\mathbf{u}_0^+$ coming from a given vertical point source $\mathbf{q}_0=(0,\delta(x-x_0))$ (in the examples below $x_0=-5$). Consequently, the energy distribution coefficients $\mu^{\pm}$ instead of $\kappa^{\pm}$ are considered in such a case. Since the poles ω_n are independent of $\mathbf{u}_0$ [as spectral points of the integral operator $\mathcal{L}$ in Eq. (12)] their manifestation in the transmission relationships (canyons) should be the same with any incident field.

Typical examples of the κ^+ dependence on two input parameters with a fixed third one are shown in Figs. 10 and 11 by means of level lines and a gray scale (here the first antisymmetric mode a_0 is taken as $\mathbf{u}_0^+$). The narrow black strips determine the parameters of resonance blocking. They are analogous to the curves of trapped modes obtained for a fluid layer with a horizontal obstacle.[1] As above, a search for the full blocking parameters was organized by numerical minimization of the function κ^+ taking start points at these strips (canyons' bottom).

As an example, the minimization from the initial point, marked in Figs. 10 and 11 by a white cross, results in the values $\omega=0.9692$, $a=1.4903$, and $d=0.3994$ for which $\kappa^+<10^{-6}$. Nevertheless, we have to note, this is not a case of pure real resonance, because the nearest spectral point ω_n still has a very small imaginary part here. Similar points were found in other parts of the (ω,a,d) space searching over the canyons' bottom. The resonance character of blocking at these points is evident from the sharp local growth of surface displacements above the crack and stress intensity factors at its tips, as well as from energy vortices appearing near the crack. In contrast to the previous example (Fig. 5), with a not distant source these vortices may not accumulate a great power, just turning away to the left the total energy flow incoming from the source [see Fig. 12 for the minimum point (ω,a,d) mentioned above]. The horizontal profile of the energy density surface $|\mathbf{e}(x,z)|$ given in the upper subplot

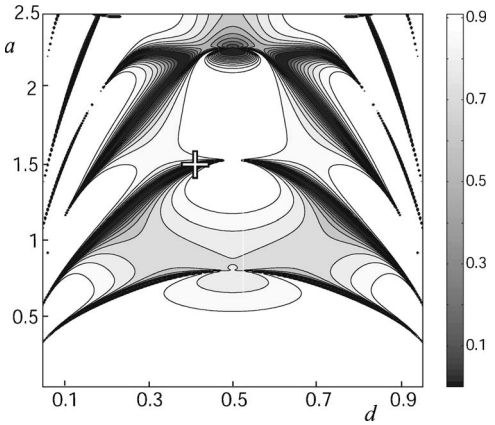


FIG. 10. Transmission coefficient κ^+ as a function of a and d at the fixed frequency $\omega=0.9692$.

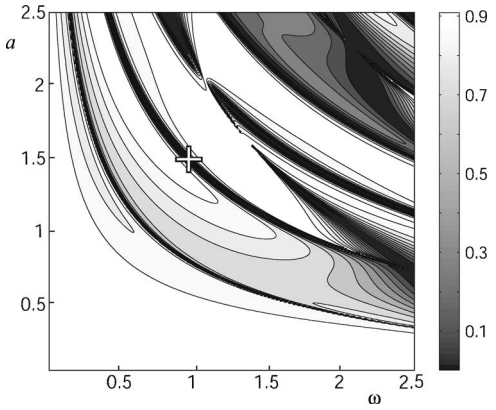


FIG. 11. Transmission coefficient κ^+ as a function of ω and a at the fixed depth $d=0.3994$.

shows very small energy amplitudes in the vortex zone. On the other hand, with a distant source or normal modes coming from infinity, a vortex energy concentration at the crack zone, similarly to the half-plane example above, is obtained for a layer as well. It is noted that the vortices, blocking energy transmission by the crack, completely disappear as soon as the point of input parameters displaces slightly across the canyon, getting up from its bottom on the brink.

Apparently, the canyons appear due to some near-real ω_n. Similarly to Fig. 8, this is demonstrated by Fig. 13, where plots of $\mathrm{Re}\,\omega_n(d)$ are put on the level line relief $\kappa^+(d,\omega)$ with $a=1$, obtained for the incident antisymmetric mode a_0, which is a Rayleigh wave analog in the layer. For small d the picture becomes similar to the related part of Fig. 8. Obviously, when the crack approaches the upper surface, the influence of the bottom side becomes negligible. For $0.5\leqslant d\leqslant 1$ the picture is mirror symmetric due to the layer symmetry.

In contrast to the half-plane, however, $|\mathrm{Im}\,\omega_n|$ do not increase monotonously together with d. Some of the poles turn upward and come to the real axis again as $d\to 0.5$. Similarly to Fig. 8, the canyons are very narrow and practically not seen under the curves $\mathrm{Re}\,\omega_n$ when ω_n are close to the real

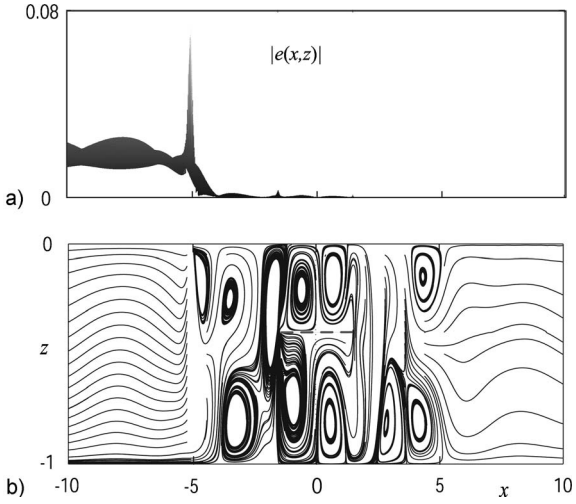


FIG. 12. A horizontal profile of the power density surface $|\mathbf{e}(x,z)|$ (a) and energy streamlines (b) from a point source at $x_0=-5$ for the layer under resonance conditions ($\omega=\mathrm{Re}\,\omega_n$, $|\mathrm{Im}\,\omega_n|\ll 1$).

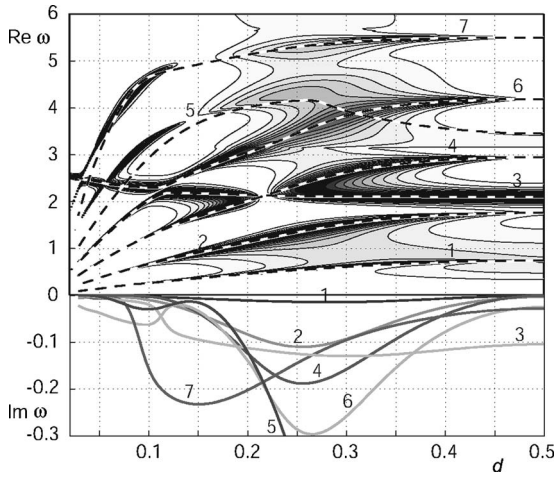


FIG. 13. Comparison of pole curves Re $\omega_n(d)$, $a=1$ with the canyon structure of the surface $\kappa^+(d,\omega)$ for the layer; the lower part depicts Im $\omega_n(d)$ in the tenfold enlarged scale.

axis. But they are present, becoming deeper the smaller Im ω_n become, so that in the limit $\kappa^+=0$ as Im $\omega_n=0$ (i.e., at a pure real resonance frequency).

To be certain of the existence of the real poles, we have followed their paths in the complex ω plane varying a at $d=0.5$. At certain points of such paths the poles touch the real axis. Examples of such touching accompanied by complete resonance stopping ($\kappa^+=\mu^+\equiv 0$) have been definitely detected (Fig. 14). Values of several pure real poles ω_n found by such a way for different a at the depth $d=0.5$ are listed in Table III together with the resonance crack widths a (the latter are also shown in Fig. 14 by arrows).

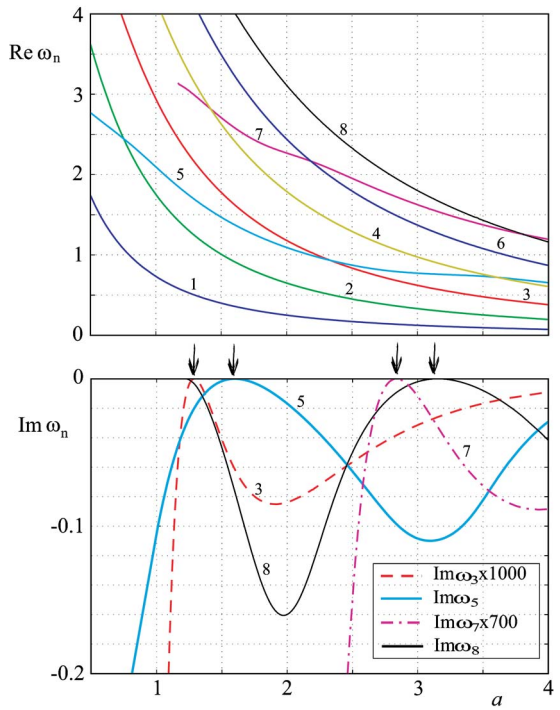


FIG. 14. (Color online) Plots Re $\omega_n(a)$ (upper) and Im $\omega_n(a)$ (lower) yielding examples of the natural resonance frequencies ω_n coming up to and touching the real axis Im $\omega=0$; $d=0.5$.

TABLE III. Real poles ω_n and associated resonance crack widths a for a median crack ($d=0.5$).

n	3	5	7	8
ω_n	2.17	1.38	1.95	1.54
a	1.29	1.60	2.84	3.15

Hence, for an elastic layer with a horizontal crack there exist real resonance points ω_n [more precisely, real spectral points in the space (ω,a,d)] at which the waveguide is locked up completely. Harmonic excitation at such frequencies gives rise to unlimited growth of oscillation amplitude and stress intensities at the crack edges as $t\to\infty$.

On the other hand, the resonance poles ω_n contribute in a transient motion $\mathbf{u}(\mathbf{x},t)$ caused by a limited pulse in time $\boldsymbol{\tau}|_{z=0}=\mathbf{q}(x)f(t)$ as residues:

$$\mathbf{u}_n(\mathbf{x},t)=\mathrm{Re}[\mathbf{u}_n(\mathbf{x})e^{-i\omega_n t}],$$

$$\mathbf{u}_n(\mathbf{x})=-2i\mathrm{res}\mathbf{u}(\mathbf{x},\omega)|_{\omega=\omega_n}F(\omega_n),$$

$$F(\omega)=\int_0^\infty f(t)e^{i\omega t}dt.$$

The pulses related to complex ω_n decay exponentially, while a real resonance pole yields undamped harmonic oscillation, indicating the trapping of wave energy from transient signals passing by the crack.

The space distribution of the trapped oscillation is described by the complex amplitude vector $\mathbf{u}_n(\mathbf{x})$, i.e., by the displacement eigenform associated with the real spectral point ω_n. It is calculated by the same way as the reflected field $\mathbf{u}_c$, but via the eigenvector $\mathbf{t}_n$: $A(\omega_n)\mathbf{t}_n=0$ instead of vector $\mathbf{t}$ obtained from the inhomogeneous system $A\mathbf{t}=\mathbf{f}$ for the usual (nonresonant) frequencies. As is to be expected, the eigenfields $\mathbf{u}_n(\mathbf{x})$ exhibit strong space localization near the obstacle. As an example, Fig. 15 depicts the space distribution of the amplitude $|\mathbf{u}_3(x,z)|$ related to the real spectral point $\omega_3=2.169$, $a=1.29$ (eigenvector $\mathbf{t}_3$ is normalized by the condition $|\mathbf{t}_3|=1$). Eigenforms for all other spectral points from Table III show similar wave localization around the crack.

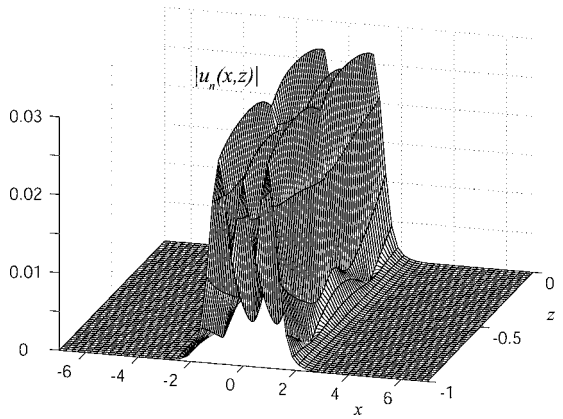


FIG. 15. The amplitude of the displacement eigenform $|\mathbf{u}_n|$ at the real resonance frequency $\omega_3=2.17$; $a=1.29$.

V. CONCLUSION

Traveling wave propagation in an elastic structure with a horizontal subsurface striplike crack may be stopped at certain frequencies due to its diffraction by the crack. There are two kinds of blocking mechanisms:

(1) nonresonant antiphase annihilation of the incoming wave with the diffracted one and
(2) resonant capturing of the energy of incident waves by the eigenforms associated with the natural resonance frequencies ω_n resulting from the crack-structure wave interaction; such blocking is featured by the wave energy localization in the energy vortices formed near the crack when the oscillation frequency ω becomes close to ω_n.

The first mechanism is peculiar for the Rayleigh wave blocking in a half-plane, while it is practically impossible to meet the conditions of antiphase annihilation with two and more traveling waves excited in a free layer waveguide. It provides a full Rayleigh wave blocking ($\kappa^+=0$) at several (ω,d) pairs (depending on Poisson's ratio ν). The second mechanism is inherent in both half-plane and layer cases, but the examples of pure real discrete spectral points ω_n entailing the full blocking by the resonance trapping mechanism have been found only for the layer.

ACKNOWLEDGMENTS

The authors are thankful to Professor V. A. Babeshko and Professor D. A. Indeitsev for fruitful discussions. This work is supported by the Russian Foundation for Basic Research, Project Nos. 04-01-00801 and 06-01-96607.

APPENDIX: ANALYTICAL REPRESENTATION FOR GREEN'S MATRICES *K* AND *M*

The calculation of a harmonic wave field in a layered structure by means of the Fourier transform with respect to the horizontal coordinate is brought to a multipoint boundary value problem for ordinary differential equations (ODEs) given in the sublayers. Then the problem is reduced to a linear algebraic system arisen from the boundary conditions with respect to unknown constants of the ODE general solutions. With a small number of sublayers this system may be solved explicitly. Hence the Fourier transform of the field in the structure can be derived analytically.

In this way the matrices K and M used in Eq. (5) are obtained via explicit representations for the sublayers S_1: $-d\leqslant z\leqslant 0$, $|x|<\infty$, and S_2: $-H\leqslant z\leqslant -d$, $|x|<\infty$, into which the crack line $z=-d$ divides the initial domain $-H\leqslant z\leqslant 0$, $|x|<\infty$.

The integral transform $\mathbf{U}_h(\alpha,z)=\mathcal{F}[\mathbf{u}_h(x,z)]$ of a wave field $\mathbf{u}_h$ in a layer of thickness h: $|x|<\infty$, $-h\leqslant z\leqslant 0$, caused by loads $\mathbf{q}^+(x)$ and $\mathbf{q}^-(x)$ applied to the surfaces $z=0$ and $z=-h$, respectively, may be written in the following form:

$$\mathbf{U}_h(z)=K_h^+(z)\mathbf{Q}^+ + K_h^-(z)\mathbf{Q}^-, \quad -h\leqslant z\leqslant 0. \tag{A1}$$

Here $\mathbf{Q}^\pm=\mathcal{F}[\mathbf{q}^\pm]$, $K_h^\pm=\|K_{h,ij}^\pm\|_{i,j=1}^2/\Delta_h$,

$$K_{h,11}^+ = -\sigma_2[\alpha^2\gamma^2(\sigma_{12}s_2+\gamma^2 s_1)-\alpha^2\gamma^4 cs_{21}+\alpha^4\sigma_{12}sc_{21} - \gamma^2\alpha^2\sigma_{12}cs_{12}+\gamma^6 sc_{12}],$$

$$K_{h,12}^+ = -i\alpha[-\sigma_{12}\gamma^2(\alpha^2 c_1+\gamma^2 c_2)-\alpha^2\sigma_{12}^2 ss_{12} + \sigma_{12}\gamma^4 cc_{12}+\gamma^2\alpha^2\sigma_{12}cc_{21}-\gamma^6 ss_{21}],$$

$$K_{h,21}^+ = -i\alpha[\sigma_{12}\gamma^2(\alpha^2 c_2+\gamma^2 c_1)-\gamma^4\sigma_{12}cc_{21}+\sigma_{12}^2\alpha^2 ss_{21} - \gamma^2\alpha^2\sigma_{12}cc_{12}+\gamma^6 ss_{12}],$$

$$K_{h,22}^+ = \sigma_1[-\alpha^2\gamma^2(\sigma_{12}s_1+\gamma^2 s_2)+\alpha^2\gamma^4 cs_{12}-\alpha^4\sigma_{12}sc_{12} + \gamma^2\alpha^2\sigma_{12}cs_{21}-\gamma^6 sc_{21}],$$

$$\Delta_h = 2\mu[-2\alpha^2\sigma_{12}\gamma^4-(\gamma^8+\alpha^4\sigma_{12})sh(\sigma_1 h)sh(\sigma_2 h) + 2\alpha^2\sigma_{12}\gamma^4 ch(\sigma_1 h)ch(\sigma_2 h)],$$

$$s_n = sh(\sigma_n z), \quad c_n = ch(\sigma_n z), \quad n=1,2,$$

$$cs_{mn}=ch(\sigma_m h)sh(\sigma_n(h+z)), \quad m,n=1,2. \tag{A2}$$

$\gamma^2=\alpha^2-\kappa_2^2/2$, $\sigma_n=\sqrt{\alpha^2-\kappa_n^2}$, $\kappa_n=\omega/v_n, n=1,2$, $v_1=\sqrt{(\lambda+2\mu)/\rho}$, and $v_2=\sqrt{\mu/\rho}$ are P and S wave velocities in the layer material; cc_{mn}, ss_{mn}, sc_{mn} are of analogous form to sc_{mn} in which c corresponds to the hyperbolic cosine ch and s to the sine sh with the same arguments. The elements of matrix K_h^- are expressed via the elements of K_h^+:

$$K_{h,ij}^-(z)=K_{h,ij}^+(-z-h)(-1)^{i+j-1}, \quad i,j=1,2.$$

For brevity, we do not show the dependence on the Fourier parameter α in Eqs. (A1) and (A2) above and furthermore below. We only indicate the z dependence where it occurs.

For a half-plane ($h=\infty$) the matrix K_∞^- is unnecessary, while K_∞^+ is of much simpler form:

$$K_{\infty,11}^+=\sigma_2(\alpha^2 e_1-\gamma^2 e_2), \quad K_{\infty,12}^+=i\alpha(\gamma^2 e_1-\sigma_1\sigma_2 e_2),$$

$$K_{\infty,21}^+=i\alpha(\sigma_1\sigma_2 e_1-\gamma^2 e_2), \quad K_{\infty,22}^+=\sigma_1(-\gamma^2 e_1+\alpha^2 e_2),$$

$$\Delta_\infty=2\mu(-\gamma^4+\alpha^2\sigma_1\sigma_2), \quad e_n(z)=e^{-\sigma_n z}, \quad n=1,2. \tag{A3}$$

The matrices K and M are expressed through the matrices $K_d^\pm$ and $K_{H-d}^\pm$. Indeed, for the layer under consideration

$$\mathbf{U}(z)=\begin{cases} K_d^+(z)\mathbf{Q}_0+K_d^-(z)\mathbf{Q}, & z\in S_1,\\ K_{H-d}^+(z+d)\mathbf{Q}, & z\in S_2. \end{cases} \tag{A4}$$

Here $\mathbf{Q}=\mathcal{F}[\mathbf{q}]$ is the Fourier transform of the unknown stress vector $\mathbf{q}(x)$ at the crack plane $z=-d$. Since $\mathbf{V}=U|_{z=-d-0}-\mathbf{U}|_{z=-d+0}$, Eq. (A4) yields

$$\mathbf{Q}=L\mathbf{V}+LK_d^+(-d)\mathbf{Q}_0, \tag{A5}$$

where the matrix

$$L=[K_{H-d}^+(0)-K_d^-(-d)]^{-1}.$$

Hence, in Eq. (5)

$$K(z)=\begin{cases} K_d^+(z)+K_d^-(z)LK_d^+(-d), & z\in S_1,\\ K_{H-d}^+(z+d)LK_d^+(-d), & z\in S_2, \end{cases} \tag{A6}$$

$$M(z)=\begin{cases} K_d^-(z)L, & z\in S_1, \\ K_{H-d}^+(z+d)L, & z\in S_2. \end{cases} \tag{A7}$$

For the half-plane case, obviously, K_{H-d}^+ must be replaced everywhere by K_∞^+.

We should remark that these representations are valid with different material properties in sublayers S_1 and S_2 (interface crack). Moreover, S_1 and S_2 may also be layered structures; in such a case $K_d^\pm$, $K_{H-d}^\pm$ are calculated by certain matrix algorithms.

It is obvious that for a homogeneous layer one can take $K(z)=K_H^+(z)$ of form (A2) instead of Eq. (A6) [the latter may be reduced to (A2) in this case]. Similarly, for a homogeneous half-space $K(z)=K_\infty^+(z)$ of form (A3).

[1]C. M. Linton and D. V. Evans, "Trapped modes above a submerged horizontal plate," Q. J. Mech. Appl. Math. **44**(3), 487–506 (1991).

[2]D. Indeitsev, "Trapping modes of oscillations in an infinitely long waveguide with submerged object in the form of a massive die," J. Acoust. Soc. Am. **105**(2), 1196 (1999).

[3]I. I. Vorovich, "Resonance properties of an elastic inhomogeneous strip," Dokl. USSR **245**(5), 1076–1079 (1979) (in Russian).

[4]I. I. Vorovich and V. A. Babeshko, *Dynamic Mixed Problems of Elasticity for Nonclassical Domains* (Nauka, Moscow, 1974) (in Russian).

[5]C. J. S. Alves and T. Ha Duong, "Numerical resolution of the boundary integral equations for elastic scattering by a plane crack," Int. J. Numer. Mech. Eng. **38**, 2347–2371 (1995).

[6]A. S. Eriksson, "Natural frequencies of a penny-shaped crack with spring boundary condition," Trans. ASME, J. Appl. Mech. **62**(1), 59–63 (1995).

[7]Ye. V. Glushkov and N. V. Glushkova, "Resonant frequencies of the scattering of elastic waves by three-dimensional cracks," J. Appl. Math. Mech. **62**(5), 803–806 (1998).

[8]Ye. V. Glushkov, N. V. Glushkova, and A. V. Ekhlakov, "A mathematical model of the ultrasonic detection of three-dimensional cracks," J. Appl. Math. Mech. **66**(1), 141–149 (2002).

[9]V. A. Babeshko, E. V. Glushkov, N. V. Glushkova, and E. V. Kirillova, "Energy localization under high-frequency resonance conditions," Sov. Phys. Dokl. **37**(8), 443–444 (1992).

[10]E. Glushkov and N. Glushkova, "Blocking property of energy vortices in elastic waveguides," J. Acoust. Soc. Am. **102**(3), 1356–1360 (1997).

[11]E. V. Glushkov, N. V. Glushkova, and E. M. Shapar, "On the Rayleigh Wave Blocking by a Subsurface Crack," Dokl. Phys. **49**(10), 608–613 (2004).

[12]E. V. Glushkov, N. V. Glushkova, and M. V. Golub, "Blocking of traveling waves and energy localization due to the elastodynamic diffraction by a crack," Acoust. Phys. **52**(3), 259–269 (2006).

[13]A. G. Sveshnikov, "The principle of radiation," Dokl. Akad. Nauk SSSR **73**(5), 917–920 (1950) (in Russian).

[14]V. A. Babeshko, E. V. Glushkov, and N. V. Glushkova, "Energy vortices and backward fluxes in elastic waveguides," Wave Motion **16**, 183–192 (1992).

[15]J. D. Achenbach, *Wave Propagation in Elastic Solids* (American Elsevier, Amsterdam, 1973).

[16]P. D. Wilcox, M. J. S. Lowe, and P. Cawley, "Mode and transducer selection for long range Lamb wave inspection," J. Intell. Mater. Syst. Struct. **12**, 553–565 (2001).

[17]J. H. M. T. van der Hijden and F. L. Neerhoff, "Diffraction of elastic waves by a sub-surface crack (in plane motion)," J. Acoust. Soc. Am. **75**(6), 1694–1704 (1984).

[18]G. R. Liu, "A combined finite element/strip element method for analyzing elastic wave scattering by cracks and inclusions in laminates," Comput. Mech. **28**, 76–81 (2002).

Analysis and design of composite/folded variable area perforated tube resonators for low frequency attenuation[a)]

T. Kar,[b)] P. P. R. Sharma,[c)] and M. L. Munjal
Facility for Research in Technical Acoustics, Department of Mechanical Engineering, Indian Institute of Science, Bangalore -560 012, India

(Received 31 August 2005; revised 6 February 2006; accepted 27 March 2006)

Variable area concentric-tube-resonator mufflers may be used effectively for attenuating the odd harmonics. This ensures a broad-band transmission loss (TL) spectrum. In this paper, configurations that consist of variable area resonators within a uniform area muffler shell for enhancing the low-frequency noise attenuation are investigated. The variable area resonator elements within a muffler allow multiple elements, where the sum of their acoustic lengths is more than the geometric length of the muffler. This shifts the overall TL spectrum toward lower frequency. The underlying idea of incorporating the extended inlets and outlets in these resonators, is to reproduce the broadband attenuation. It has the effectiveness of dissipative ducts without the use of any absorptive materials. The results have been validated by means of three-dimensional finite element analysis.
 [DOI: 10.1121/1.2197791]

PACS number(s): 43.20.Hq, 43.20.Mv [DKW] Pages: 3599–3609

I. INTRODUCTION

It is important to design acoustical mufflers so they have broadband transmission loss (TL), including attenuation of low-frequency harmonics. Absorptive material lining is known to offer such a spectral attribute. With absorptive material lined on the inside of the ducts, an absorptive muffler when used on an engine, may get clogged with unburnt carbon particles or undergo thermal cracking. The use of absorptive material is further restricted as the fibrous lining gets blown away by the grazing flow. The plug-muffler, that appears to overcome this problem, has to compromise with the detrimental effect of increased back pressure, the loss in stagnation pressure across the flow junctions, which results in a decreased volumetric efficiency of the engine. So there is a need for muffler configuration with all reactive components that would offer a reasonable low-frequency attenuation for a considerable bandwidth.

Thus, it is possible to attenuate the odd harmonics, in principle the troughs related to the axial modal frequency, by means of variable area resonators.[1,2] This raises the troughs of the transmission loss (TL) spectrum, which is more important than the corresponding peaks because of the implication of the antilogarithmic addition of sound pressure level. Howard *et al.* made use of the rhomboidal shaped resonators that permit longer quarter wave tubes within the fixed lateral space of the reactive stack silencers.[3] They limited their investigation that incorporates variable area resonators to finite element analysis. The longer elements offer an improved low-frequency attenuation.

Multiple resonators can achieve a broadband attenuation, but they do not assure a low frequency attenuation. A folded resonator, restricted to coaxial geometries and uniform area ducts, gives a good low-frequency tonal attenuation. But it does not offer a broadband spectrum.[4] The need to analyze the higher-order modes, except the effect of end corrections, lacks clarity in the light of its applications to curb low-frequency tonal noise. The narrow spectral peaks of the folded resonator fail to cover all the harmonics of the fundamental frequency of the noise source. Multiple chamber resonators are known to offer broadband attenuation for a considerable frequency range.[5,6] But they remain ineffective at lower frequencies.

The slope of the TL spectrum is controlled by the geometrical parameters, i.e., length and area ratio. It is known in the existing literature that the geometric length decides the resonance frequency of a resonator. The longer the element, the lower the resonance frequency. The radial dimensions of the resonator control the amplitude of the spectrum. The acoustic length, rather than the geometric length, of a multiple resonator muffler decides the position of the peaks/troughs at the lower frequency. The position of the first peak on the frequency coordinate depends upon the total acoustic lengths of the resonator elements. For a uniform cross-sectional resonator, the acoustic length equals the geometric length. This is, in fact, not true for a variable area resonator. The length of a muffler is a design constraint. Although it is possible to relocate the position of the spectral troughs at middle and high frequency, any kind of axial sharing between the multiple muffler elements does not help a designer to get a low-frequency attenuation. This is the conventional or classical approach to muffler design.

Thus, to achieve a better low-frequency attenuation, from the conventional design point of view, one needs to increase the slope of the TL spectrum at lower frequency by increasing the amplitude of the spectrum. This may be

[a)]Part of this manuscript has been presented at the Internoise-2005, Rio, Brazil.
[b)]Author to whom correspondence should be addressed. Electronic mail: kar@mecheng.iisc.ernet.in
[c)]Now at the NVH Laboratory, Engineering Research Center, Tata Motors Limited, Pimpri, Pune-411 015, India.

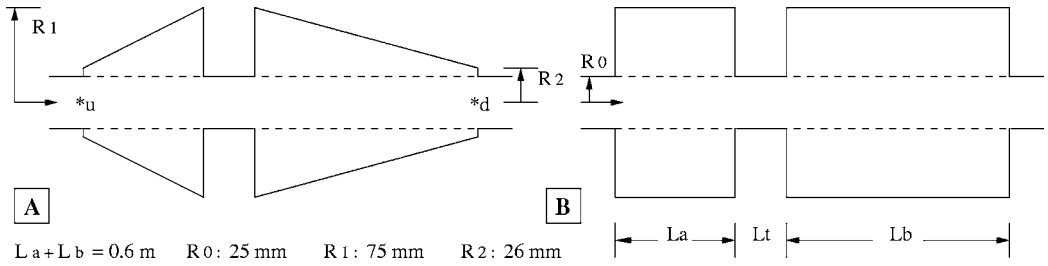


FIG. 1. (a) Schematic diagram of the two-segment conical concentric tube resonator. (b) The corresponding simple CTR.

achieved by increasing the radial dimension of the resonator elements. This, however, keeps the profile of the TL spectrum intact for the plane wave propagation. The purpose could be served if it were possible to shift the peaks toward the lower frequency. To do so, one may need to have longer elements; acoustic elements with longer acoustic lengths. This could be achieved by incorporating variable area resonators.

Here, a couple of configurations have been analyzed. One of them is a composite variable area concentric-tube resonator (C-VCTR); this consists of a tuned conical concentric tube resonator (CCTR) and a simple CTR with a variable area extended inlet. The other one is a variable area folded resonator (VAFR). All the perforated tube resonator elements, involved in the overall silencer elements, have been solved by making use of the matrizant analysis of matrix calculus.[7,8] The results have been validated with their three-dimensional finite element analysis (FEA).

II. SEGMENTED RESONATORS

Here, a brief overview of the segmented resonators and their spectral analysis is given. Resonator lengths L_a, L_b, and L_t of Fig. 1 are among the parameters that govern the optimization process. Transfer matrices for individual sections of the resonator are derived independently and multiplied sequentially to produce the overall transfer matrix.

Let the perforated tube and the outer/annular cavity for a CCTR be denoted by subscripts 1 and 2, respectively. The emphasis is laid on identification of the impact of individual parameters over the TL spectrum and synthesis of an optimized configuration. The discussion hereafter, in connection with the transfer matrix, will be based on the position of the spectral extrema, namely, the troughs and the peaks. The sound attenuation map for a two chamber resonator, separated by an intermediate duct of length L_t=3 cm, of Fig. 1 is shown in Fig. 2. Inlet resonator length L_a is varied while keeping the overall length of the chambers (L_a+L_b=0.6 m) constant.

Figure 2(**L**) shows the sound attenuation map for a two chamber VCTR, whereas the corresponding one for a uniform CTR ($R_1=R_2$) is outlined in Fig. 2(**R**). Here, the positions of the troughs and peaks are emphasized rather than their absolute values. Thus, the associated radial dimensions are of least interest for the present section. Dark lines represent a lower transmission loss when compared to the bright quadrilateral patches. Both the subplots are worked out for a stationary medium and thus, are symmetric about the ordinate for $L_a=L_b$=0.3 m. This sound attenuation map is least descriptive about the actual TL values. The intersecting points of these dark spectral lines are the absolute troughs, where the harmonics of the characteristic resonance frequencies of both the chambers concur. These resonance lines are more acute for a uniform CTR than the VCTR as the latter attenuates the odd harmonics. This attribute reflects the raised troughs for the conical chambers. The VCTR has a shorter effective length, attributed to the frequency-dependent characteristic impedance for a variable area duct, for the first axial modal resonance frequency of the individual ducts and thus, the positions of the intersecting points appear at relatively higher frequencies when compared to those of CTR.

The corner points (A-B-C-D) of the quadrilateral are absolute minima and represent zero transmission loss for the CTR. In contrast, the minima for a VCTR have nonzero TL

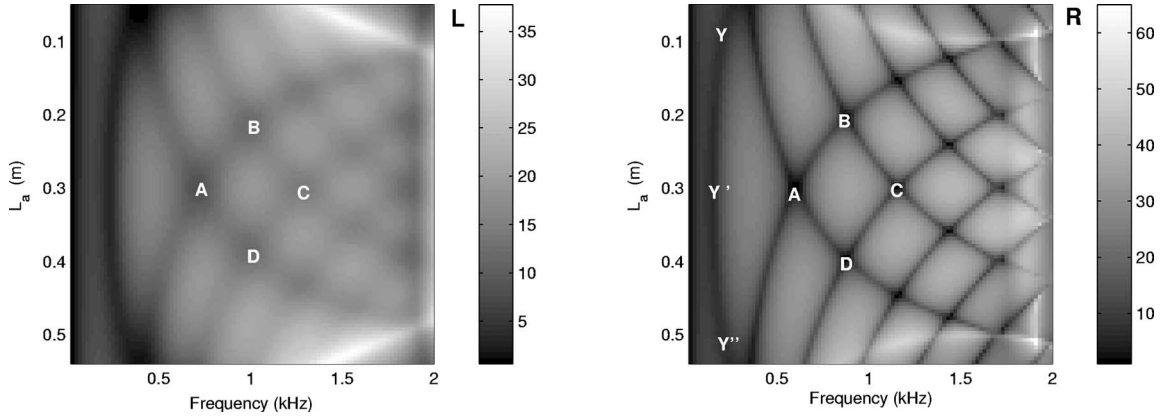


FIG. 2. Sound attenuation map for a two chamber VCTR (**L**) and the corresponding one for a two chamber simple CTR (**R**).

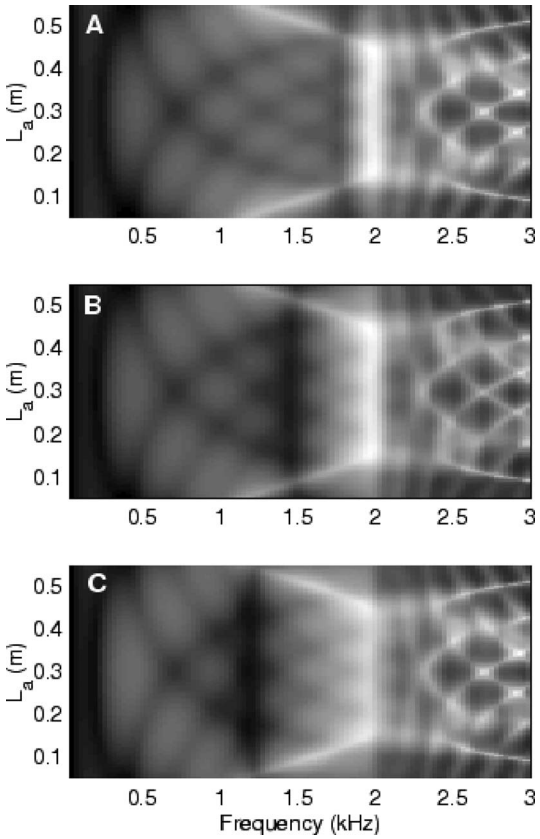


FIG. 3. Effect of the duct length L_t over the sound attenuation map of a VCTR of Fig. 1. (a) 3 mm, (b) 6 mm, and (c) 9 mm.

values. The position of an extremum is independent of the radial dimensions of the resonators as long as the ratio of the annular cross sectional areas at the ends for each of the cavities is kept constant. The line $(Y-Y'-Y'')$ of Fig. 2(**R**) represents the first trough of the TL spectrum and is inherent to any segmented resonators. Its position with respect to the frequency coordinate depends upon the acoustic lengths (length corresponding to the fundamental resonance frequency) of the segments. The overall axial dimension is one of the dimensional constraints. Henceforth, the task is cut out to make an optimized choice of parameters for a broadband TL around a given frequency range. Given a frequency range $[f_1, f_2]$ for sound attenuation, one would search for the best attainable muffler configuration within the physical constraints. It is desirable to avoid the sharp troughs between $[f_1, f_2]$. The resonance effect, characterized by the black band of the intermediate short duct, appears before the primary resonance frequency in Fig. 3.

III. LOW FREQUENCY ATTENUATORS

Unlike its absorptive counterpart, the spectrum of a reactive muffler is characterized by several extrema. Although the sound attenuation over a wide frequency range depends upon the overall volume of the silencer chamber, yet the position of such attenuation peaks may be designed as per need. The conventional resonator elements of Fig. 1 are not very effective at lower frequencies and thus fail to achieve a remarkable attenuation at the firing frequency. The length of the individual segments of the resonator of Fig. 1 reduces

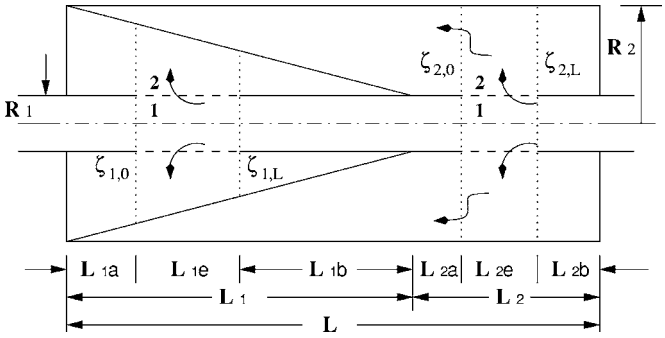


FIG. 4. Schematic diagram of a composite VCTR (C-VCTR), a configuration for low frequency noise attenuation.

due to its segmentation and thus the position of the fundamental peak (first peak) drifts away to higher frequency. Segmental lengths of the resonator may be increased within an overall length of the muffler by incorporating VCTRs and thus improving the low frequency performance of a muffler.

A. Composite variable area concentric-tube resonator (C-VCTR)

One such configuration is shown in Fig. 4, where the muffler element consists of a VCTR of length $L_1=L_{1a}+L_{1e}+L_{1b}$, and a uniform CTR of length $L_2=L_{2a}+L_{2e}+L_{2b}$, with extended inlet of variable area. The two parts are connected by a uniform tube of length $L_t=L_{1b}+L_{2a}$. Before the performance of the composite VCTR is discussed, a brief theoretical overview of the evaluation of the performance spectrum is outlined.

1. Governing equations

The first higher-order mode gets cut on for a shell radius of 0.293 times the wavelength λ, whereas the corresponding value for the first radial mode is 0.61λ. As the present analysis emphasizes the low-frequency attenuation ($f<1$ kz), the wave propagation in a muffler of maximum radial dimension of 75 mm would be within the plane wave cut-off limit. A plane wave propagation in both the interacting variable area ducts, and a grazing mean flow through the inner tube, is assumed. Details of the mathematical modeling may be gleaned from the literature,[2] the rigors of which have been avoided by assuming the absence of temperature gradients and viscous effect of the medium. Thus, the mass continuity equations for an incompressible moving medium in the duct of the variable cross-sectional area $S_i(z)$ may be given in terms of ρ_i, the space average density perturbation over the mean air density ρ_0, and u_i, the corresponding particle velocity perturbation, as

$$\frac{D\rho_i}{Dt}+\rho_0\frac{\partial u_i}{\partial z}+u_i\rho_0\frac{dS_i(z)}{S_i(z)dz}+\frac{2\pi R_1(z)u^*\rho_0}{S_i(z)}=0,\quad i=1 \text{ and } 2, \tag{1}$$

where $S_1=\pi R_1^2$ and $S_2(z)=\pi[R_2^2(z)-R_1^2]$ are the effective cross-sectional areas of the tube and the cavity, respectively. The particle velocity through the perforation u^* transfers the

acoustic energy along the radial direction to resonate the cavity. It is related to the pressure difference across the perforated plate as

$$u^* = \frac{p_1(z) - p_2(z)}{\rho_0 c_0 \zeta}, \tag{2}$$

where ζ is the dimensionless specific acoustic impedance of the perforate and c_0 is the speed of sound in air. The conservation of momentum equations, with the assumption of a dynamically passive interface, are given as

$$\rho_0\left[\frac{Du_i}{Dt}\right] + \frac{\partial p_i}{\partial z} - (\rho_0 u_i + \rho_i U_i)\frac{U_i dS_i(z)}{S_i(z)dz} = 0, \quad i = 1 \text{ and } 2. \tag{3}$$

The time dependence of all variables may be taken as harmonic ($e^{j\omega t}$). Using the condition of isentropicity, the density and the pressure perturbations are related by

$$p_i = \rho_i c_0^2; \quad i = 1,2. \tag{4}$$

2. Solution

A brief description of the matrizant analysis to solve the set of four coupled differential equations, Eqs. (1) and (3), will be rendered here. Equation (4) may be used to replace the density perturbations ρ_i by the corresponding acoustic pressures p_i to rearrange the derivative of each state variable as a linear combination of the state variables (p_i and u_i). Thus,

$$\frac{d\{\mathbf{Y}\}_i}{dz} = \sum_{j=1}^{4} [\mho]_{ij}\{\mathbf{Y}\}_j; \quad i = 1 \text{ to } 4, \tag{5}$$

where the state vector $\{\mathbf{Y}\}$ of Eq. (5) may be written in the normalized form

$$\{\mathbf{Y}\} = \begin{bmatrix} p_1 & p_2 & (\rho_0 c_0)u_1 & (\rho_0 c_0)u_2 \end{bmatrix}^T. \tag{6}$$

Details of the process of obtaining the constituent elements of the square matrix $[\mho]$ are left out of the current manuscript; they have been covered in an earlier work.[2,9] There exists a definite integral and thus an approximated solution for the first-order differential equation of Eqs. (5) over a finite range of variables for the continuous functions $[\mho]_{ij}$. The total path of the variable z may be segmented into a finite number $\mathbf{N}$ of intervals of length $L_s = (L/\mathbf{N})$, and the functions for each of these segments are to be independently worked out. Thus, for any such interval that spans (z_0, z_1), the state vectors at the boundaries are related as[2]

$$\{\mathbf{Y}\}_{z=z_0} = \mathbf{e}^{[\mathbf{\Gamma}]}\{\mathbf{Y}\}_{z=z_1}; \quad [\mathbf{\Gamma}] = -\int_{z_0}^{z_1} [\mho] dz, \tag{7}$$

where $(z_1 - z_0) = L_s$. The transfer matrix $\mathbf{e}^{[\mathbf{\Gamma}]}$, that acts as a linear operator between the downstream and upstream state vectors of one segment may be expanded using the Maclaurin's series

$$\mathbf{e}^{[\mathbf{\Gamma}]} = [\mathbf{I}_N] + \sum_{n=1}^{\infty} \frac{[\mathbf{\Gamma}]^n}{n!} \equiv [\mathbf{\Phi}], \text{ (say)}. \tag{8}$$

The transfer matrix $\mathbf{e}^{[\mathbf{\Gamma}]}$ is diagonalized using the corresponding modal matrix $[\mathbf{\Psi}]$ and the eigenvalues of the matrix $[\mathbf{\Gamma}]$ as follows:

$$[\mathbf{\Gamma}] = [\mathbf{\Psi}][\lambda][\mathbf{\Psi}]^{-1}, \quad [\mathbf{\Phi}] = [\mathbf{\Psi}][\mathbf{e}^{\lambda}][\mathbf{\Psi}]^{-1}, \tag{9}$$

where

$$[\mathbf{e}^{\lambda}]_{ik} = \begin{cases} \mathbf{e}^{\lambda_i} & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}, \tag{10}$$

where λ_i is the ith latent root. The process is repeated for all segments and the corresponding transfer matrices are determined and sequentially multiplied to produce the overall transfer matrix

$$[\mathbf{\Omega}] = \left[\prod_{i=1}^{\mathbf{N}} [\mathbf{\Phi}]_i\right]. \tag{11}$$

3. Boundary conditions

A natural option to have broadband attenuation is to add extended inlets (of length L_{1a} and $L_1 + L_{2a}$) and outlets (of length L_{1b} and L_{2b}).[10] The central tube is perforated over a length of L_{1e} and L_{2e}. The overall transfer matrix $[\mathbf{\Omega}]$ of Eq. (11) is derived over the perforated length. Boundary conditions in terms of impedance expressions of the extended cavities are applied to the system matrix. They may be given as

$$\frac{p_2}{u_2} = \begin{cases} \zeta_{i,0}; \\ \zeta_{i,L;} \end{cases} \quad i = 1 \text{ and } 2. \tag{12}$$

Unlike the uniform area cavities, closed form impedance expressions of the variable-area cavities are complex in nature. They have the property of rightward shifting of the resonance frequency at the lower frequencies.[11,12] They may be determined by means of the matrizant method described earlier. Application of boundary conditions of Eq. (12) in succession to the corresponding transfer matrix $[\mathbf{\Omega}]$ leads to respective four-pole parameter transmission matrix $[\mathbf{T}_i]$ that relates the acoustic pressures and particle velocities at the inlet and outlet of the individual muffler component. The sequential multiplication of these matrices produces the overall transmission matrix. Transmission loss (TL) can be evaluated from the absolute sum of the four-pole parameters[5]

$$\text{TL} = 20\log_{10}\left[\frac{1}{2}\left(\frac{S_d}{S_u}\right)^{0.5}\left(\frac{1+M_u}{1+M_d}\right) \times |T_{11} + T_{12} + T_{21} + T_{22}|\right], \tag{13}$$

where S_d and S_u are the cross-sectional areas of, and M_d and M_u are the mean flow Mach numbers at, the downstream end (outlet) and the upstream end (inlet) points, respectively.

The concept of a double-tuned resonator for a VCTR looks unconventional for a variable-area cavity. The theoret-

ical basis of a double-tuned resonator is based upon the resonance frequencies and their harmonics.[13] A little intuition works better than the analytical rigors. The effective/acoustic length (length corresponding to the fundamental resonance frequency) L_e of a variable-area duct is less than its geometric length L_g and their difference $(L_g - L_e)$ depends upon the terminal area ratio (A_0/A_L). This area ratio over a unit length for the extended inlet of length L_{1b} of Fig. 4 is higher than the corresponding value of the extended outlet L_{1a}, whereas for the perforated segment of length L_{1e}, it lies in between. The higher the area ratio, the larger the difference.[2] As a result, the above difference as a fraction of the geometric length $[(L_g - L_e)/L_g]$ for the extended inlet is more than the extended outlet. To double tune the resonator here, the effective/acoustic lengths of the extended outlet and the extended inlet are kept at half and a fourth of the effective length of the VCTR, respectively. The exact analytical expression of the effective lengths for all three parts $(L_{1a}, L_{1b}$, and $L_{1e})$ of the resonator would amount to equating the corresponding differential equations. Here, the extended parts will be considered as single duct cavities, whereas the perforated part is essentially a wave coupling between the inner tube and the surrounding cavity. This would lead to the determination of the correction parameters ϵ_a and ϵ_b associated with the geometric lengths of the extended inlet and the outlet.

$$\mathrm{L}_{1a} = \mathrm{L}_1/4 - \epsilon_a, \quad \mathrm{L}_{1b} = \mathrm{L}_1/2 + \epsilon_b. \tag{14}$$

The effort that goes in for deriving the analytical expressions for ϵ_a and ϵ_b outweighs the gain that might come with the theoretical exactness. The length corrections ϵ_a and ϵ_b are taken as 5 and 10% of L_1 intuitively for the above muffler configuration.

4. 3D numerical verification

The one-dimensional (1D) theory that makes use of matrizant analysis approach has been verified with its 3D numerical counterpart, the commercially available software SYSNOISE.[14] The acoustic domain has been discretized using eight-noded brick elements. A unit velocity excitation is imposed on the inlet face along with an anechoic termination at the outlet. An admittance relation is imposed at the perforate interface.[15] Assuming a continuity of the radial particle velocity at the boundary of the interacting ducts, acoustic variables on the faces of two sets of unconnected finite elements are related by

$$-u_{n1} = u_{n2} = u_n, \tag{15a}$$

$$\beta(p_2 - p_1) = u_n, \tag{15b}$$

where the suffix "n" represents the normal direction and β is the perforate admittance $[1/(\rho_0 c_0 \zeta)]$. The empirical expression for the perforate impedance depends on the flow configuration and is the same as that used in the 1D matrizant analysis. This model, with the appropriate boundary conditions is then solved for a given frequency range to yield the nodal acoustic properties, acoustic pressure "p" and particle velocity "u" as a function of frequency. The transmission loss at each frequency is then calculated using the formula

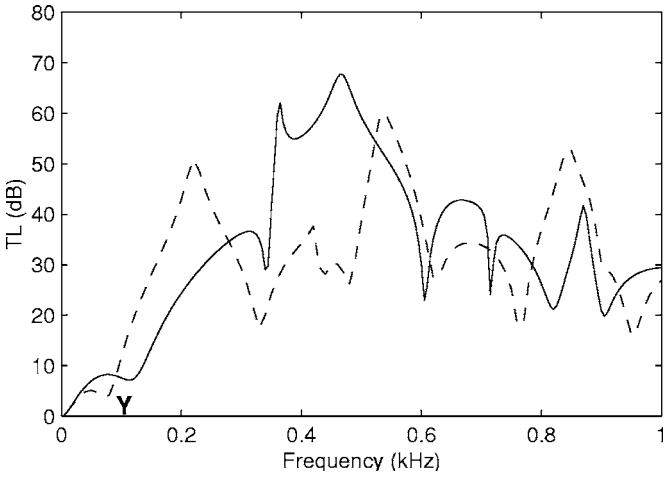


FIG. 5. TL spectrum for the muffler configuration of Fig. 4. – – –, $L_1 = 0.5$ m, $L_2 = 0.1$ m, $L_{1a} = 0.1$ m, $L_{1b} = 0.3$ m, $L_{2a} = L_{2b} = 0.0$ m, and ——, $L_1 = L_2 = 0.3$ m, $L_{1a} = 0.075$ m, $L_{2a} = 0.0$ m, $L_{1b} = L_{2b} = 0.15$ m.

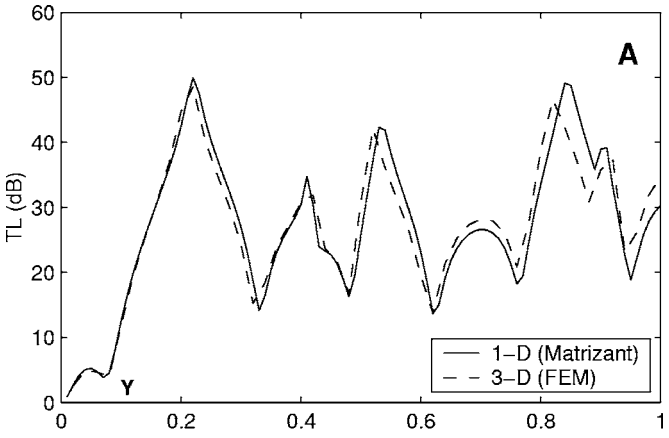
$$\mathrm{TL} = 20 \log_{10} \left| \frac{p_i + \rho_0 c_0 u_i}{\rho_0 c_0 u_o} \right|, \tag{16}$$

where subscript "i" refers to the inlet face and "o" refers to the outlet face.

5. Application

The use of absorptive materials in an air conditioning system is not advised owing to health hazards. This reactive muffler configuration may be applied for noise abatement of a propeller-type airhandling fan of constant rotational speed in terms of revolutions per minute (RPM) used in heating ventilation and air-conditioning (HVAC) systems. The parameters that dictate the TL spectrum should be optimized accordingly so as to achieve the desired attenuation at the blade passing frequency (BPF) and its next couple of harmonics. Here, two such cases with BPF at 150 and 100 Hz will be considered for which the desired TL/IL values are of the order of 20–30 dB. An optimized design must place the BPF and its first two harmonics under the spectral peaks.

The TL spectrum of this configuration for a BPF of 150 Hz is shown in Fig. 5, where R_1 and R_2 are taken as 25 and 75 mm, respectively. The dashed line represents the TL spectrum of a configuration with $L_1 = 0.5$ m and $L_2 = 0.1$ m. The spectrum indicates an attenuation of 28 dB at the BPF and 29 and 30 dB for the next two harmonics. Position of a peak on the frequency coordinate is a function of the effective lengths of the segments, whereas the magnitude of a peak depends upon the volume shared by each of the subelements, namely, the VCTR and the CTR. Effective low-frequency attenuation occurs for a shorter L_2. For a shorter L_2, noise propagating through the duct fails to interact with the cavity of the CTR at higher frequencies. Thus, the additional volume added to the cavity of the CTR in the form of an extended inlet (variable area) fails to produce any significant effect on the TL spectrum. The spectral peak moves toward the right on the abscissa for an increasing value of L_2. As shown by the solid line of Fig. 5, such a configuration



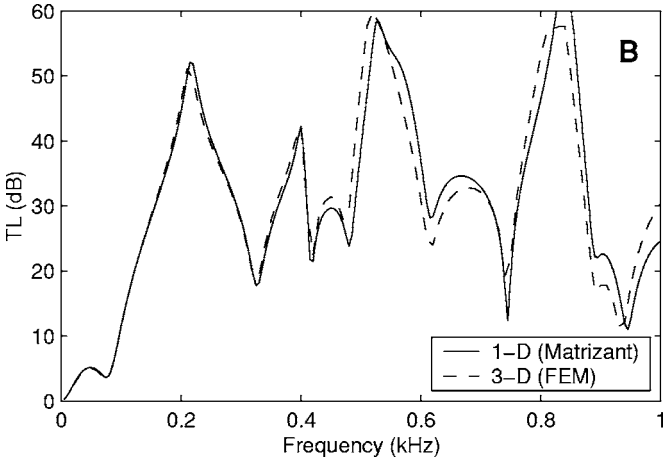


FIG. 6. TL spectra of C-VCTR of Fig. 4 for L_1=0.5 m, L_2=0.1 m, L_{1b} =0.3 m, L_{2a}=L_{2b}=0.0 m. (a) L_{1a}=0.0 m and (b) L_{1a}=0.1 m.

with L_1=L_2 provides a better attenuation in the medium frequency range. The low-frequency attenuation is not attainable by simply partitioning the cavity of a CTR. This is evident from the fact that each of the segments, when compared to the segments of the C-VCTR, will have smaller acoustic length.

6. Parametric study

In case the lengths of the extended inlet and the outlet are altered, the resonator performance at lower frequencies will be affected. This is due to the fact that the wave propagating through the pipe, in that case, will interact with a decreased cavity volume. If the first and second harmonics of the BPF lie well within the first trough, the performance may be enhanced by using only an extended outlet. Then, the length of the extended outlet may be taken in excess of 80% of L_1; i.e., L_{1a}=0.0 m and L_{1b}=0.4 m for the present configuration.

Any parametric variation is least effective at very low frequencies. The effectiveness builds up gradually as the frequency increases. Thus, the parametric study must target an improved performance for the first few harmonics rather than the BPF alone. Figure 6 outlines the effect of the extended inlet of the VCTR L_{1a}. The results are verified by their 3D field element method (FEM) counterparts. The accuracy of the present analysis may be noted from the comparisons.

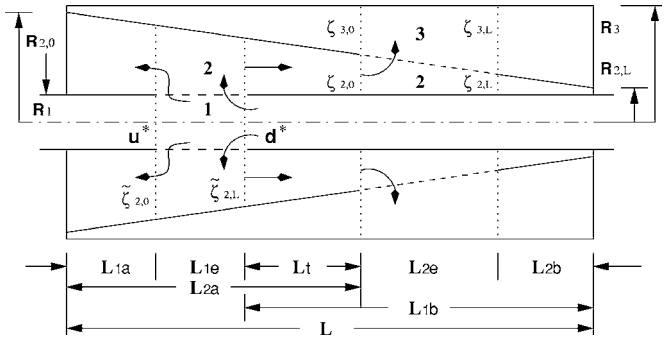


FIG. 7. A variable area folded resonator configuration for low frequency noise attenuation. The schematic diagram, basically a three duct perforated element, comprises of a couple of VCTRs within an overall length of length L with variable area extended inlets and outlets.

B. Variable area folded resonator (VAFR)

The trough, represented by the letter "Y" in Fig. 6, at a very low frequency, is a feature of the segmented resonators. This inevitable effect is independent of the length sharing among the segments. It is denoted by the arc (Y–Y′–Y″) in Fig. 2(**R**). This curtails the applicability of the composite resonators at very low frequencies. The folded resonator is used to overcome this shortcoming.[4] Here, a similar configuration but closely related to the VCTR, is the subject of discussion.

1. Analysis

The analysis proceeds on a similar track to that of the VCTR. With the partially perforated ducts, the muffler configuration of Fig. 7 is not exactly a folded resonator as it fails to serve the cardinal criterion of a unidirectional cavity volume. The perforation of the innermost duct is necessary in order to avoid the back pressure that would accompany sudden area changes, whereas it may be totally removed from the intermediate duct to provide a slitlike leakage into the outermost cavity. Here, however, the analysis presumes a partial perforation for both the ducts, as shown in Fig. 7.

The above configuration comprises a couple of two-duct VCTRs of perforated lengths L_{1e} and L_{2e}, respectively. Making use of the matrizant analysis, the transmision matrices $[\mathbf{\Omega}]$ of Eq. (11) for both the VCTRs may be determined. Let $[\mathbf{\Omega}^1]$ and $[\mathbf{\Omega}^2]$ be the 4×4 transmission matrices of the VCTRs, respectively. Applying the boundary conditions to the matrix $[\mathbf{\Omega}^2]$ in terms of the impedance expressions $\zeta_{3,0}$, $\zeta_{3,L}$, and $\zeta_{2,L}$, the impedance $\zeta_{2,0}$ may be obtained.

$$\zeta_{2,0} = \frac{\mathbf{A}_1\mathbf{C}_2 - \mathbf{B}_1\mathbf{C}_1}{\mathbf{A}_2\mathbf{C}_2 - \mathbf{B}_2\mathbf{C}_1}, \tag{17}$$

where

$$\mathbf{A}_1 = [\mathbf{\Omega}^2]_{1,1}\zeta_{2,L} + [\mathbf{\Omega}^2]_{1,3} \quad \text{and} \quad \mathbf{A}_2 = [\mathbf{\Omega}^2]_{3,1}\zeta_{2,L} + [\mathbf{\Omega}^2]_{3,3}, \tag{18a}$$

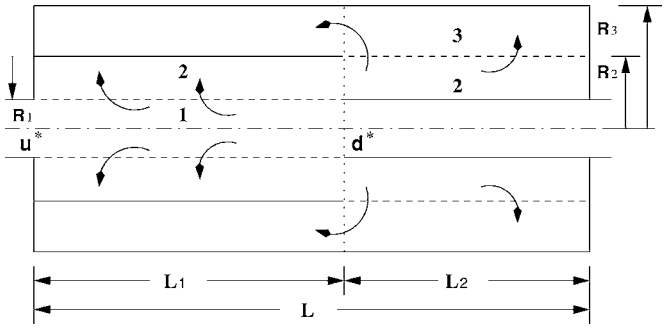


FIG. 8. Schematic diagram of a uniform area folded resonator. L=0.5 m, L_{1a}=L_{2b}=0.0, L_{1b}=L_{2a}=0.25 m, $R_{2,0}$=$R_{2,L}$=50 mm.

$$\mathbf{B}_1 = [\mathbf{\Omega}^2]_{1,2}\zeta_{3,L} + [\mathbf{\Omega}^2]_{1,4} \quad \text{and} \quad \mathbf{B}_2 = [\mathbf{\Omega}^2]_{3,2}\zeta_{3,L} + [\mathbf{\Omega}^2]_{3,4}, \tag{18b}$$

$$\mathbf{C}_1 = [([\mathbf{\Omega}^2]_{2,1}\zeta_{2,L} + [\mathbf{\Omega}^2]_{2,3}) - \zeta_{3,0}([\mathbf{\Omega}^2]_{4,1}\zeta_{2,L} + [\mathbf{\Omega}^2]_{4,3})], \tag{18c}$$

$$\mathbf{C}_2 = [([\mathbf{\Omega}^2]_{2,2}\zeta_{3,L} + [\mathbf{\Omega}^2]_{2,4}) - \zeta_{3,0}([\mathbf{\Omega}^2]_{4,2}\zeta_{3,L} + [\mathbf{\Omega}^2]_{4,4})]. \tag{18d}$$

Applying the expression for the impedance $\zeta_{2,0}$ to the transfer matrix $[\boldsymbol{\mathcal{T}}]$, of the variable area duct of Length L_t, the corresponding impedance expression $\tilde{\zeta}_{2,L}$ may be determined

$$\tilde{\zeta}_{2,L} = \frac{\boldsymbol{\mathcal{T}}_{1,1}\zeta_{2,L} + \boldsymbol{\mathcal{T}}_{1,2}}{\boldsymbol{\mathcal{T}}_{2,1}\zeta_{2,L} + \boldsymbol{\mathcal{T}}_{2,2}}. \tag{19}$$

The impedance expression $\tilde{\zeta}_{2,L}$ of Eq. (19) along with the other impedance $\tilde{\zeta}_{2,0}$ of the extended inlet of length L_{1a} forms the set of two boundary conditions, later applied to the transmission matrix $[\mathbf{\Omega}^1]$ to generate the desired four-pole transmission matrix connecting the state variables at the upstream point u^* to those at the downstream point d^* in Fig. 7.

2. Results and discussion

The variable area folded resonator is designed to perform at a very low frequency. Thus, similar to the C-VCTR of Sec. III A. (Fig. 4), the performance analysis of this muffler will be limited to low-frequency range. To bring forth a comparison between both the configurations, the overall dimensions, unless mentioned explicitly for individual cases, will be the same as those of the C-VCTR. The dimensions R_1 and R_3 are fixed at 25 and 75 mm, respectively. The corresponding first higher-order mode cut-on frequency is 1328 Hz, whereas the first radial mode gets cut on at 2764 Hz. As done before to all other configurations, the present one-dimensional theoretical approach is verified against its three-dimensional FEM counterpart. The comparison for a uniform area folded resonator (UAFR) of Fig. 8 ($R_{2,0}$=$R_{2,L}$=R_2) is shown in Fig. 9. Figure 10 depicts the corresponding TL spectra for a variable area folded resonator. The difference between the results of the present approach and the FEM analysis is negligible and hard to make out.

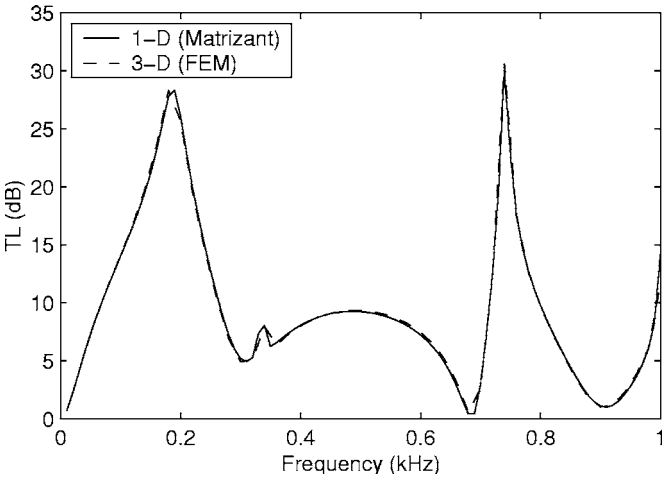


FIG. 9. Comparison of the adopted matrizant analysis with its 3D FEM counterpart for the uniform area folded resonator of Fig. 8.

As predicted before, the absence of the trough (denoted by "Y" in Fig. 6) characterized by the segmentation of a muffler enhances the very low-frequency performance. The advantages as well as the limitations of the variable area folded resonator will be discussed in detail in the next section. With L_{1a}=L_{2b}=0, the configuration may be considered as double cavity, single-tuned resonators. The incorporation of the double-tuned cavities will be its generic extension. The performance curve of the VAFR is shown in Fig. 11. The spectrum shows an improvement with a broadband peak between 300 and 500 Hz, which may cater a good attenuation for the third to fifth harmonics of a BPF of 75–85 Hz, but collectively, on a whole it has very little improvement at the desired frequency range (0–150 Hz). This however does not mean an instant negation of the concept and incidentally will be justified later.

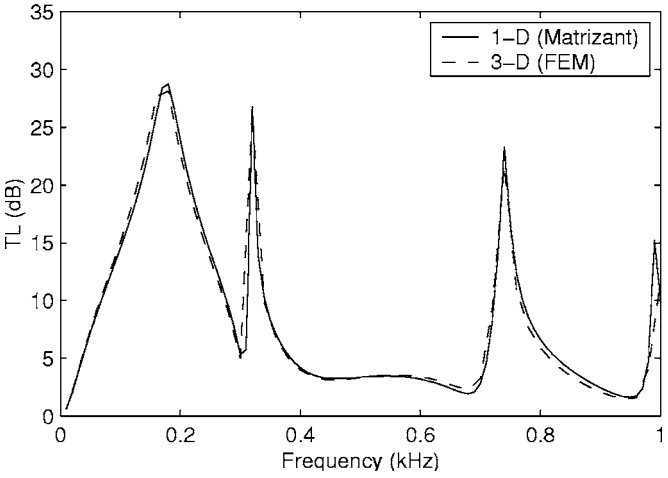


FIG. 10. Comparison of the adopted matrizant analysis with its 3D FEM counterpart for a VAFR of Fig. 7. L=0.5 m, L_{1a}=L_{2b}=0.0, L_{1b}=L_{2a}=0.25 m, $R_{2,0}$=73 mm and $R_{2,L}$=27 mm.

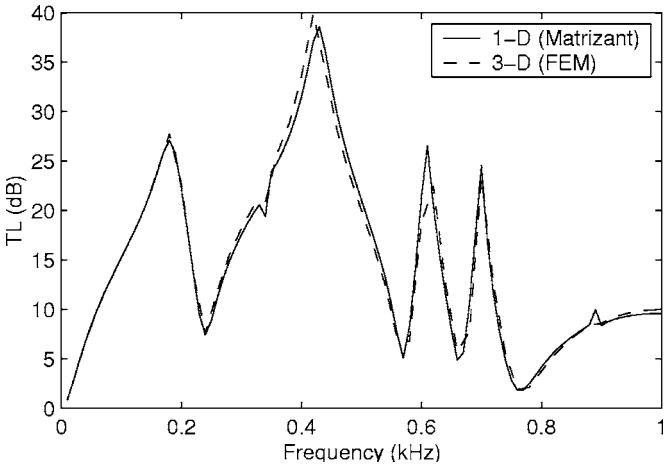


FIG. 11. Comparison of the present 1D theory with its 3D FEM counterpart for a VAFR of Fig. 7 with double-tuned cavities. L=0.6 m, L_{1a}=L_{2b}=0.15, L_{1b}=L_{2a}=0.3 m, $R_{2,0}$=73 mm, $R_{2,L}$=27 mm.

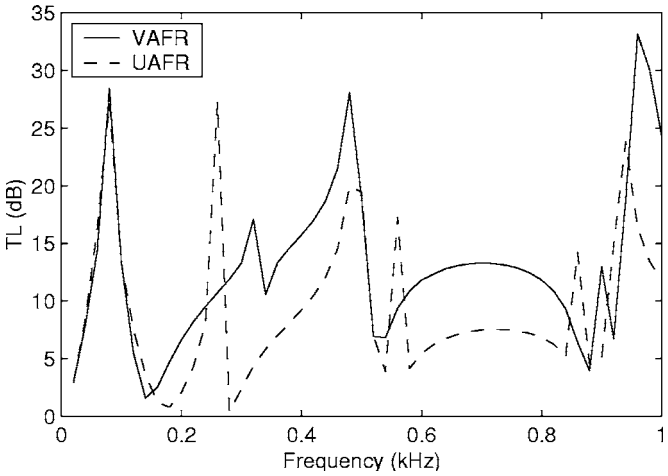


FIG. 13. TL spectra of folded resonators with nonzero L_t(L_t=0.3 m). L =0.6 m, L_{1a}=L_{2b}=0 and L_{1b}=L_{2a}=0.45 m. ——, $R_{2,0}$=73 mm, $R_{2,L}$ =27 mm and - - - -, $R_{2,0}$=$R_{2,L}$=50 mm.

3. Parametric study

Although an extensive analysis can cover a wide range of parametric study, here the analysis is restricted to a limited version of the configuration where the intermediate duct L_t is assumed to be absent. This assumption will be reasoned out later in this section. With an objective of low-frequency attenuation, the sound energy associated with the firing frequency and its first couple of harmonics is of primary concern. Thus, the performance curves of the C-VCTR of Fig. 6(b) and a couple of VAFR configurations are compared for a frequency less than 150 Hz in Fig. 12. Here, the different VAFR configurations have different values of extended inlet and outlet. Perforated length L_{1e} as well as the overall dimensions are kept the same. VAFR-1 is the same as in Fig. 11, whereas VAFR-2 has only the extended outlet. In the latter configuration, the sound wave interacts with a single large cavity and thus shifts the fundamental peak towards the lower frequency further.

Although it deteriorates at the higher frequency, the performance of the variable area folded resonator, when compared to the corresponding C-VCTR, is a much better option for very low-frequency noise attenuation. It may be noted that the VAFR-1 offers a 5–10 dB more attenuation than the C-VCTR for a BPF of 60–100 Hz, and more particularly around 75–85 Hz, where it is in excess of 10 dB. The low-frequency performance for VAFR-2, with the dimensions L_{1a}=0.45, L_{2b}=0.15, and L_{1b}=L_{2a}=0.0 m, is even better when compared to the other configurations. Compared to the C-VCTR, it gives a 7 dB additional attenuation at 60 Hz, and an absolute attenuation in excess of 22.5 dB at 100 Hz. If the noise source is of constant speed (ventilation fan), then the latter configuration may be useful to attenuate the next two harmonics, which stands at 11 and 13 dB for the first and second harmonics, respectively, for the present set of parameters. The enhanced performance of VAFR-2 at the lower-frequency range is obtained at the cost of its performance at higher frequencies.

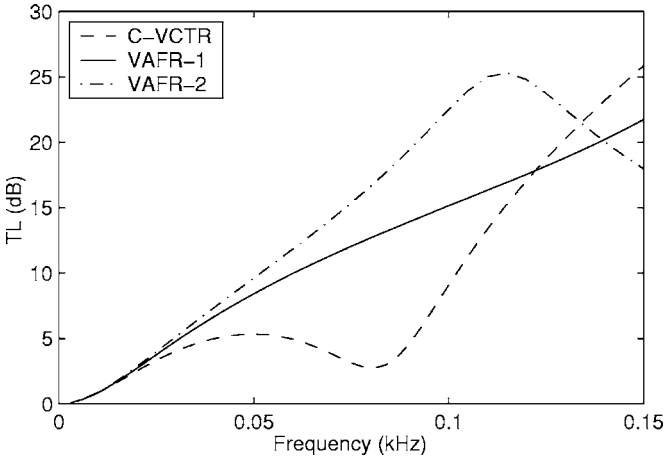


FIG. 12. Comparison of the performance curves of a C-VCTR with its VAFR counterparts at very low frequency (f<150 Hz). - - - -, (C-VCTR) of Fig. 6(b); ——(VAFR-1) of Fig. 11; and –·–·–(VAFR-2), L=0.6, L_{1a}=0.45, L_{2b}=0.15, and L_{1b}=L_{2a}=0.0 m, $R_{2,0}$=73 mm, and $R_{2,L}$=27 mm.

The effect of a nonzero L_t is shown in Fig. 13. The configuration has a partial perforation for the ducts 1 and 2, z=0 to z=15 mm for the inner tube, and z=45 to z =60 mm for the intermediate duct. The fundamental peak appears at very low frequency due to a longer effective length. Here, the performance of a VAFR is compared with the corresponding UAFR.[4] Although they coincide at very low frequency, the VAFR, when compared to UAFR, has a better broadband attenuaion spectrum at higher frequencies.

IV. DESIGN ILLUSTRATION

The design of mufflers is a combination of art (experience and intuition) and science. As has been shown above, two variable area resonator configurations (C-VCTR and VAFR) are particularly efficient for low-frequency attenuation. These are used hereunder to illustrate their performance against different sources of noise.

A. Constant speed source

HVAC systems with the air handling fan as the noise source often make use of absorptive materials. However, in some thermal power station applications, the use of absorp-

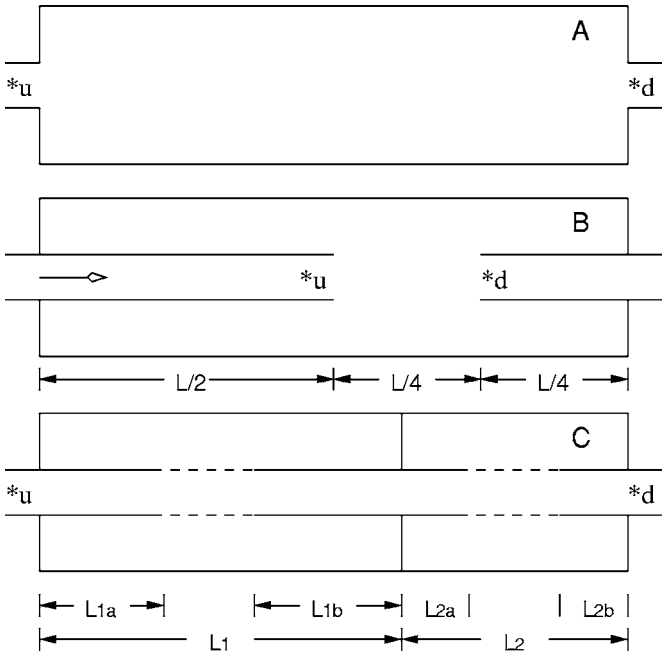


FIG. 14. Schematic diagrams of (a) simple expansion chamber and (b) double-tuned expansion chamber.

tive materials may be inadvisable because of high temperatures and/or high amount of suspended particulates. In such-like situations, one may contemplate use of reactive mufflers. Here, the fan specifications of 1500 RPM and seven blades correspond to the blade passing frequency (BPF) of 175 Hz. Noise for such applications remains within the BPF and first two harmonics. Thus, the analysis should cover the attenuation band of 175–525 Hz. This may either be achieved by a broadband transmission loss spectrum or by the tonal attenuation at certain frequencies. The length of the muffler is 0.6 m. TL spectrum of a C-VCTR is compared with those of simple and double-tuned expansion chambers of Fig. 14 and is shown in Fig. 15.

The maximum TL amplitude of the simple expansion chamber, a function of the ratio of the radial dimensions R_1 (inlet/outlet) and R_2 (cavity), for the present set of dimensions is 13.2 dB and the troughs touch zero decibels. Thus, neglecting the performance of the simple expansion chamber, a quantitative comparison over the predicted performances (b) and (c) of Fig. 15 is made in Table I. It is of utmost importance to obtain better attenuation at the BPF. The performance of the C-VCTR is far more superior than that of the double-tuned expansion chamber. In order to avoid the back pressure, a highly perforated bridge is usually provided to connect the inlet and the outlet tube of the tuned expansion chambers, which accounts for a relatively lower trough. Also the C-VCTR provides a better attenuation for the first and second harmonics.

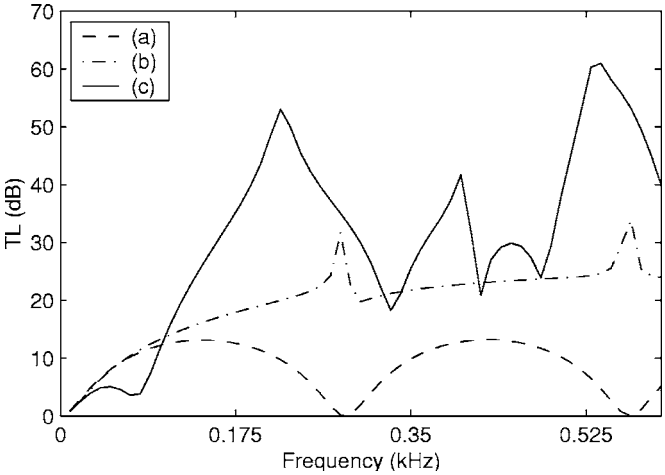


FIG. 15. TL spectra of different configurations for an HVAC system. (a) Simple expansion chamber [Fig. 14(a)], (b) double-tuned expansion chamber [Fig. 14(b)] and (c) C-VCTR with L_1=0.5, L_2=0.1, L_{1a}=0.1, L_{1b}=0.3, and L_{2a}=L_{2b}=0.0 m.

TABLE I. TL values (dB) of the spectra (b) and (c) of Fig. 15 for a frequency range of 175–525 Hz.

Frequency (Hz) → Spectra ↓	175	350	525
(b)	18.0	21.8	24.2
(c)	35.2	25.5	56.5

Figure 16 shows the sound attenuation map of the spectrum (c) of Fig. 15 with L_{1b} as parameter for L_{1a}+L_{1b} =0.4 m. TL peaks and troughs are characterized by the bright and dark patches, respectively. It may be noted that the variation of L_{1b} (and hence L_{1a}) has little effect on the attenuation at the BPF. But the performance then onward, particularly for the first harmonic, depends upon these parameters and they may be varied accordingly. Although the noise reduction at the BPF is quite insensitive to the minor parametric changes, an attenuation in excess of 40 dB is quite achievable with the present overall dimensions.

B. Variable speed source

Automotive engines are variable speed engines. A good tonal attenuation is no solution for such applications and a broadband noise reduction is a necessary criterion. Here, a four cylinder, four stroke cycle diesel engine with a rated speed of 1000 RPM is considered. Figure 17 shows a typical sound pressure level (SPL) spectrum of the diesel engine at no-load condition.[16] In this case, the firing frequency is

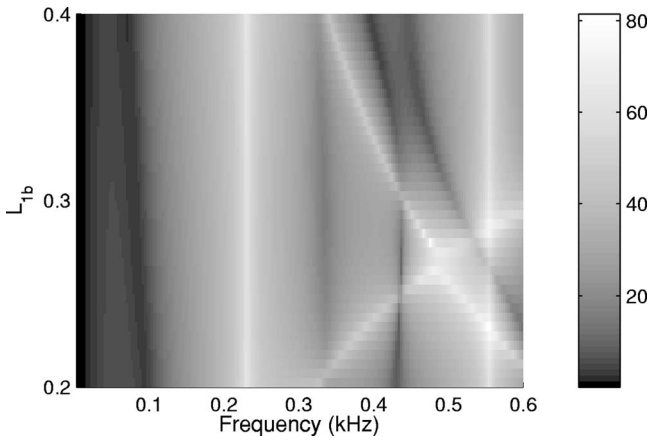


FIG. 16. Sound attenuation map of the spectrum (c) of Fig. 15 with L_{1b} as parameter for L_{1a}+L_{1b}=0.4 m.

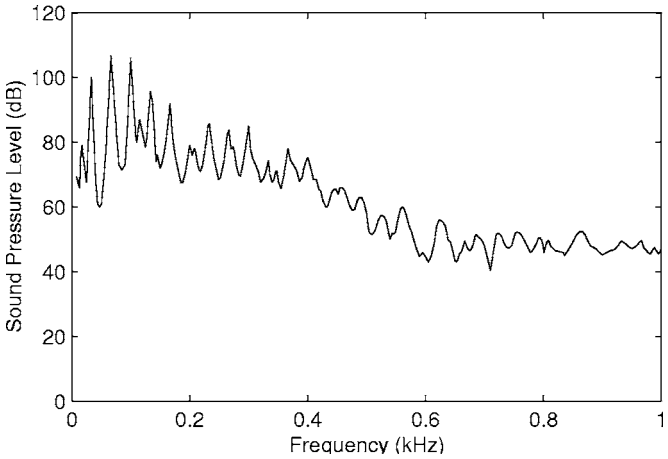


FIG. 17. Sound Pressure Level of a four-cylinder, four stroke cycle diesel engine with a rated speed of 1000 RPM with no-load condition.[16]

33.3 Hz. The sound pressure level for first six harmonics is around 95–110 dB. The first few harmonics have distinct peaks, whereas they become less prominent as the frequency increases. The absolute value of the SPL decreases with the increase of frequency, making the low-frequency attenuation more vital. Due to the implication of the antilogarithmic addition of the SPL for the octave bands, an attenuation spectrum has to be designed to cover the harmonics with higher pressure level, i.e., first few harmonics. At very low frequency (firing frequency and its first harmonic), the A-weighted sound level becomes insignificant. Considering the sound power distribution of the noise source over a frequency range, the challenge to curb the noise for the next few harmonics ($f_h \in [100-250]$ Hz) becomes more important.

Here, the analysis will be performed for a muffler length of 0.6 m over a frequency range of 0–1 kHz, that comprises several harmonics of the firing frequency. The performance curves for two different configurations, C-VCTR and segmented double-tuned CTR, are shown in Fig. 18. TL spectra for both the muffler configurations are shown in Fig. 18(a), whereas Fig. 18(b) shows the difference between their performance in terms of TL. The quantitative difference may be noticed from the zone "**Z**" of Fig. 18(b), where the C-VCTR offers a better TL than the segmented double tuned CTR in excess of 5–14 dB.This is the most significant advantage of the variable area resonators over its uniform area resonator counterparts.

V. CONCLUSIONS

In this paper, two rather novel configurations have been introduced and investigated for enhancing the low-frequency performance of mufflers for the constant-speed machines like DG sets and HVAC systems. Some of the salient observations or conclusions are as follows:

- The most significant advantage of the composite variable-area concentric tube resonator (C-VCTR) is the absence of any trough up to about 500 Hz, a feature that makes it ideal for HVAC systems.

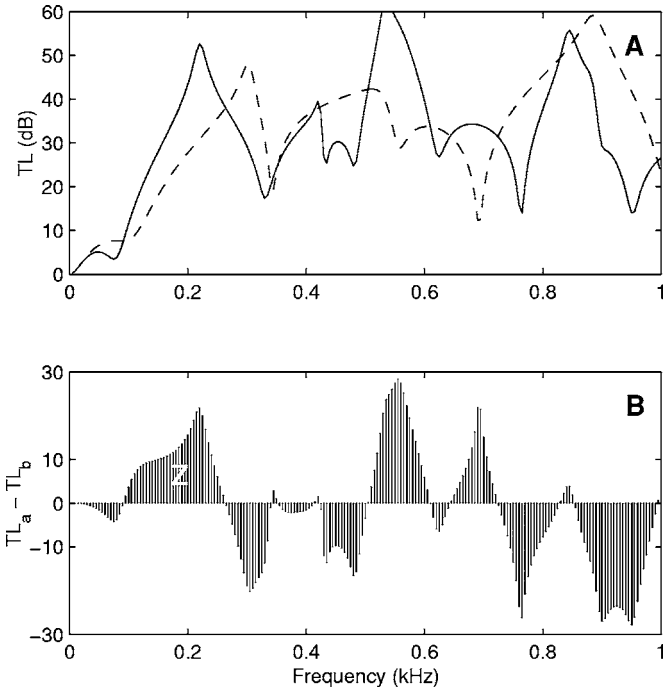


FIG. 18. TL spectra for two different configurations. (a) (a)——, muffler configuration of Fig. 4. L_1=0.5 m, L_2=0.1 m, L_{1a}=0.1 m, L_{1b}=0.3 m, L_{2a} =L_{2b}=0.0 m and (b) – – –, configuration of Fig. 14(c): L_1=0.5 m, L_2 =0.1 m, L_{1a}=0.125 m, L_{1b}=0.25 m, L_{2a}=L_{2a}=0.0 m. (b) $TL_a - TL_b$.

- The TL spectra beyond the primary resonance frequency remain the same and are quite insensitive to the length of the intermediate tube.
- Application of VCTR inside an overall uniform cross-sectional area configuration enhances the performance of a muffler at lower frequencies.
- The above configurations can be fine tuned by incorporating an extended outlet into the CTR. But the performance at lower frequencies is more or less insensitive to these variations, as the volume of the extended inlet to the CTR dominates the effectiveness of this addition.
- Although the configuration for a HVAC system with blade passing frequency (BPF) of 175 Hz has been shown, the corresponding ones for other frequencies may be designed on a similar line.
- For a system that carries acoustic energy (noise) over a wider frequency range, the ratio of the segmental lengths may be altered to attain a broad band attenuation.
- The concept of the folded resonators (VAFR/UAFR) may find application for very low frequency attenuation.

ACKNOWLEDGMENT

The authors would like to place on record their appreciation of the financial sponsorship of FRITA by the Department of Science and Technology of the Government of India.

[1]R. B. Bourne, "Sound attenuating devices," U.S. Patent No. 2,017,744(1934).
[2]T. Kar and M. L. Munjal, "Analysis and design of conical concentric tube resonators," J. Acoust. Soc. Am. **116**(1), 152, 74–83 (2004).
[3]C. Q. Howard, B. S. Cazzolato, and C. H. Hansen, "Exhaust stack silencer

design using finite element analysis," Noise Control Eng. J. **48**(4), 113–120 (2000).
[4]R. Glav, P. L. Regaud, and M. Abom, "Study of a folded resonator including the effects of the higher order modes," J. Sound Vib. **273**(4–5), 777–792 (2004).
[5]M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, l987).
[6]M. L. Munjal, in "Muffler acoustics," *Formulas of Acoustics*, edited by F. P. Mechel (Springer-Verlag, Berlin, 2002), Chap. K.
[7]R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices and Some Applications to Dynamics and Differential Equations* (Cambridge University Press, Cambridge, England, 1952).
[8]T. Kar, "Plane wave analysis of variable area perforated-tube resonators and acoustic wedges," Ph.D. dissertation, IISc, Bangalore, 2005.
[9]T. Kar and M. L. Munjal, "An inherently stable boundary-condition-transfer algorithm for muffler analysis," J. Acoust. Soc. Am. **116**(1), 74–83 (2004).
[10]T. Kar, P. P. R. Sharma, and M. L. Munjal, "Analysis of multiple-duct variable area perforated tube resonators," Int. J. Acoust. Vib. **11**(1), 19–26 (2006).
[11]A. G. Webster, "Acoustical impedance, and the theory of horns and of the phonograph," Proc. Natl. Acad. Sci. U.S.A. **5**(7), 275–282 (1919).
[12]M. F. Hamilton, Y. A. Ilinskii, and E. A. Zabolotskaya, "Linear and nonlinear frequency shifts in acoustical resonators with varying cross sections," J. Acoust. Soc. Am. **110**(1), 109–118 (2001).
[13]M. L. Munjal, "Design of passive silencers ," *International Conference on Computational and Experimental Engineering and Sciences*, Corfu, Greece (July 2003).
[14]*LMS SYSNOISE Rev. 5.6 veriosn 1.0*, LMS International , Leuven, Belgium (March 2003).
[15]M. L. Munjal, S. Krishnan, and N. M. Reddy, "Performance of perforated elements with application to design," Noise Control Eng. J. **40**(1), 159–167 (1993).
[16]Timothy D Whitehead, M.E. thesis, Dept. of Mech. Eng., Chirstchurch, University of Canterbury, NZ (2005).

Investigation of a three-phase medium with a negative parameter of nonlinearity

D. H. Trivett,[a] Hervé Pincon, and Peter H. Rogers
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 30 September 2005; revised 21 March 2006; accepted 27 March 2006)

Nonlinear acoustic properties of a composite medium, consisting of hollow microspheres suspended in Castor Oil, at elevated hydrostatic pressure are experimentally investigated. The acoustic nonlinear parameter B/A of the medium is found to be highly dependent upon hydrostatic pressure. B/A varies from a small negative value near ambient pressure to a very large negative value (around -6000) in the vicinity of 7×10^4 Pascals, above ambient pressure. With a further increase in hydrostatic pressure the magnitude of B/A decreases passing through zero and finally becomes positive. Finite amplitude wave propagation in this medium at low hydrostatic pressures is characterized by waveform steepening in the backward direction leading to rarefactive shockwaves and, at high hydrostatic pressures, by steepening in the forward direction leading to compressive shockwaves. Estimates of B/A are obtained from both the measurement of thermodynamic properties and from waveform distortion during propagation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197804]

PACS number(s): 43.25.Ba, 43.25.Cb, 43.25.Zx [MFH] Pages: 3610–3617

I. INTRODUCTION

Nonlinear propagation in typical fluids, from the point of view of the nonlinear acoustician, is an agonizingly slow process of the accumulation of small nonlinear perturbations. These perturbations distort the waveform and the rate is characterized by the acoustic nonlinear parameter B/A.[2] For liquids, B/A varies between 3 and 13 with the largest reported value of 13.68 (Ref. 3) having been measured in Fluorocarbon 43 at a temperature of 0 °C. The values of the nonlinear parameter in liquids are higher than those typical for gases, which are in the range of 0.2 to 0.7. However, in both cases one must either operate at high frequency, generally in the MHz region, or over long ranges to realize significant distortion. While this has not prevented the application of nonlinear acoustics in many technical areas, such as nonlinear ultrasonic imaging, where the use of high frequency is advantageous, the low levels of nonlinearity in liquids has restrained practical applications at more moderate frequencies. This is especially true in underwater acoustics where the only significant application of nonlinear acoustics is Westervelt's[4] parametric array and even this application suffers from the low efficiency of difference frequency generation due to the small value of B/A in water.

There have been attempts to increase the magnitude of B/A in liquids to allow for more compact applications of nonlinear acoustics. The introduction of ultrasonic contrast agents in conjunction with the use of high frequency, near the resonant frequency of the microspheres, has been found to significantly increase the nonlinearity of the resulting composite medium. A value of 2678 was reported for B/A in a liquid solution containing an undisclosed void fraction of the contrast agent Echovist® 300.[5] The authors also reported that higher values were obtained with increased void fractions; however, the enhancement was relatively short lived with the value of B/A decreasing to 16 in approximately 45 minutes as the contrast agent dissolved. Wu and Zhu[6] reported a value for B/A between 10^4 and 10^5 in tap water containing cylindrical air bubbles trapped in membranes. The range was due to uncertainty in the value of the sound speed in the composite medium. These measurements were also performed at high frequency, 950 kHz, near the resonant frequency of the trapped bubbles. While Wu and Zhu utilized high frequency in their measurements they noted that there were two distinct physical mechanisms involved in the enhancement of the B/A due to the presence of bubbles. One is the nonlinear oscillation of the bubbles when driven near their resonant frequency and the second is the hydrostatic dependence of the static properties of the composite medium, which can enhance the nonlinearity at a much lower frequency.

The low frequency enhancement of nonlinearity, due to the presence of bubbles, is well known. Apfel[7] developed a method for calculating the effective B/A of immiscible fluids, later corrected by Everbach *et al.*,[8] which when applied to free air bubbles well below resonance size in water yields enhancement at all void fractions below 0.3. Figure 1 contains a plot of B/A for air bubbles in water, and a peak value of 7.7×10^3, depending upon the value of the sound speeds, is obtained when the void fraction is 6.5×10^{-5}. Kobelev and Sutin[9] and Kozyaev and Naugol'nykh[10] theoretically investigated bubbly liquids as a means of enhancing the efficiency of parametric arrays. Although the theoretical investigations were promising, practical realization has been held back by the significant attenuation which occurs in bubbly media near the resonant frequency of the bubbles. However, a preliminary experiment with a bubble layer[11] demonstrated that enhancement was possible by observing difference frequency several times larger than would be obtained with pure water.

[a]Electronic mail: david.trivett@me.gatech.edu

 0001-4966/2006/119(6)/3610/8/$22.50

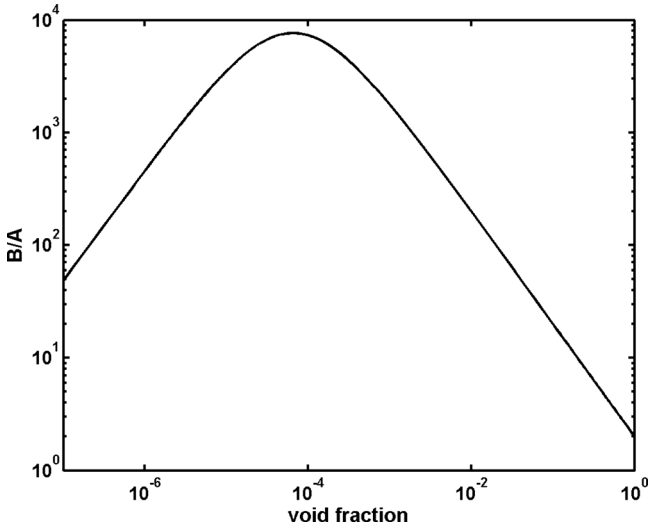


FIG. 1. The effective B/A of a gas-liquid mixture versus gas void (volume) fraction.

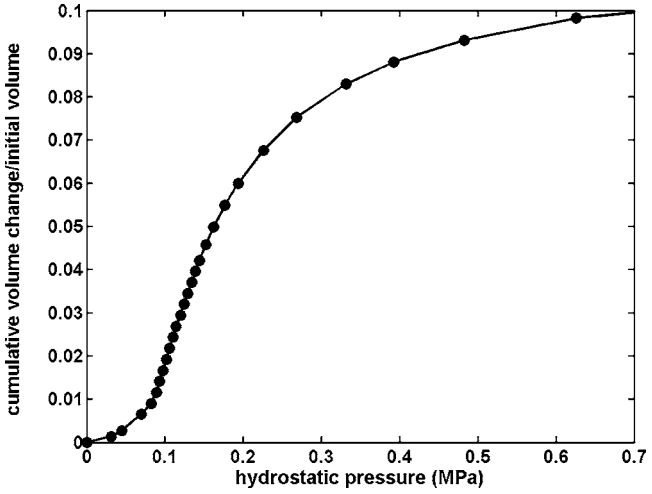


FIG. 2. Normalized volume change of composite medium as a function of hydrostatic pressure. The open circles are measurements and the solid line simply connects these values.

More recently Druzhinin, Ostrovsky, and Prosperetti[12] proposed that parametric array enhancement could be obtained by exploiting the resonances of a bubbly liquid layer where the difference frequency is chosen to be at a thickness resonance of the layer. Karpov, Prosperetti, and Ostrovsky[13] corrected and expanded the investigation and suggested that, in order to avoid the practical difficulties associated with free air bubbles, one might look to other systems. These would need to be media in which the static properties are highly dependent upon the hydrostatic pressure. One such medium which we have been investigating consists of a mixture of microspheres and Castor oil. This medium has unusual static properties that are highly dependent upon hydrostatic pressure leading to some interesting nonlinear properties and regions of high nonlinearity. The following is a description of the properties of this medium and of measurements of the parameter of nonlinearity at certain hydrostatic pressures.

II. STATIC PROPERTIES OF THE MEDIUM

The medium consists of an arbitrarily chosen volume ratio of 90% Castor oil[14] and 10% Expancel® 091 DE 80 d30 (Ref. 1) microspheres. Castor oil is commonly used in underwater acoustics due to its good impedance match to water, high electrical resistivity, and low cost. It has a remarkably constant composition for a naturally occurring compound and is an ideal fill fluid for underwater transducers. As such, it has been extremely well characterized by the sonar community.[15,16] Castor oil also has a significantly higher viscosity than water and was chosen to reduce migration of the microspheres, which are positively buoyant. In any practical application the Castor oil would be replaced with a gelatin or soft polyurethane, however, the necessity of making propagation measurements at various locations was simplified by using a fluid.

The Expancel microspheres used in this application are spherical, hollow shells with diameters that vary between 35 and 55 μm and a shell thickness of approximately 0.1 μm, encapsulating a gas that is either isobutene or isopentane. The microspheres are commonly used as light-weight filler for automotive underbody coatings, paints, ceramics, and cultured marble. As such, there is not a lot of quality control applied to their production.

The isothermal, static bulk modulus of the medium was measured in a dilatometer, which consists of a pressure vessel that is housed in an environmental chamber. The pressure vessel is a massive, steel cylinder with a lid bolted onto the top that is sealed with a single-use, hollow steel o-ring to minimize the compliance. The dilatometer is initially filled with Castor oil and its compliance is obtained by measuring the volume change (volume of oil introduced by a precision screw pump) as a function of hydrostatic pressure and comparing it to a value calculated from the known properties of Castor oil. This allows one to obtain the volume expansion of the dilatometer as a function of hydrostatic pressure and then to subtract this quantity from measurements of unknown materials. Into this measurement system the microvoided-Castor oil mixture was introduced and the measured volume change as a function of hydrostatic pressure and corrected for the small compliance of the dilatometer, is shown in Fig. 2. Initially the compliance of the medium is relatively low resulting in a small volume change with increasing hydrostatic pressure. In the hydrostatic pressure range of 70 to 90 kPa, the compliance dramatically increases and remains nearly constant until the pressure reaches 150 kPa. Above this pressure the compliance steadily decreases and asymptotically approaches that of the Castor oil. This behavior is entirely due to the geometry and size distribution of the microspheres. The microspheres are perfectly spherical, hollow shells. They have a high bulk modulus or low compliance at low hydrostatic pressures due to the hoop stiffness of the shell. As the hydrostatic pressure is increased a critical pressure, P_{cr}, given by[17]

$$P_{cr} = \frac{2Et^2}{r^2\sqrt{3(1-\sigma^2)}}, \tag{1}$$

where E is the Young's modulus of the shell material, t is the shell thickness, r is the radius and σ is Poisson's ratio, is

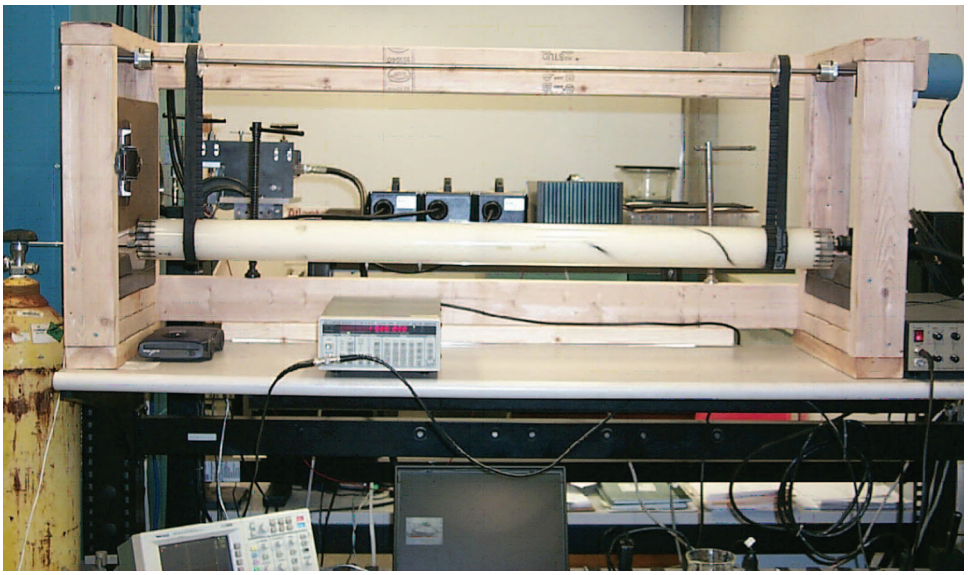


FIG. 3. Photograph of the sound speed measurement system, the plexiglass tube, suspended from the automatic rotation apparatus. The tube is filled with the composite medium and the projector cable is visible on the right-hand side along with the connection to the hand pressure pump. The rod extending from the left end cap is attached to the hydrophone and allows for positional adjustments.

reached where the microsphere buckles. Once buckled, the hoop stiffness vanishes and the compliance of the microsphere is approximately that of a gas bubble. After all of the microspheres have buckled, as determined by the geometric size distribution, the medium behaves in a manner similar to a bubbly fluid. The compliance decreases with further increase of hydrostatic pressure as the voids collapse and the void fraction is reduced ultimately reaching a value equal to that of pure Castor oil. This process is reversible and repeatable as long as the microspheres are not held at very high hydrostatic pressures for extended periods of time.

Clearly, the static compliance of this composite medium is highly dependent upon hydrostatic pressure and in the vicinity of the buckling of the microspheres it has the unusual property of increasing compliance with increasing hydrostatic pressure. In this region the sound speed decreases with increasing hydrostatic pressure and should result in a negative coefficient of nonlinearity. In addition, the observed change in compliance is very large leading one to suspect that the coefficient of nonlinearity is also very large. For these reasons, we undertook the experimental determination of coefficient of nonlinearity of this medium as a function of hydrostatic pressure.

III. THERMODYNAMIC MEASUREMENT OF *B/A*

Following the derivation of Beyer,[2] the difference between the instantaneous pressure p and the hydrostatic pressure p_0 is expanded in a Taylor series, along an isentrope, in terms of $(\rho-\rho_0)$:

$$p = p_0 + \left(\frac{\partial p}{\partial \rho}\right)_{\rho_0,s}(\rho-\rho_0) + \frac{1}{2}\left(\frac{\partial^2 p}{\partial \rho^2}\right)_{\rho_0,s}(\rho-\rho_0)^2 + \cdots . \tag{2}$$

Ignoring terms higher than quadratic, the equation can be rewritten as

$$p = p_0 + A\left(\frac{\rho-\rho_0}{\rho_0}\right) + \frac{B}{2}\left(\frac{\rho-\rho_0}{\rho_0}\right)^2, \tag{3}$$

where

$$\frac{B}{A} = \frac{\rho_0}{c_0^2}\left(\frac{\partial^2 p}{\partial \rho^2}\right)_{\rho_0,s} = 2\rho_0 c_0\left(\frac{\partial c}{\partial p}\right)_{\rho_0,s}. \tag{4}$$

While Eq. (4) is a deceptively simple expression for the acoustical nonlinear parameter, it is very difficult to measure adiabatic processes. Thus, using thermodynamic identities, we rewrite the expression in terms of variables that are easier to control, obtaining

$$\frac{B}{A} = 2\rho_0 c_0\left(\frac{\partial c}{\partial p}\right)_{\rho_0,T} + \frac{2\eta T c_0}{c_p}\left(\frac{\partial c}{\partial T}\right)_{\rho_0,p}, \tag{5}$$

where T is the temperature, c_p is the specific heat at constant pressure, and η is the coefficient of thermal expansion,

$$\eta = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p. \tag{6}$$

Three fundamental measurements are required to thermodynamically determine the value of B/A in Eq. (5). These are the specific volume, the isothermal hydrostatic pressure dependence of the sound speed, $(\partial c/\partial p)_T$, and the isobaric temperature dependence of the sound speed, $(\partial c/\partial T)p$.

A. Measurement of $(\partial c/\partial p)_T$

Measuring the isothermal, hydrostatic pressure dependence of the sound speed of the composite medium in free field is impractical. The requirement of varying the hydrostatic pressure necessitates the use of a closed container with the knowledge that the measurements will require corrections to obtain free-field values. We chose to perform the measurements in a thick-walled plexiglass tube, see Fig. 3, having an inner diameter of 5.07 cm, a wall thickness of 1.27 cm, and a length of 122 cm. Plexiglass, as opposed to steel, was chosen to allow one to visually determine that the medium was well mixed prior to making measurements. The

microspheres are buoyant and the medium requires regular attention to maintain a homogeneous mixture. This was accomplished by automatically rotating the tube 180 degrees, to change the direction of migration, every 2 minutes. Testing had shown that this time period was sufficient to keep the medium well mixed. A fixed acoustic source, a 1.9 cm diameter PZT hollow sphere with a 0.254 cm wall thickness, similar to the U.S. Navy standard Type F42 transducer,[18] was placed 30 cm from one end of the tube and a movable receiver, a Bruel and Kjaer model 8103 hydrophone, was inserted through the other end attached to a rod to allow for positional adjustment. The tube ends were capped and one was attached to a hand operated pressure pump[19] by a flexible hose to allow for rotation of the tube, which was used to adjust the hydrostatic pressure.

The sound speed in a tube deviates from that in the free field. For high bulk modulus fluids the difference between the propagation speed in free field and that in the plexiglass tube is considerable. However, for the medium under investigation, in the hydrostatic pressure range of interest, the difference will be relatively small. The low frequency propagation speed can be calculated from[20]

$$c = \sqrt{\frac{K}{\rho\left[1 + \frac{DK}{Ee}(1 - v^2)\right]}}, \tag{7}$$

where K is the bulk modulus of the fluid, D is the inner diameter of the tube, E is the Young's modulus of the tube material, e is the tube wall thickness, ρ is the density of the fluid, and v is Poisson's ratio for the tube material. Measurements made when the tube was filled with pure Castor oil, at room temperature, and the hydrostatic pressure varied between 100 and 800 kPa found the sound speed to be 810 ± 2 m/s over the entire pressure range. This speed compares with a theoretical value of 811 m/s obtained with $E=3$ GPa and $v=0.4$,[21] and a free-field sound speed of 1499 m/s.[15] The composite medium has a density of 867 kg/m^3 and is expected to have a sound speed on the order of 330 m/s in the low pressure region of interest. Thus, the sound speed corrections required to obtain free-field sound speeds are expected to be approximately 10% and are relatively small.

The sound speed of the composite medium was measured at various hydrostatic pressures at a frequency of 1.5 kHz. The hydrophone was initially placed 10 cm away from the projector and the received waveforms, gated sinusoids, were digitized and recorded for each hydrostatic pressure. The hydrophone was then moved 10 cm further away from the projector and the process repeated. The measured sound speeds, after correcting for the compliance of the tube, are plotted in Fig. 4. If all of the microspheres were initially unbuckled and behaving as rigid spheres the propagation speed measured in the tube should be 810 m/s and would be corrected to a free-field sound speed of the order of 1499 m/s, the same as Castor oil. Our measurements find that the free-field sound speed, near ambient pressure, is on the order of 450 m/s. This unexpected low value may be due to the presence of a very small number of broken or prebuckled microspheres. Using Wood's equation[22] one can determine the required void fraction of broken or prebuckled microspheres required to explain the low sound speed. The result is that only seven out of each 1000 microspheres need to be broken to reduce the uncorrected sound speed in the tube, at ambient pressure, by a factor of one-half. This fraction of broken microspheres is observed routinely when samples are examined with a microscope. As previously stated, the quality control on the manufacturing process of the microspheres is relatively minor.

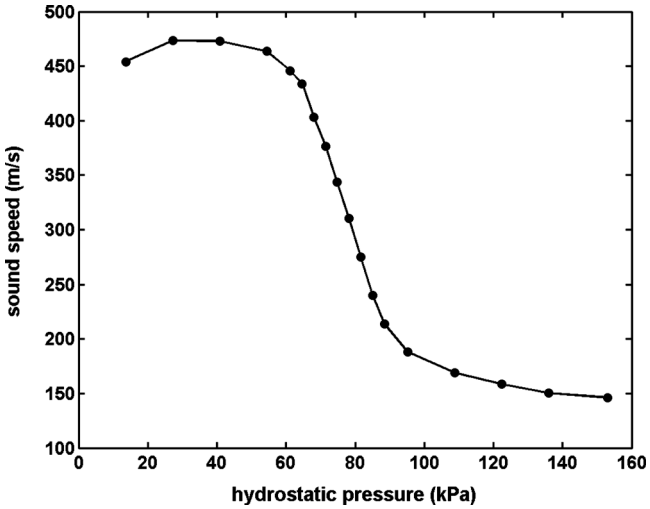


FIG. 4. The free-field sound speed of the composite medium as a function of hydrostatic pressure. The measurements obtained in the plexiglass tube, solid circles, have been corrected for the compliance of the measurement system, using Eq. (7), to yield the free-field values.

The measured sound speed initially increases, with increasing hydrostatic pressure, as the broken or initially buckled microspheres collapse. However, in the vicinity of 60 kPa the sound speed decreases rapidly as the initially intact microspheres begin to buckle. The sound speed decrease is nearly linear between 64 and 88 kPa and in this region $(\partial c/\partial p)_T$ is 9.7×10^{-3} m/s/Pa, which is approximately 6000 times larger than the corresponding increase for pure water. As the hydrostatic pressure is increased further, the rate of increasing compliance decelerates until maximum compliance is reached in the vicinity of 150 kPa. With further increase in hydrostatic pressure the sound speed increases and asymptotically approaches that of Castor oil, although not shown in Fig. 4. Note, also that the slope of the curve is zero in two locations, in the vicinity of 35 kPa and 150 kPa.

The static properties illustrated in Fig. 2 were used to calculate the density of the composite medium as a function of hydrostatic pressure. These results along with the data plotted in Fig. 4 were used to calculate the first term on the right-hand side of Eq. (5). The significance of the first term is that, in general, the second term is negligible in comparison and thus, the ratio B/A, is simply given by the first term. The results are plotted in Fig. 5 where the initial large positive value at 20 kPa is believed to be associated with compression of the damaged or prebuckled microspheres. These microspheres should totally compress quickly and have little impact on the measured values at higher hydrostatic pressures. The measured values between 50 and 100 kPa are un-

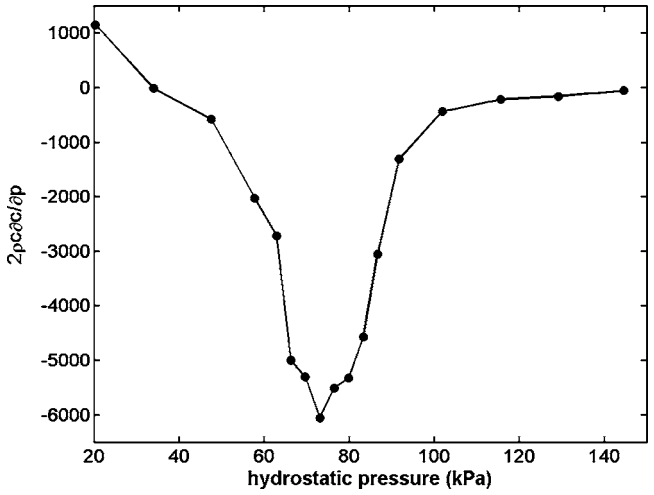


FIG. 5. The value of the first term on the right-hand side of Eq. (5), $2\rho c(\partial c/\partial p)_T$, as a function of hydrostatic pressure. Note the large negative values between 40 and 100 kPa. The second term on the right-hand side of Eq. (5) is negligible resulting in this being a plot, in essence, of B/A as a function of hydrostatic pressure.

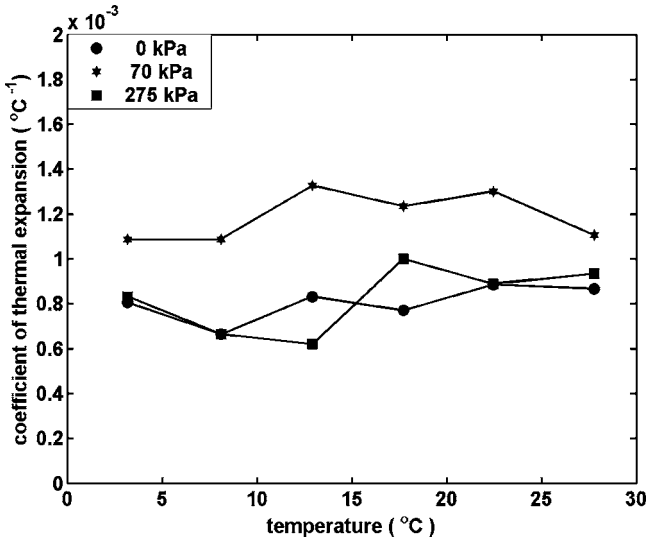


FIG. 6. The measured coefficients of thermal expansion of the composite medium versus temperature for three different hydrostatic pressures. The lowest pressure is where none of the microspheres are buckled; the intermediate hydrostatic pressure is close to that of the maximum compliance of the medium; and at the highest hydrostatic pressure the microspheres are totally collapsed.

usual in that they are very large and negative. This hydrostatic pressure range is adjustable by varying the size and shell thickness distribution of the microspheres, as can be done with the use of one of the 10 different grades of microspheres manufactured by Expancel, and the magnitude of the effect can be varied by changing the initial void fraction.

B. Measurement of the coefficient of thermal expansion

The coefficient of thermal expansion of the composite medium was measured in the dilatometer. A mixture of 90% Castor Oil and 10% microspheres was introduced and the volume of the sample measured, at a fixed hydrostatic pressure, as the temperature was varied. This was accomplished by adjusting the hand pump to maintain a fixed hydrostatic pressure after each temperature change occurred and while the sample stabilized. The temperature was varied in approximately five (5) degree Centigrade steps and the volume change, after subtracting out the calibrated contribution due to the expansion of the dilatometer, was used to approximate the derivative, $(\partial v/\partial T)_p$. The measurements were performed at three distinct hydrostatic pressures to investigate the behavior of the composite medium. The first measurement was at ambient pressure where the majority of the microspheres are spherical; the second measurement was at 70 kPa, in a region where the some of the microspheres are buckled and the medium is highly compliant; and the third was at 275 kPa, well beyond the highly nonlinear region where all of the microspheres have buckled and most have been fully compressed. The lower and upper pressure measurements should be similar to that of pure Castor oil which has a coefficient of thermal expansion of 7.3×10^{-4} K^{-1} and the measurements, shown in Fig. 6, are close to this value. The measurements at 70 kPa are higher than those at the other hydrostatic pressures and are close to that given by a simple mixture rule for 10% air and 90% Castor oil. Using these measurements and a simple mixture rule for air and Castor oil to approximate the specific heat of the medium, the coefficients in front of $(\partial c/\partial T)_p$ are approximately 0.1 K m/s. Given that the sound speed is on the order of 300 m/s in the highly nonlinear region, it is unlikely that the sound speed variation with temperature can be of such a magnitude that the second term on the right-hand side of Eq. (5) is significant. A simple experiment was designed to verify this conclusion.

C. Measurement of $(\partial c/\partial T)_p$

Rather than construct an environmental chamber adequate to house the sound speed measurement apparatus and determine the variation in speed as a function of temperature, a simple shell was constructed to immerse the plexiglass tube in an ice-water mixture. With this configuration, the sound speed at 0 °C could be measured and compared to that at ambient temperature to obtain the sound speed change as a function of temperature. This value, while over a large change in temperature, could be used to approximate the order of magnitude of the derivative. Of interest was the sound speed in the unbuckled region, at 25 kPa, and at the hydrostatic pressure of maximum nonlinearity of the medium, 72 kPa. The measured sound speeds at 0 °C, corrected for tube compliance, where 314 m/s at 25 kPa and 108 m/s at 71 kPa. Using these values $(\partial c/\partial T)_p$ was approximated and found to be 7 m/s/K at 25 kPa and 11 m/s/K at 71 kPa. These values indicate that the second term on the right-hand side of Eq. (5) is of order unity and insignificant relative to the first term, in the hydrostatic pressure region of high negative nonlinearity. Thus, Fig. 5 is essentially a plot of B/A for the composite medium and a detailed investigation of the temperature dependence of the sound speed in the composite medium is unwarranted. Note that the maximum value of B/A is approximately 6000 and is negative. This should be compared with a positive value of 5.5 for water.

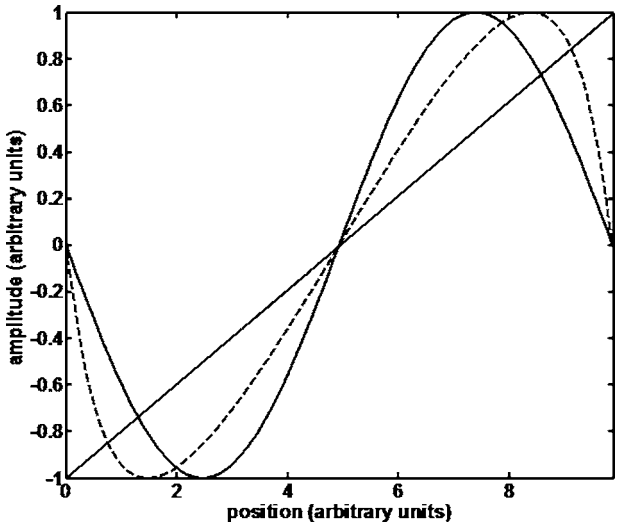


FIG. 7. Propagation of an initially sinusoidal waveform in a medium with a positive coefficient of nonlinearity. The waveform is propagating to the right and the observed is moving with the infinitesimal amplitude sound speed c_c. The positive amplitude portion of the initial sinusoidal waveform moves faster than c_c while the negative amplitude portion moves at a slower speed, as illustrated by the dashed curve. Ultimately, if linear attenuation is low enough, the distortion leads a compressive shock wave or N wave, solid line.

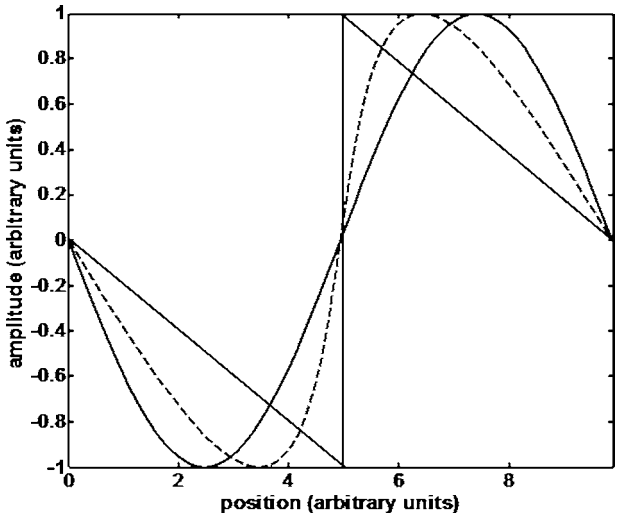


FIG. 8. Propagation of an initially sinusoidal waveform in a medium with a negative coefficient of nonlinearity. The waveform is propagating to the right and the observed is moving with the infinitesimal amplitude sound speed c_c. The negative amplitude portion of the initial sinusoidal waveform moves faster than c_c while the positive amplitude portion moves at a slower speed, as illustrated by the dashed curve. Ultimately, if linear attenuation is low enough, the distortion leads a rarefactive shock wave, solid line.

D. Nonlinear propagation measurements

While we have determined, from thermodynamic relations, the approximate value of B/A we have not demonstrated that the medium dynamically behaves in a nonlinear manner. Only quasistatic measurements were used to calculate B/A. It has not been demonstrated that the microspheres buckle and unbuckle quickly enough to impact the properties of the medium as a low frequency signal passes. It has also not been demonstrated that the coefficient of nonlinearity is negative and would lead to rarefactive shockwaves, if the attenuation was sufficiently low. This can only be demonstrated by measuring the distortion of the waveform or the growth of harmonics as the waveform propagates.

The difference between a positive and negative value of the coefficient of nonlinearity determines the manner in which a finite amplitude wave distorts as it propagates. Figure 7 illustrates the distortion of an initially sinusoidal finite amplitude waveform as it propagates in a medium with a positive coefficient of nonlinearity. The waveform is propagating toward the right of the plot and the observer is moving with the infinitesimal propagation speed, c_0. The positive coefficient of nonlinearity results in the higher pressure portion (positive pressure in plot) of the waveform propagating faster than c_0 and the lower pressure portion propagating slower than c_0, as illustrated by the dashed curve. Ultimately, if the attenuation is low enough, a compressive, shockwave or N wave is achieved. Figure 8 illustrates the distortion of the same sinusoidal finite amplitude waveform as it propagates in a medium with a negative coefficient of nonlinearity. In this case the lower pressure portion of the waveform propagates faster, and the positive pressure propagates slower than c_0, as illustrated by the dashed curve. In this case, if the attenuation is sufficiently low, a rarefactive shockwave is achieved, as illustrated.

The ultimate demonstration of the magnitude and sign of the coefficient of nonlinearity of the composite medium would be to measure the distortion of an initially sinusoidal waveform, noting the direction in which the waveform distorts, and compare the rate of distortion with a nonlinear propagation model. The discontinuity distance,[23] l, is the point at which an initially sinusoidal waveform would shock if linear absorption were absent and for a plane wave is given by

$$l = \frac{c_0^2}{(1 + B/2A)U_0\omega_0}, \tag{8}$$

where U_0 is the initial particle velocity of the sinusoidal waveform. Driving the acoustic source in the propagation tube with maximum voltage obtainable with the electronics, 1000 volts at 1.5 kHz, the discontinuity distance is approximately 1.5 m. Given the geometry of the tube, the maximum propagation distance that could be measured is on the order of 40 centimeters which would be sufficient to observe distortion in the absence of linear absorption. However, the relative importance of nonlinear effects is given by the Goldberg number,[24] $1/\alpha l$, where α is the coefficient of linear attenuation. If the Goldberg number is greater than unity the nonlinear effect is important and shocks are likely. If the Goldberg number is much less than unity linear attenuation dominates and significant distortion of the waveform does not occur. Unfortunately, in this medium even though the coefficient of nonlinearity is very large, the Goldberg number, at 1.5 kHz, is approximately 0.095. Significant distortion of the waveform will not be observed over 40 centimeters of propagation. Thus, the demonstration of the dynamic nonlinearity of the medium can only be accomplished by measuring the growth of the harmonics as the waveform propagates and comparing the results to a nonlinear propagation model. The sign

of the coefficient of nonlinearity will be determined by the relative phase of the harmonics. A Fourier decomposition of the distorted waveform in Fig. 7 consists of a series of harmonics all of which are of the same phase. The same decomposition for the distorted waveform in Fig. 8 consists of a series of harmonics in which the odd harmonics are of opposite phase relative to the even harmonics. Thus, an observation of the magnitude and phase of the first harmonic of the waveform, as a function of distance, will provide a measure of the magnitude and sign of the coefficient of nonlinearity.

In the sound speed measurement system, both the acoustic source and the hydrophone are encapsulated. When the two are positioned as close as possible, the acoustic centers are two (2) centimeters apart. Even though the two encapsulating materials that are in contact have low coefficients of nonlinearity, there is sufficient medium surrounding the two transducers to produce significant harmonics in this short distance. Thus, one cannot start with a single frequency tone and observe the harmonic production as the signal propagates. Instead, we chose to drive the source with four (4) different high voltage levels. The waveforms were recorded as close as possible to the source and again after 20 centimeters of propagation. The amplitudes of the harmonics in each waveform were then obtained by performing a least-square error fitting of a Fourier series expansion and the results compared to numerical predictions.

The drive frequency was arbitrarily chosen to be 1.5 kHz and the linear attenuation in the composite medium at the primary, second, and third harmonic frequencies was measured with the lowest level signal possible. This information was required as input into the numerical algorithm used to determine the value of B/A from the propagation measurements. The linear attenuation at 3 kHz is not inconsequential with the measured attenuation coefficient being −6.93 Np/m. Thus, significant harmonic generation is necessary simply to overcome the linear loss. To maximize the harmonic generation, the nonlinear measurements were performed at a hydrostatic pressure of 71 kPa where the nonlinear coefficient was determined, on the basis of quasistatic measurements, to have the largest value.

The propagation was modeled using the finite amplitude wave algorithm of Trivett and Van Buren.[25] The propagation was assumed to be spherical until the duct walls are reached and then to propagate as a plane wave. The initial amplitude of the first three harmonics was input and the amplitude of the first harmonic, after propagating 20 cm, was determined for each of the four drive levels. A least-square error analysis with steps of 50 was performed to obtain the best value of β, $(1+B/2A)$, that fit the measured data. In Fig. 9 we have plotted the magnitude of the first harmonic, after 20 cm of propagation, versus the initial amplitude of the primary 1.5 kHz signal. We have chosen to plot the magnitude of the first harmonic in Fig. 9. It is, however, opposite in phase relative to the fundamental as one would expect for a negative coefficient of nonlinearity. Also included in the plot is the best fit numerical curve obtained with a value of −3200 for β and a curve corresponding to the expected magnitude of the first harmonic if it had propagated linearly subject

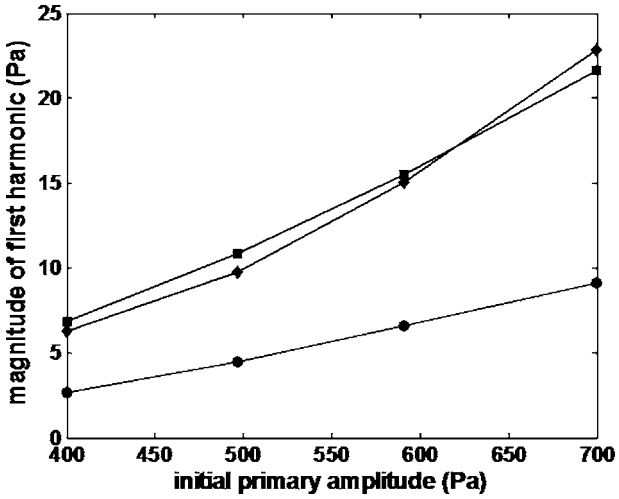


FIG. 9. The magnitude of the first harmonic after 20 centimeters of propagation, solid squares, versus the initial amplitude of the primary frequency. The measured results are compared with those of a numerical nonlinear propagation analysis, solid diamonds, using a value of −6398 for B/A, which was obtained from a least-square error analysis, and with value expected if the initial measured amplitude of the first harmonic had propagated linearly, solid circles.

only to spreading loss and linear attenuation. The numerical predictions are in good agreement with the measured results and the value of −3200 for β, or −6398 for B/A, is in reasonable agreement with the results of the quasistatic investigation plotted in Fig. 5. The results of the propagation measurements indicate that, at least up to1.5 kHz, the microspheres buckle and unbuckle quickly enough to impact the properties of the medium: the composite medium is highly nonlinear; and that the coefficient of nonlinearity is negative.

E. Conclusion

We have investigated the nonlinearity of a composite medium which has physical properties that are highly dependent upon hydrostatic pressure. As a result, there is a hydrostatic pressure range which has a very high, negative coefficient of nonlinearity. For the composite medium investigated, the region of high nonlinearity extended from approximately 40 to 100 kPa and B/A was measured both quasistatically and dynamically to have a peak value around −6000. Both the magnitude of the effect and the hydrostatic pressure range over which it occurs can be adjusted by varying the void fraction and the size distribution of the microspheres. In addition, while Castor oil was used in this investigation, we have used both Xanthan gelatin[26] and several different low durometer Urethanes as matrix materials and obtained similar results. The composite medium made with the gelatin has exactly the same physical properties as the Castor oil based medium, while the physical properties of the Urethanes impact the hydrostatic pressure range over which the microspheres buckle. The shear modulus of the Urethane must be taken into account when developing a material for a specific application. All of these materials have been used to enhance the difference frequency generation of a parametric source and will be the subject of a forthcoming paper.

ACKNOWLEDGMENT

This work was supported by the Rae and Frank H. Neely Endowment to the Woodruff School of Mechanical Engineering in the Georgia Institute of Technology.

[1]Expancel® 091 DE 80 d30 microspheres were used. Expancel®, Box 13000, S-850 13 Sundsvall, Sweden is part of Caseo Products a business unit of Akzo Nobel.

[2]R. T. Beyer, "Parameter of nonlinearity in fluids," J. Acoust. Soc. Am. **32**, 719–721 (1960).

[3]W. M. Madigosky, I. Rosenbaum, and R. Lucas, "Sound velocities and B/A in fluorocarbon fluids and in several low density solids," J. Acoust. Soc. Am. **69**, 1639–1643 (1981).

[4]P. J. Westervelt, "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537 (1963).

[5]X. F. Gong, S. G. Ye, D. Zhang, S. S. Feng, R. Q. Zhang, R. T. Wang, Z. Z. Xu, L. M. Liu, and K. L. Ha, "Acoustical nonlinearity parameter of liquids with microbubbles," J. Acoust. Soc. Am. **103**, 2960–2961 (1998).

[6]J. Wu and Z. Zhu, "Measurements of the effective nonlinearity parameter B/A of water containing trapped cylindrical bubbles," J. Acoust. Soc. Am. **89**, 2634–2639 (1991).

[7]R. E. Apfel, "The effective nonlinearity parameter for immiscible liquid mixtures," J. Acoust. Soc. Am. **74**, 1866–1868 (1983).

[8]E. C. Everbach, Z. Zhu, P. Jiang, B. T. Chu, and R. E. Apfel, "A correct mixture law for B/A," J. Acoust. Soc. Am. **89**, 446–447 (1990).

[9]Yu. A. Kobelev and A. M. Sutin, "Difference-frequency sound generation in a liquid containing bubbles of different size," Sov. Phys. Acoust. **26**, 485–487 (1981).

[10]E. F. Kozyaev and K. A. Naugol'nykh, "Parametric sound radiation in a two-phase medium," Sov. Phys. Acoust. **26**, 48–51 (1980).

[11]L. M. Kustov, V. E. Nazarov, L. A. Ostrovsky, A. M. Sutin, and S. V. Zamolin, "Parametric acoustic radiation with a bubble layer," Acoust. Lett. **6**, 15–17 (1982).

[12]O. A. Druzhinin, L. A. Ostrovsky, and A. Prosperetti, "Low frequency acoustic wave generation in a resonant bubble layer," J. Acoust. Soc. Am. **100**, 3570–3580 (1996).

[13]S. Karpov, L. A. Ostrovsky, and A. Prosperetti, "Nonlinear wave interactions in bubble layers," J. Acoust. Soc. Am. **113**, 1304–1316 (2003).

[14]Ricinus oil, CA208 from Spectrum Chemical Mfg. Corp., 14422 S. San Pedro St., Gardena, CA 90248.

[15]R. W. Timme, "Speed of Sound in Castor Oil," J. Acoust. Soc. Am. **52**, 989–992 (1972).

[16]J. M. Stallard, "Some thermodynamic properties of castor oil vs. temperature and pressure," Naval Ordinance Laboratory Technical Report No. 66-68, Naval Ordnance Laboratory, White Oak, Md, April 1966.

[17]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon, New York, 1959).

[18]I. D. Groves, Jr., "Twenty Years of Underwater Electroacoustic Standards," Naval Research Laboratory Report No. 7735, 21 February 1974.

[19]Hand operated pressure pump with vernier, Model 87-6-5, High Pressure Equipment, Company, 1222 Linden, Erie, PA 16505.

[20]J. A. Swaffield and A. P. Boldy, *Pressure Surge in Pipe and Duct Systems* (Ashgate Publishing Company, Brookfield, Vermont 1993), p. 31.

[21]http://me203.stanford.edu/materials/technical_properties.html#Young's

[22]A. B. Wood, *A Textbook of Sound*, 2nd ed. (MacMillan, New York, 1941), p. 361.

[23]R. T. Beyer, *Nonlinear Acoustics* (Naval Sea Systems Command, Washington, DC, 1974), p. 104.

[24]Z. A. Goldberg, Sov. Phys. Acoust. **3**, 346 (1956).

[25]D. H. Trivett and A. L. Van Buren, "Propagation of plane, cylindrical, and spherical finite amplitude waves," J. Acoust. Soc. Am. **69**, 943–949 (1981).

[26]P. A. Hwang, R. A. Roy, and L. A. Crum, "Artificial bubble cloud targets for underwater acoustic remote sensing," J. Atmos. Ocean. Technol. **12**, 1289–1302 (1995).

Strong on-axis focal shift and its nonlinear variation in low-Fresnel-number ultrasound beams

Y. N. Makov
Department of Acoustics, Faculty of Physics, Moscow State University, 119899 Moscow, Russia

V. Espinosa, V. J. Sánchez-Morcillo, J. Ramis, J. Cruañes, and F. Camarena
Departament de Física Aplicada, Escuela Politécnica Superior de Gandia, Universitat Politècnica de València, Crta. Natzaret-Oliva s/n, 46730 Grau de Gandia, Spain

(Received 19 April 2005; revised 20 March 2006; accepted 27 March 2006)

We examine the case of focusing transducers with a strong initial (in linear regime) shift of the main on-axis peak pressure maximum from the geometrical focal point toward the transducer. Such transducers are characterized, using the concepts introduced in this paper, by a low Fresnel number. The displacement of this initially shifted on-axis pressure maximum point toward the geometrical focus, and its backward motion as the driving transducer voltage increases until highly nonlinear regimes, has been experimentally observed. The simultaneous monitoring of the temporal wave-form distortion determines the real roles and interplay between different nonlinear effects (refraction and attenuation) in the observed dynamics of the on-axis pressure maximum. The numerical solution of the corresponding mathematical model confirms the physical interpretation of the observed phenomenon. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197805]

PACS number(s): 43.25.Cb, 43.25.Jh [MFH] Pages: 3618–3624

I. INTRODUCTION

Today, focused ultrasonic beams are well known and traditional objects of theoretical (analytical and numerical)[1–3] and experimental[4] investigations, mainly because of their utilization as multifunctional instruments in many applied technologies, for example, in medical therapy,[5] nondestructive testing,[6] and others. But the interplay among the main factors like diffraction, nonlinearity, and focusing still keep some open questions and misunderstood details concerning sound beams. Furthermore, the definition of the role of each factor, without taking into account their joint action, can lead to incomplete and as a consequence to seemingly contradictory conclusions. For example, in Refs. 3 (p. 915) and 7 (p. 978) a shift of the on-axis pressure maximum away from transducer is predicted, while in Ref. 8 (see p. 160) the opposite effect (a shift toward the transducer) is described. In both cases, the common effect of shock front formation was claimed as being responsible of such shifts. Therefore there is a stimulus for the clarification of the situation with the change of the on-axis pressure maximum position under the transition from linear to nonlinear beam propagation regimes as the driving transducer voltage is increased. This is the main aim of the present paper. But for the realization of this aim it is necessary to start in linear regime with the perceptible (sufficiently large) initial shift of the maximum pressure position from the geometrical focus. Such situation is intrinsic to the so-called low-Fresnel-number focused transducers. This problem has been treated in optics and is known as the focal shift effect.[9,10] The concept of Fresnel number is a widely used characteristic in optical systems[11] and, although it coincides formally (with a proportionality factor of π^{-1}) with the linear gain of focused acoustic beams, its use for describing the beam and transducer properties is more convenient than the value of the gain. This will be discussed in the next section. Also we present an improvement of the estimation for the on-axis maximum pressure location of a focused transducer in the linear regime, which has been given previously by other authors.[12,13]

Our experiments show that, as the driving transducer voltage is increased, the position of the point of maximum pressure shifts in the beginning (for moderate driving voltage) from its initial linear position toward the geometrical focus, but this tendency is inverted and the motion is backward as the nonlinear regime become stronger. The parallel observation of the temporal wave form (profile) and its modifications served as the best indicator of nonlinearity degree. For the assurance that such dynamics of maximum pressure point results only from the above-mentioned three factors but not on others (e.g., thermal effects) the experimental results were confirmed by numerical solutions of a nonlinear Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation, which accounts for focusing, diffraction, and nonlinearity. It is interesting to note that a similar dynamics of on-axis pressure maximum point can be found as a secondary result in Ref. 14, although this effect was interpreted as resulting from the imperfection of the measuring hydrophone (i.e., the real cause was not indicated).

The paper is organized as follows: In Sec. II some facts about the Fresnel number as a characteristic of focused acoustic beams or transducers, in parallel with the determination of the correct dependence of on-axis pressure maximum point location versus the Fresnel number of the transducer, are given. In Sec. III the experimental procedure is described and its results are presented. In Sec. IV the numerical results are reported. The discussion and interpretation of all results are contained in Sec. V. Finally, Sec. VI represents the conclusion.

 0001-4966/2006/119(6)/3618/7/$22.50

II. NOTES ABOUT CHARACTERISTICS OF FOCUSED TRANSDUCERS AND BEAMS IN THE LINEAR REGIME

The basic characteristics of a focused transducer are the aperture dimension (e.g., its radius a), which in certain relation with the wavelength λ (or wave number k) gives the diffractional (Rayleigh) length $L_d=ka^2/2$, and the radius of curvature of the transducer (geometrical focal length) R. The quotient of these characteristics L_d/R is the simplest nondimensional parameter characterizing the relative but oppositely directed influences of diffraction and focusing on the structure of the irradiated beam.

This parameter has a twofold significance in the present context. On one hand it connects directly with the Fresnel number N_F, well known in the optical context.[11] For a focused transducer, the depth h (axial distance from transducer to the aperture plane), in the Fresnel approximation $a \ll R$ is given by

$$h = R - \sqrt{R^2 - a^2} \approx \frac{a^2}{2R}. \tag{1}$$

Then, the Fresnel number N_F corresponds to the number of half wavelengths contained in h, which using Eq. (1) can be expressed as

$$N_F = \frac{h}{\frac{\lambda}{2}} = \frac{a^2}{\lambda R} = \frac{1}{\pi}\frac{L_d}{R}. \tag{2}$$

In this interpretation the condition $N_F \gg 1$ corresponds to a large initial wave-front curvature (i.e. the effect of strong focusing) and the geometrical approach with the maximum of the field at the centre of curvature F, where all rays converge, is justified. In the opposite case $N_F \ll 1$ the effect of the initial wave-front curvature (i.e., the focusing effect) is weak, and the structure of the irradiated field approaches the result from a plane circular transducer.

On the other hand, in the theory of focused acoustic beams the general and direct meaning (as relation between diffraction and focusing) of this quotient is traditionally diminished by the more specific and narrow interpretation as the linear gain of a focused transducer, determined through the ratio of pressure in the geometrical focal point to the pressure near the transducer surface, and is equal to (see, e.g., Refs. 2 and 13)

$$G = \frac{p(0,R)}{p_0} = \frac{L_d}{R}, \tag{3}$$

although, from Eq. (3), it is clear that the value of G also reflects the relation of diffraction and focusing. Recall that the expression (3) for the linear gain follows from the solution of the wave equation for the pressure distribution.[13]

In spite of the direct proportionality between both characteristics ($N_F=\pi^{-1}G$) there is a reason for the preferable use of the Fresnel-number conception as the common diffractional-focusing characteristic of transducers, and the classification of focusing transducers (and the corresponding irradiated beams) into high- and low-Fresnel-number cases is justified. The term "gain" remains for its original and narrow meaning for focusing transducers and, furthermore, the real value of the gain for low-Fresnel-number transducers notably differs from the traditional simplest expression (3), as will be shown below.

Our interest here concerns the location of the on-axis maximum pressure point in the linear regime of a focused transducer and how this point deviates from the geometrical value when the Fresnel number is decreased. Of course, it is well known in sound beam theory that the on-axis maximal pressure does not coincide with the geometrical focus, since this would occur only in the unrealistic case of vanishing diffraction. However, the case of strong shift for low-Fresnel-number transducers has not been explicitly considered in acoustics, and the well-known approximate analytical estimation for this shift proposed in Refs. 12 and 13 is not correct for the low-Fresnel-number case. This result follows from the exact analytical solution for the pressure complex amplitude $A(r,z)$, determined by the ordinary wave equation in the parabolic approximation[13]

$$A(r,z) = -\frac{ik}{z}\exp\left(\frac{ik}{2z}r^2\right)\int_0^a \exp\left(\frac{ik}{2z}r'^2\right)J_0\left(\frac{k}{z}rr'\right) \times A(r',0)r'dr', \tag{4}$$

with the acoustic pressure being $p(r,z,t)=A(r,z)e^{ikz-i\omega t}$.

In the simplest case of constant pressure along the aperture, and parabolic phase accounting for the focusing effect, the initial condition reads

$$A(r',0) = p_0 \exp\left(-\frac{ikr^2}{2R}\right), \tag{5}$$

for which the on-axis pressure distribution can be calculated from Eq. (4), and results

$$\frac{p(0,\tilde{z})}{p_0} = \left|\frac{2}{1-\tilde{z}}\sin\left(\frac{\pi N_F}{2}\frac{1-\tilde{z}}{\tilde{z}}\right)\right|, \tag{6}$$

where the bars denote absolute value, p_0 is the constant pressure along the transducer aperture, and $\tilde{z}=z/R$ is a dimensionless coordinate along the beam axis. Taking the derivative of Eq. (6) and equating it to zero, the locations of the points of pressure extremum can be found as the solutions of the transcendental equation

$$\frac{tg(X)}{X/\tilde{z}} = 1, \tag{7}$$

where $X(\tilde{z})=\frac{\pi N_F}{2}\frac{1-\tilde{z}}{\tilde{z}}$.

The root of Eq. (7) for which the pressure distribution in Eq. (6) takes the largest value, corresponds to the location of the main ("focal") maximum of on-axis pressure. The result of the numerical calculation of this root ($\tilde{z}_{\max}$) from Eq. (7) depending on the Fresnel-number N_F is shown in Fig. 1. For comparison, also the approximate result $\tilde{z}_{\max}=1-12/(\pi N_F)^2$ given by Eq. (20) from Ref. 13 is presented. From Fig. 1 it follows: First, for $N_F>6$ (the high-Fresnel-number focusing transducers) the main pressure maximum is very close to the geometrical focus, but for $N_F<3$ (low-Fresnel-number focusing transducers with realistic actual geometrical parameters) the difference between these points

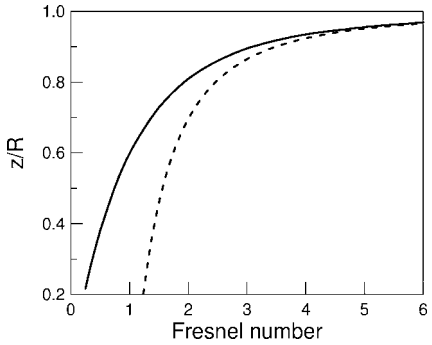


FIG. 1. Dependence of the position of the on-axis main pressure maximum on Fresnel number N_F according to the exact solution (solid curve) and the approximate estimation from Ref. 13 (dashed curve).

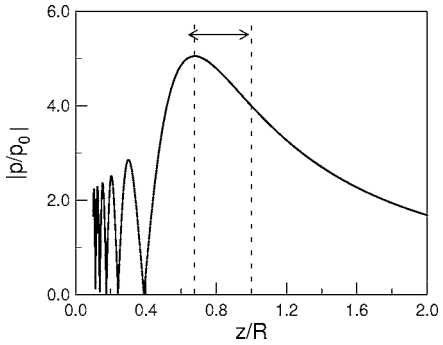


FIG. 3. On-axis pressure distribution in linear regime for the transducer used in experiments. The arrow shows the shift with respect to the geometrical focus.

is large, and the point of the main pressure maximum is strongly shifted toward the transducer. Second, the indicated estimation for the shift from Ref. 13 underestimates the value shift in the low-Fresnel-number region.

At the same time it follows from Eq. (6) and from the solution of Eq. (7) that the maximum pressure normalized to transducer surface's pressure $p(0,\tilde{z}_{\max})/p_0$ (which in fact represents the real gain) is not equal but higher (very notably in the low-Fresnel-number region) than the traditional G as defined in Eq. (3) at the geometrical focal point, where the pressure is not maximum. Figure 2 illustrates these statements.

In the next section we validate experimentally these results for the linear regime, and explore the behavior of focal point location in the weakly and strongly nonlinear regimes. For this aim, a transducer where a large shift of the maximum pressure point from geometrical focus occurs has been chosen. From the previous theory, this situation corresponds to small-Fresnel-number values. Note that combining the usual condition $a/\lambda \gg 1$ and considering a low-Fresnel-number transducer leads to

$$N_F = \frac{a}{\lambda}\frac{a}{R} = \frac{a}{\lambda}\sin\alpha < 3, \tag{8}$$

which imposes a condition on the half aperture of the transducer. Under these requirements, we consider for our experimental work a serial focused transducer worked in water at frequency $f=1$ MHz ($\lambda=0.15$ cm) with $a=1.5$ cm and $R=11.7$ cm, for which $N_F=1.28$, and according to the previous theoretical results, the main maximum pressure point in the linear regime localizes at a distance $0.67R=7.8$ cm from the transducer. The complete distribution of the on-axis pressure for this concrete transducer is shown in Fig. 3, obtained on the basis of Eq. (6). Here the difference between the values of real and traditional gains is clearly appreciated.

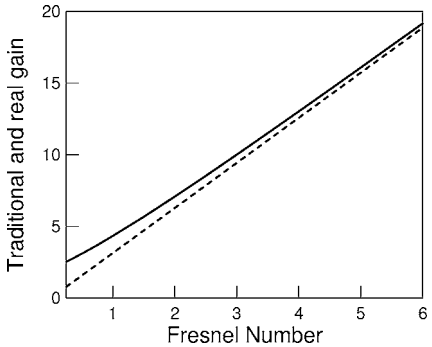


FIG. 2. Dependence of the focusing linear gain with the Fresnel number. The solid curve represents the real gain, as obtained numerically from Eq. (6). The dashed curve represents the traditional value of the gain, with a linear dependence.

III. EXPERIMENTAL RESULTS

The experimental setup follows the classical scheme of confronted emitting transducer and receiving hydrophone in a water tank. We use a Valpey-Fisher focused transducer based on a piezoelectric spherical shell of curvature radius of 11.7 cm and diameter of 3 cm working at 1 MHz. The transducer is driven by the signal provided by a programmable Agilent 33220 function generator, amplified by a broadband rf power amplifier either (depending on the power requirements) ENI 240L (40 W, +30 dB) or ENI 500A (500 W, +60 dB), which permits to deliver voltage amplitudes at the transducer terminals up to 750 Vpp without distortion, using an impedance matching filter. The emitted signal is detected by a calibrated membrane hydrophone NTR/Onda Corp. MH2000B, with a sensing aperture of 0.2 mm and a flat frequency response between 1 and 20 MHz, which allows for the registration of the wave-form profile confidently. All the signal generation and acquisition process is based on a National Instruments PXI-technology controller NI8176, which also controls an OWIS GmbH two-axis motorized system that moves the hydrophone with a repeatability of 10 μm: The tank, with dimensions $25\times25\times50$ cm^3, is made of plexiglass and it is filled with degassed water to avoid cavitation effects. The synchronized generation and capture of a tone burst of 20 cycles at 1 MHz with a very low repetition rate (2 Hz) ensures the detection of the direct signal without the superposition with reflections coming from the walls, and protects the transducer and the hydrophone from overheating.

The emitter and the hydrophone have been aligned by repetitively scanning the transversal plane at two different distances from the source, at both sides of the focal area, recording the maximum pressure amplitude in steps of 0.5 mm. The alignment of the transducer has been then con-

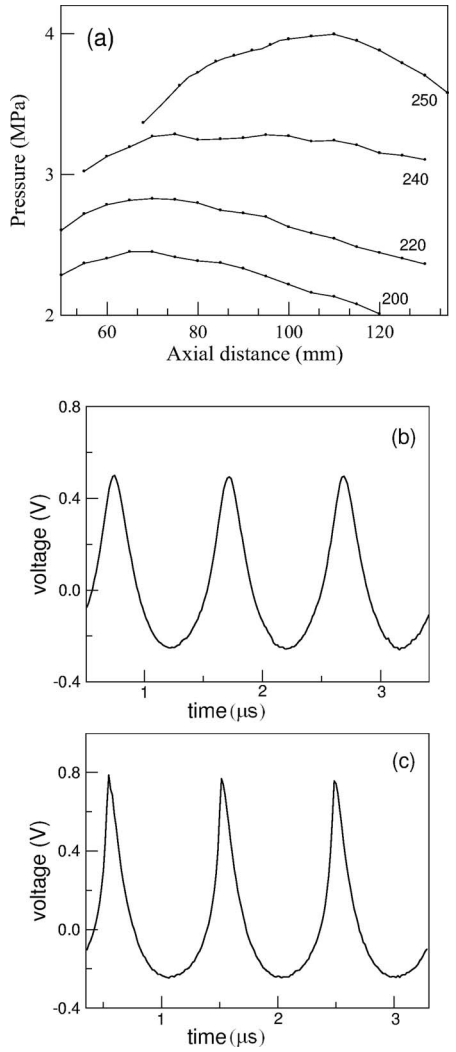


FIG. 4. Experimental on-axis pressure distributions, showing the forward motion of its maximum for increasing input voltages in the range 200–250 Vpp (a) and the corresponding wave forms in the neighborhood of maxima for the lower (b) and upper (c) curves.

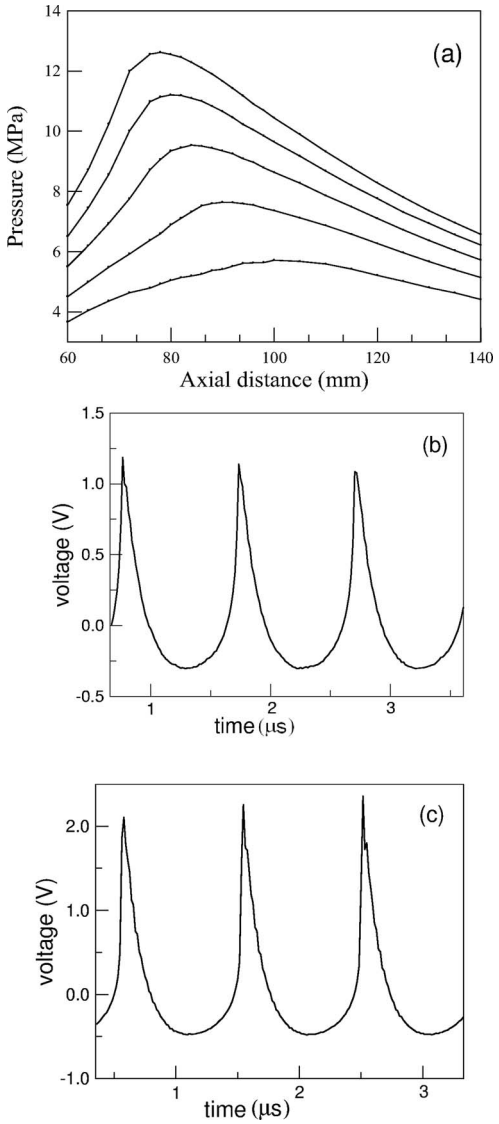


FIG. 5. Experimental on-axis pressure distributions, showing the backward motion of its maximum for increasing input voltage in the range 300–500 Vpp (a) and the corresponding wave forms with shock fronts for the lower (b) and upper (c) curves. Strong nonlinear regime.

sequently corrected until the displacement along the z axis from the near to the far field has corresponded to the maximum pressure line. Every point measurement is repeated ten times and averaged to minimize possible fluctuations due to noise.

The experimental results are summarized in Figs. 4 and 5, which show the measured on-axis pressure amplitude distributions (peak values) for increasing voltage inputs at the transducer terminals, in the range from 200 to 500 Vpp, together with the corresponding temporal profiles at selected driving values. Two qualitatively distinct regimes in the behavior of the maximum pressure point are disclosed. Figure 4(a) shows how this maximum moves initially away from the transducer until the neighborhood of geometrical focus, when the input voltage is increased from 200 to 250 Vpp. Figures 4(b) and 4(c) show the wave forms (temporal profiles) corresponding to the lower and upper curves in Fig 4(a), measured in the neighborhood of on-axis pressure maxima. These profiles experience, when increasing the driving amplitude, incremental nonlinear distortions but without the forming of shocks.

At higher driving voltages, the on-axis pressure distribution curves show maxima shifting to the opposite direction (toward the transducer), as illustrated in Fig. 5(a) for input voltage values from 300 to 500 Vpp. The wave forms corresponding to the lower and upper curves, measured in the neighborhood of maxima, are shown in Figs. 5(b) and 5(c), respectively. In this voltage range shocks with increasing amplitude developed in the wave forms. We conclude that, for the examined transducer, the input voltage level of 250 Vpp corresponds to the beginning of shock wave formation and, at the same time, the motion of the axial pressure maximum inverts its direction. The hydrophone marks indirectly the formation of shocks at the wave forms by its lack of sensitivity for detecting the very high frequencies at the wave-form maximum.

IV. NUMERICAL RESULTS

The phenomena reported in the previous section (strong focal shift and its dynamic behaviour) are caused by the

main effects such as focusing, diffraction, and nonlinearity, whereas other effects (e.g., thermal refraction) did not play an essential role. The demonstration of this assumption is provided by the numerical results obtained from a mathematical model which involves only these three effects.

The propagation of high intensity focused ultrasonic beam can be modeled with great accuracy by the Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation,[15] which for axisymmetric beams is

$$\frac{\partial^2 p}{\partial t' \partial z} = \frac{c_0}{2}\left(\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r}\right) + \frac{\delta}{2c_0^3}\frac{\partial^3 p}{\partial t'^3} + \frac{\beta}{2\rho_0 c_0^3}\frac{\partial^2 p^2}{\partial t'^2}, \quad (9)$$

where $t'=t-z/c_0$ is a retarded time, c_0 the propagation speed, δ the sound diffusivity, $\beta=1+B/2A$ the coefficient of nonlinearity, and ρ_0 the ambient density of the medium. Equation (9) is valid in the paraxial approximation ($ka \gg 1$) and takes into account nonlinearity, diffraction, and thermoviscous absorption by its corresponding terms, and the focusing through initial condition. Since the thermoviscous absorption of water is very small (as a physical factor), the corresponding term merely plays a mathematical role for obtaining single-valued solutions (for the exclusion of many valuedness in solutions).

Several numerical schemes have been proposed for solving Eq. (9). We have used the time domain algorithm described in Ref. 16. For a focused source of angular frequency ω_0 and with radius a and focal distance R, the following boundary condition is considered:

$$p = p_0 E(t)\sin\left[\omega_0\left(t + \frac{r^2}{2c_0 R}\right) + \Phi(t)\right] H\left(1 - \frac{r}{a}\right), \quad (10)$$

where p_0 is the peak pressure, $E(t)$ and $\Phi(t)$ are amplitude and phase modulations, respectively, of a sinusoidal signal and H is the Heaviside unit step function, which denotes the finite size of the source. The finite temporal duration of the signal in the experiment is modeled with the envelope function $E(t)=\exp[-(2t/T)^{2m}]$, where T is the effective duration of the pulse (containing 20 cycles in our case, according to the experiment) and m determines the rise time ($m=5$ in our case).

The different parameters appearing in Eqs. (9) and (10) were taken from the experimental conditions: for water at 20 °C, $\rho_0=998$ Kg/m^3, $c_0=1486$ m/s, and $\beta=3.5$. The geometrical parameters and frequency corresponded to the transducer used in the experiment, with initial pressures p_0 from 150 kPa to 1 Mpa. The results of the simulation are shown in Fig. 6, for increasing source amplitudes. These results are in agreement with the experimental observations, and confirm that the observed dynamic shift of the maximum pressure point is related only to the interplay between focusing, diffraction, and nonlinearity.

V. DISCUSSION

We have shown that the numerical results, in good agreement with the experiment, demonstrate that only the effects of diffraction and focusing (through the Fresnel number) and the nonlinearity are involved in the on-axis pressure maximum motion. On the basis of these effects we next discuss the features of this motion.

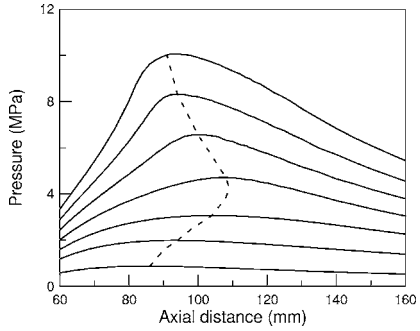


FIG. 6. On-axis pressure distributions obtained from numerical solutions of KZK equation.

The first feature is a strong initial (in linear regime of beam formation and propagation) shift of the on-axis pressure maximum from the geometrical focal point for low-Fresnel-number transducers, i.e., under the condition

$$N_F \equiv \frac{a^2}{\lambda R} \ll 1. \quad (11)$$

On one hand, condition (11) implies that the influence of the initial wave-front curvature (i.e., the focusing effect) is insignificant, and the structure of the irradiated field approaches that of a circular plane radiator. On the other hand, it is well known (see, e.g., Ref. 2) that for a plane circular transducer the location z_0 of the first (the furthest from transducer) on-axis field maximum is determined by the relation

$$N_F(z_0) \equiv \frac{a^2}{\lambda z_0} = 1. \quad (12)$$

From relations (11) and (12) it follows that, in the low-Fresnel-number case

$$z_0 \ll R, \quad (13)$$

i.e., the coordinate of the maximum pressure point is distinctly smaller than the geometrical focusing distance. This is the qualitative explanation of on-axis maximum field shift effect for the transducers with small values of N_F.

The second feature, of nonlinear nature, is the motion of the main on-axis pressure maximum from its initial position to the right (toward the geometrical focal point) under the increase of the input voltage in the range from 200 to 250 Vpp [see Figs. 4(a)]. Increasing driving voltage in this range results in an incremental nonlinear distortion of the wave form [see Figs. 4(b) and 4(c)] owing to the appearance of higher harmonics, but without shock front formation. The last circumstance implies the weakness of the nonlinear absorption effect (the thermoviscous absorption of water is very small and does not play an essential role here). The on-axis pressure distributions of the new nascent harmonics have maxima located closer to the geometrical focal point (but do not cross it). This tendency can be seen from the second-order solution (in the quasilinear approximation) of

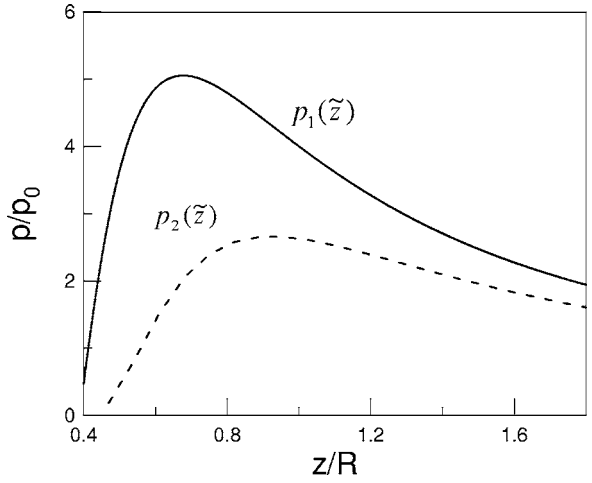


FIG. 7. On-axis plots for the first (solid curve) and second (dashed curve) pressure harmonics as analytical solutions of KZK equation in the quasilinear approximation.

the Khokhlov-Zabolotskaya equation. Under the initial condition (5) this solution for pressure on the axis has the following expression[17]

$$p_2(\tilde{z},0) = \frac{i\beta a^2 k^2 p_0^2}{\rho_0 c_0^2} \int_0^1 \int_0^1 xy e^{i(x^2+y^2)\pi N_F(\tilde{z}^{-1}-1)} \times \left(\int_0^\infty e^{it/2(x^2+y^2)} \frac{J_0(xyt)}{t\tilde{z}(\pi N_F)^{-1}+1} dt \right) dxdy, \quad (14)$$

where β is the parameter of nonlinearity and $J_0(x)$ the zeroth-order Bessel function of the first kind. In Fig. 7 the plots of the fundamental mode $p_1(\tilde{z})$ on the basis of Eq. (6), and the second harmonic $p_2(\tilde{z})$ on the basis of Eq. (14) are shown in the neighbourhood of their maximum values. Taking into account that during the process of nonlinear profile deformation (i) the number of higher-order harmonics and their maximum pressure values grow, with the position of the maximum of each harmonics located closer to the geometrical focus (see Fig. 7), and (ii) the maximum value of the fundamental pressure mode is decreased, it follows that the maximum of the resulting plot (on-axis pressure distribution), as composition of the first and rest of the harmonics, will approach the geometrical focus as new harmonics appear, as observed in the experiment. It is also possible to justify, nonrigorously but clearly, the considered effect (shift of the maximum to the right) by the increase of the virtual frequency (decrease of virtual wavelength) of the radiated beam due to the appearance of higher-order harmonics, resulting in an increase of the Fresnel number [see Eq. (1)] and consequently in an approach of the maximum to geometrical focus (see Sec. II and Fig. 2).

The third feature is the backward motion (toward the transducer) of the main pressure maximum along the axis [see Fig. 5(a)] under further increase of input voltage (within 250–500 Vpp range in our case). The distinctive effect in this stage is the growing shock front formation on the wave form [see Figs. 5(b) and 5(c)]. This strong nonlinear effect provokes the backward motion of pressure maximum. Under the gradual increase of the input voltage the formation of shock fronts on the wave form begins at the on-axis maximum pressure region and causes a high-frequency nonlinear absorption at this location (in a lossless medium, absorption is negligible for smaller distances). Therefore, in the initial stage, the beam amplitude in the region extending from the point of maximum pressure (located near the geometrical focus) to longer distances decreases in amplitude, and the part of the beam not affected by nonlinear absorption (closer to transducer) becomes the new maximum. As a consequence, the maximum shifts to the left during this process, as seen in Fig. 5(a).

VI. CONCLUSIONS

In this paper we have reported a large shift of the on-axis pressure maximum in the initial linear regime for ultrasound transducers. Also, a large range of movement of this maximum in the nonlinear regime has been observed and analyzed. The primary cause of the reported focal shift is the low-Fresnel number of the transducer. Therefore the classification of all transducers into low- and high-Fresnel-number cases is proposed here.

The parallel experimental observation of the varying on-axis pressure distribution and the temporal wave form, under the increase of the input voltage, allows us to disclose the role of nonlinearity at the different stages of its evolution, in the occurrence of one or another direction of maximum pressure shift.

The effects reported in this article, concerning the changes of on-axis pressure maximum in different regimes of focused acoustic beam propagation were confirmed several times during experiments, not only with the case presented here but also with other low-Fresnel-number transducers.

ACKNOWLEDGMENTS

The authors express their gratitude to Dr. Mark F. Hamilton for his advice during the preparation of the article. This work was supported by the Russian Foundation for Basic Researches (Grant No. 04-02-17009), and by the MEC of the Spanish Government, under Project No. FIS2005-07931-C03-02. The results reported in this article were confirmed by one of the authors (V.E.) during a research stay in the Hydroacoustics Laboratory of the Department of Physics and Technology of the University of Bergen. V.E. acknowledges Professor Halvor Hobaek for his hospitality and collaboration.

[1] N. S. Bakhvalov, Ya. M. Zhileikin, and E. A. Zabolotskaya, *Nonlinear Theory of Sound Beams* (American Institute of Physics, New York, 1987).
[2] M. F. Hamilton, "Sound beams" in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, CA, 1998), Chap. 8.
[3] O. V. Rudenko and O. A. Sapozhnikov, "Self-action effects for wave beams containing shock fronts," Phys. Usp. **174**, 907–922 (2004).
[4] C. M. Langton, *Principles and Applications of Ultrasound* (Institute of Physics Publishing, University of Reading, Berkshire, 2004).
[5] F. A. Duck, A. C. Baker, and H. C. Starritt, eds., *Ultrasound in Medicine* (Advanced Medical Publishing, Bristol, UK, 1998).
[6] J. Blitz and G. Simpson, *Ultrasonic Methods of Non-destructive Testing* (Chapman and Hall, London, 1996).
[7] O. V. Rudenko, "Nonlinear sawtooth-shaped waves," Phys. Usp. **38**, 965–990 (1995).
[8] K. Naugolnykh and L. A. Ostrovsky, *Nonlinear Wave Processes in Acous-*

tics (Cambridge University press, Cambridge, England, 1998), Sec. IV.
[9]Y. Li and E. Wolf, "Focal shifts in diffracted converging spherical waves," Opt. Commun. **39**, 211–215 (1981).
[10]For ultrasound transducers we prefer to speak about the shift of pressure maximum point, relating the terms about focal dynamics (including shift) to the case of active change of parameters (e.g., by a phased array).
[11]M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, Cambridge, England, 1999).
[12]H. T. O'Neil, "Theory of focusing radiators," J. Acoust. Soc. Am. **21**, 516–526 (1949).
[13]B. G. Lucas and T. G. Muir, "The field of a focusing source," J. Acoust. Soc. Am. **72**, 1289–1296 (1982).
[14]M. A. Averkiou and M. F. Hamilton, "Nonlinear distortion of short pulses radiated by plane and focused circular pistons," J. Acoust. Soc. Am. **102**, 2539–2548 (1997).
[15]V. P. Kuznetsov, "Equation of nonlinear acoustics," Sov. Phys. Acoust. **16**, 467–470 (1970).
[16]Y.-S. Lee and M. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[17]B. G. Lucas and T. G. Muir, "Field of a finite-amplitude focusing source," J. Acoust. Soc. Am. **74**, 1522–1528 (1983).
[18]P. T. Christopher and K. J. Parker, "New approaches to nonlinear diffractive field propagation," J. Acoust. Soc. Am. **90**, 488–499 (1991).

The effect of reflector geometry on the acoustic field and bubble dynamics produced by an electrohydraulic shock wave lithotripter

Yufeng Zhou[a] and Pei Zhong
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708

(Received 23 December 2005; revised 10 March 2006; accepted 13 March 2006)

A theoretical model for the propagation of shock wave from an axisymmetric reflector was developed by modifying the initial conditions for the conventional solution of a nonlinear parabolic wave equation (i.e., the Khokhlov–Zabolotskaya–Kuznestsov equation). The ellipsoidal reflector of an HM-3 lithotripter is modeled equivalently as a self-focusing spherically distributed pressure source. The pressure wave form generated by the spark discharge of the HM-3 electrode was measured by a fiber optic probe hydrophone and used as source conditions in the numerical calculation. The simulated pressure wave forms, accounting for the effects of diffraction, nonlinearity, and thermoviscous absorption in wave propagation and focusing, were compared with the measured results and a reasonably good agreement was found. Furthermore, the primary characteristics in the pressure wave forms produced by different reflector geometries, such as that produced by a reflector insert, can also be predicted by this model. It is interesting to note that when the interpulse delay time calculated by linear geometric model is less than about 1.5 μs, two pulses from the reflector insert and the uncovered bottom of the original HM-3 reflector will merge together. Coupling the simulated pressure wave form with the Gilmore model was carried out to evaluate the effect of reflector geometry on resultant bubble dynamics in a lithotripter field. Altogether, the equivalent reflector model was found to provide a useful tool for the prediction of pressure wave form generated in a lithotripter field. This model may be used to guide the design optimization of reflector geometries for improving the performance and safety of clinical lithotripters. [DOI: 10.1121/1.2195074]

PACS number(s): 43.25.Cb, 43.25.Yw [CCC] Pages: 3625–3636

I. INTRODUCTION

Since its invention in the early 1980s, shock wave lithotripsy (SWL) has revolutionized the treatment for renal and upper urinary stones.[1] Currently, approximately 75% of stone patients are treated by SWL alone and another 20% by SWL in conjunction with endoscopic procedures.[2] In spite of this great success and the fact that several new generations of lithotripters have been introduced for clinical use, no fundamental improvements in SWL technology that can lead to better treatment efficiency with reduced tissue injury have been accomplished in the past two decades.[2–4] The first-generation Dornier HM-3 lithotripter is still regarded by most urologists in the US as the gold standard of SWL.[3]

Both clinical and basic studies have shown that SWL can produce acute renal injury, such as hematuria, kidney enlargement, renal and perirenal hemorrhage and hematomas, especially with the much higher pressure output of the third-generation lithotripters.[5–9] Following SWL, although most young adult patients recover well, a subgroup of patients, such as pediatric and elderly patients, and patients with pre-existing renal function impairment, are much more susceptible to SWL-induced chronic injury.[6] Efforts have been under way to improve the design of lithotripter shock wave sources to improve stone comminution while reducing concomitantly the collateral renal injury.[4,10–13]

Theoretical modeling of the propagation and focusing of lithotripter shock wave (LSW), and bubble dynamics in a lithotripter field is important for understanding the working mechanisms of SWL and for design optimization of the lithotripters. Attempts of modeling the acoustic field of an electrohydraulic shock wave lithotripter have been carried out by several groups. Hamilton developed a linear model to simulate the pressure wave form along the lithotripter axis, which provides useful insight on the evolution of different LSW components during wave converging toward the lithotripter focus.[14] Coleman and colleagues used the Burger's equation, which is a one-dimensional version of the Khokhlov–Zabolotskaya–Kuznestsov (KZK) equation to model the focused LSW and diffractive effects were considered by assuming the sound field as a Gaussian beam.[15] Christopher developed a nonlinear propagation model to account for wave diffraction, absorption, dispersion, and nonlinearity, though a limit on peak negative pressure needs to be set in the model calculation.[16] Averkiou and Cleveland used geometrical acoustics to describe the wave propagation within a reflector bowl, and subsequently the calculated pressure distribution at the reflector aperture was fed into a two-dimensional KZK model as source condition to simulate the propagation and focusing of LSWs.[17] Colonius and Tanguay

[a]Electronic mail: yufeng.zhou@duke.edu

developed a model of shock wave propagation using an ensemble averaged two-phase flow model for the bubbly mixture combined with a high-order accurate shock capturing technique.[18,19] Szeri and colleagues modeled the reflector as an interface with a density jump across it, and used a finite-volume scheme to solve the reflection and steepening of a pressure wave based on the Euler equations coupled with the Tait equation of state.[20,21] Although these previous works have led to a better understanding of the shock wave propagation and evolution in a lithotripter field, there are still some discrepancies between the measurements and the simulated results, such as in the wave form profile, pressure distribution, and the evolution of different wave components. In addition, there are significant discrepancies in the assumed pressure wave form generated by the spark discharge of the HM-3 lithotripter in the previous model calculations.[16–19]

In this study, the shock wave emanated from the spark discharge between the electrode tips in an HM-3 lithotripter was measured by a fiber optic probe hydrophone (FOPH) and used as the source condition for the model calculation. An equivalent reflector model was developed and coupled with the KZK equation to simulate the acoustic field of an HM-3 lithotripter using either the original or an upgraded reflector. The simulated results were compared with the experimental measurements and also with the results calculated using the Averkiou and Cleveland model (referred to in this paper as the AC model) both along and transverse to the lithotripter axis. Furthermore, the effects of the reflector insert geometry on LSW profile and bubble dynamics were evaluated.

II. EXPERIMENTAL MATERIALS AND METHODS

A. Lithotripter

The experiments were carried out in a Dornier HM-3 lithotripter with a 80 nF capacitor and a truncated brass ellipsoidal reflector with a semi-major axis $a=138$ mm, a semi-minor axis $b=77.5$ mm, and half-focal length $c=114$ mm. A thin shell ellipsoidal reflector insert, developed in our previous studies,[4,13,22] was also used. The outer surface of the insert matches with the original HM-3 reflector and the inner surface fits an ellipsoidal surface with $a'=132.5$ mm, $b'=71.5$ mm, and $c'=111.5$ mm. The lower edge of the reflector insert was 4 mm above the focal plane at F_1 where the tips of the electrode are centered.[4,13]

B. Pressure field mapping

The pressure wave forms produced by the HM-3 lithotripter with either the original or the upgraded reflectors were measured using a FOPH-500 (RP Acoustics, Leutenbach, Germany). The sensing probe of the hydrophone, a 100 μm optical fiber, was placed inside a holder and attached to a three-dimensional position system (step motors: VX M-2, lead screws: BiSlide-M02, Velmex, Bloomfield, NY), which has a minimum step size of 5 μm and a maximum scanning range of 250 mm. The hydrophone was tilted at 14° so that the probe could be aligned parallel to the lithotripter axis (Fig. 1). A mechanical pointer was used to aid the alignment of the probe tip with F_2. A program was used to control the automatic scan of the probe in the lithotripter field at a minimum step size of 1 mm. At each location, at least six pressure wave forms were recorded by a digital oscilloscope (LeCroy 9314M, Chestnut Ridge, NY) operated at 100 MHz sampling rate and the data were subsequently transferred to a PC for offline analysis.

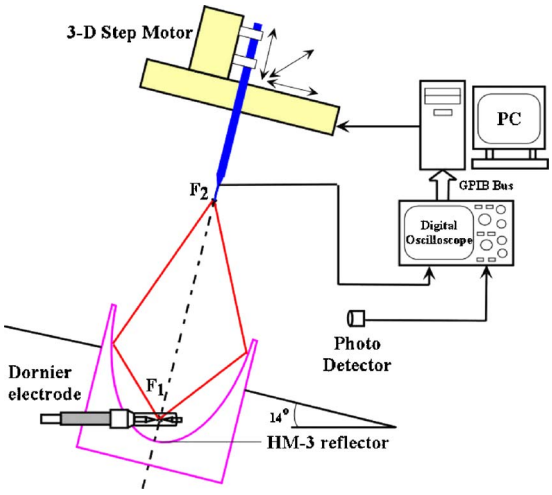


FIG. 1. The experimental setup for pressure mapping in a Dornier HM-3 lithotripter by fiber optical probe hydrophone.

C. Equivalent reflector model

The ellipsoidal reflector of an HM-3 lithotripter can be considered to be equivalent to a spherical reflector with a focal length of $2a$ and with its focus at F_2 (Fig. 2). In this model shock waves will emanate from the surface of the equivalent reflector, instead of bouncing off from the ellip-

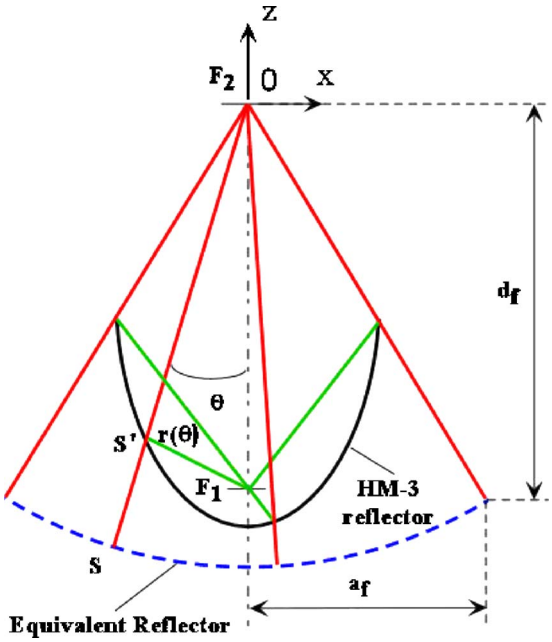


FIG. 2. The schematic diagram of the equivalent reflector model. The solid line shows the ellipsoidal reflector of an HM-3 lithotripter, and the dashed line shows the equivalent reflector of it.

soidal reflector before focusing toward F_2. The distance between a point on the ellipsoidal surface and the electrode tip (F_1) is

$$r(\theta) = 2a - \frac{a(1-\varepsilon^2)}{1+\varepsilon\cos\theta}, \tag{1}$$

where ε is the eccentricity of the ellipse, θ is the incident angle of an acoustic ray with respect to the lithotripter axis towards F_2. In addition, the spark discharge at F_1 is assumed to be a spherical source, which satisfies

$$|p(r)| = \frac{a-c}{r(\theta)} \cdot p_0, \tag{2}$$

where p_0 is the sound pressure at a distance of a-c from F_1, r is the traveling distance. Combining Eqs. (1) and (2) the pressure amplitude at the point S' is equal to

$$|p'(\theta)| = \frac{(1-\varepsilon)(1-\varepsilon\cdot\cos\theta)}{1-2\varepsilon\cdot\cos\theta+\varepsilon^2} \cdot p_0. \tag{3}$$

Then the pressure amplitude at the point S can be estimated as $|p(\theta)|=|p'(\theta)|\cdot[2a-r(\theta)]/2a$ by using the ray theory. Therefore, pressure wave form at the aperture of the equivalent reflector can now be expressed as

$$p(\theta) = \frac{(1-\varepsilon)^2(1+\varepsilon)}{2(1-2\varepsilon\cdot\cos\theta+\varepsilon^2)} \cdot p_0 \cdot f(\tau+G\rho^2)\cdot h(1-\rho), \tag{4}$$

where $f(^*)$ is the source wave form of electrohydraulic spark discharge, $\rho=x/a_f$ and $\tau=\omega_0 t'=\omega_0(t-z/c_0)$ are normalized variables, x and z is the coordinate transverse to and along the axis of the sound beam, respectively, t is the propagation time from F_1, a_f is the aperture radius, c_0 is the sound speed of small amplitude signal in water, ω_0 is a characteristic angular frequency, G is the focusing gain, and $h(^*)$ is the step unit function.

The pressure wave form in the lithotripter field is simulated by using an axisymmetric form of the KZK equation solved in dimensionless form:[17,23,24]

$$\frac{\partial P}{\partial\sigma} = \frac{1}{4G}\int_{-\infty}^{\tau}\left(\frac{\partial^2 P}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial P}{\partial\rho}\right)d\tau' + A\frac{\partial^2 P}{\partial\tau^2} + \frac{NP}{(1+\sigma)}\frac{\partial P}{\partial\tau}, \tag{5}$$

where $P=p/p_0$, $\sigma=z/d_f$, $\tau=\omega_0 t'$, and d_f is the focal length. The three items on the right of Eq. (5) account for the effect of diffraction, absorption, and nonlinearity phenomena on the progressive propagation of the LSW. The dimensionless parameters that describe the relative importance of those phenomena are

$$G = z_0/d_f, \quad A = \alpha_0 d_f, \quad N = d_f/\bar{z}, \tag{6}$$

respectively, where $z_0=\omega_0 a_f^2/2c_0$ is the Rayleigh distance, α_0 is the thermo-viscous attenuation coefficient, $\bar{z}=\rho_0 c_0^3/\beta\omega_0 p_0$ is the plane wave shock formation distance, ρ_0 is the ambient density of the fluid, β is the nonlinearity coefficient. The corresponding parameters for water used in the numerical code are α_0 (f=0.25 MHz)=0.0016 Np/m, β=3.5, ρ_0 =1000 kg/m^3, c_0=1500 m/s.[17] In calculating the lithotripter field produced by the upgraded reflector, the initial conditions for the KZK equation have two parts, one from the inner surface of the reflector insert ($0°\leqslant\theta\leqslant 11.8°$) and the other from the uncovered bottom surface of the original HM-3 reflector ($11.8°<\theta\leqslant 32.5°$).

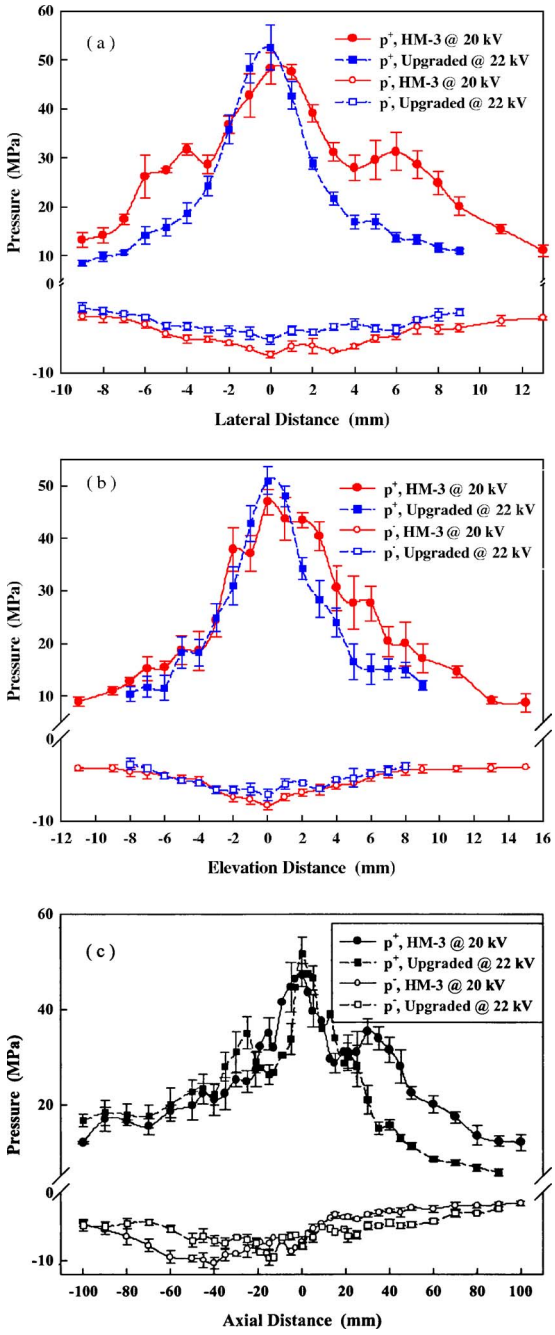


FIG. 3. Distribution of the peak pressure of the lithotripter shock waves in the (a) foot-head, (b) left-right, and (c) axial direction generated by using the original HM-3 reflector at 20 kV and the upgraded reflector at 22 kV.

The normalized KZK equation was solved numerically in the time domain using finite-difference operators to approximate the derivatives. The diffraction, absorption, and nonlinearity algorithms were implemented independently over each step $\Delta\sigma$ because the order in which these three effects were included is not important. In the near field region, a fully implicit backward finite difference method was used to effectively reduce numerical oscillations. Beyond the

TABLE I. Comparison of −6 dB beam size of lithotripter field produced by the original HM-3 reflector at 20 kV and the upgraded reflector at 22 kV.

	HM-3 reflector at 20 kV		Upgraded reflector at 22 kV	
Reflector configuration	p^+	p^-	p^+	p^-
Focal width, head-foot (mm)	14.3	18	5.9	16.2
Focal width, left-right (mm)	9.6	14.8	6.6	15
Focal extent (mm)	81.4	118.6	62.1	90

oscillatory region, a Crank-Nicolson finite difference method with a second-order approximation in $\Delta\sigma$ was applied which permits larger step sizes.[17,23] It was found that the r.m.s. error of the algorithm is proportional to $(\Delta\sigma)^{1.3}$, and the expected convergence rate is almost proportional to $(\Delta\sigma)^{1.0}$.[23] The computation time for simulating a full lithotripter field with the equivalent reflector in a P4 1.0 GHz 256 MByte memory PC is about 10 h.

D. Modeling of bubble dynamics in a lithotripter field

The Gilmore model was applied to estimate the radial oscillation of a single spherical bubble with consideration of the compressibility of the liquid.[25]

$$R\left(1-\frac{U}{C}\right)\frac{dU}{dt}+\frac{3}{2}\left(1-\frac{U}{3C}\right)U^2 = \left(1+\frac{U}{C}\right)H+\frac{1}{C}\left(1-\frac{U}{C}\right)R\frac{dH}{dt}, \quad (7)$$

where R is the bubble radius, and U $(=dR/dt)$ is the velocity of the bubble wall, C is the speed of sound in the liquid at the bubble wall, and H is the enthalpy difference between the liquid at pressure $P(R)$ and pressure P_∞. The gas diffusion during bubble oscillation is not considered since we are most interested in the expansion phase of the bubble dynamics as it is related to the propensity of tissue injury.[22] The Gilmore model was solved by using the fifth-order Runge–Kutta–Fehlberg method with a step-size control algorithm.[26]

III. RESULTS

A. Pressure field and distribution

The distributions of the peak positive (p^+) and peak negative (p^-) pressure, both along and transverse to the lithotripter axis at F_2, are shown in Fig. 3 and the −6 dB beam sizes are listed in Table I. It is found that the beam sizes of p^- are larger than the corresponding values of p^+ because the tensile component of LSW has much lower frequency (~125 kHz) than the compressive wave (~0.5 MHz); here the frequency of LSW component is defined as the half reciprocal of the wave component duration time. The pressure distribution in the focal plane of the original HM-3 reflector is not axisymmetric. Larger beam width was found in the head-foot direction than in the left-right direction [Figs. 3(a) and 3(b)], which is presumably caused by the lateral truncation of the original HM-3 reflector to accommodate the bi-planar fluoroscopic imaging system for stone localization. In comparison, the pressure distribution of the upgraded reflector was found to be more axisymmetric and with smaller beam size.

B. Source condition

The pressure wave form produced by the spark discharge of the HM-3 lithotripter at 20 kV is shown in Fig. 4(a). The wave form can be fitted in an exponential decay curve [dashed line in Fig. 4(a)], which is characteristic of an underwater explosion:[27]

$$p(t)=A\cdot\exp\left[-\frac{t-t_0}{T}\right]\cdot h(t-t_0), \quad (8)$$

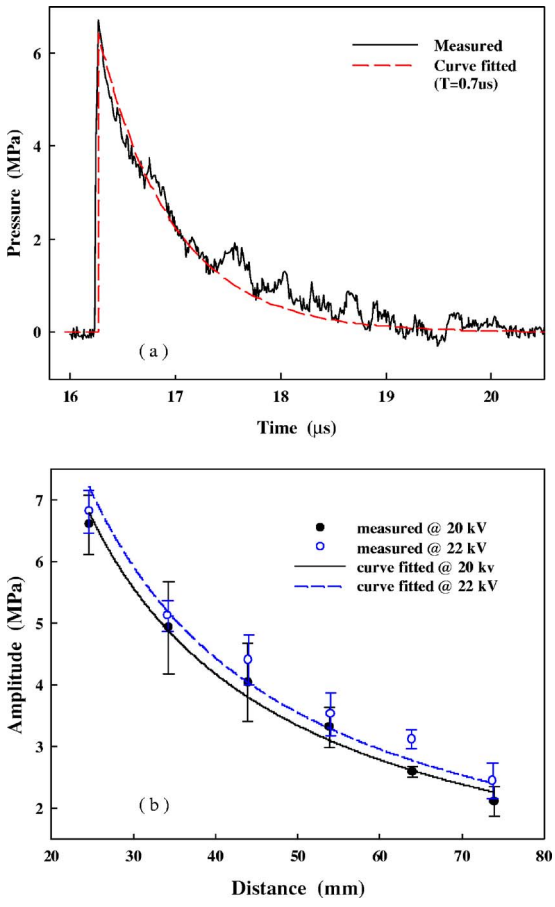


FIG. 4. (a) The pressure wave form produced by the electrode spark discharge of HM-3 lithotripter at 20 kV at about 24 mm away from F_1. The dashed line is the fitted curve in an exponential decay format and (b) the pressure amplitude is found to be inversely proportional to the propagation distance at both 20 and 22 kV.

where A is the peak positive pressure, T is a time constant, t_0 is the retarded time. Furthermore, the pressure amplitude was found to be inversely proportional to the traveling distance from F_1, $c_0 t_0$ [Fig. 4(b)]. For the equivalent model calculation, the source pressure p_0 was estimated to be 6.8 MPa at 20 kV and 7.3 MPa at 22 kV, respectively.

C. Simulated pressure wave forms

1. Comparison between the equivalent reflector model and the AC model

Figures 5 and 6 show the simulated LSW profiles both transverse to (Fig. 5) and along the lithotripter axis (Fig. 6), using the equivalent reflector model (solid line in the right column) and the AC model (the dashed line in the right column), respectively, together with the pressure wave forms measured experimentally (the left column). Because shock wave propagation and focusing are considered inside the reflector bowl in the equivalent reflector model, the simulated LSW always travels in advance of the corresponding LSW predicted by the AC model. In addition, the time delay between the arrival of the LSW and the estimated value using small-signal acoustic speed in water increases with the propagation distance. At $z=-70$ and 0 mm the LSW arrives at a retarded time of -0.08 and -0.58 μs based on the AC model and -0.43 and -1.12 μs using the equivalent reflector model, respectively. There are also significant differences in the profile of the LSWs predicted by the two models. First, at F_2 [Fig. 5(a)] the tensile component of the LSW predicted by the AC model has a prolonged duration near p^- before quickly returning to the equilibrium pressure. In comparison, the pressure wave form predicted by the equivalent reflector model reaches p^- earlier and then gradually returns to zero, which resembles more closely the measured pressure wave forms[28–30] although the tensile duration is slightly shorter than the measured wave form. Second, using the AC model the rise time of the compressive wave increases substantially from 31 ns at $x=0$ to 1.04 μs at $x=6$ mm and the wave form becomes almost sinusoidal at $x=9$ mm. In comparison, at $x=6$ mm the simulated wave form from the equivalent reflector model has a rise time of 0.2 μs, and overall, resembles more closely the measured pressure wave forms (Fig. 5). Furthermore, the equivalent reflector model clearly predicts the separation of a single peak in the LSW profile into two distinct peaks along the transverse direction of the lithotripter field. In comparison, the peak separation is not clearly predicted using the AC model [Fig. 5(d)]. Third, beyond the focal point the diffracted wave originated from the aperture edge of the reflector will overtake the center wave of the LSW.[14] The general characteristics of the edge wave evolution are well predicted by both models and are in good agreement with the linear model prediction, i.e., the edge wave moves toward the central wave and merges with the wake when propagating toward the focal point, and subsequently, it overtakes the central wave beyond F_2.[14] However, the advancing speed of the edge wave in the AC model was found to be much faster than that predicted by the equivalent reflector model [Figs. 6(g) and 6(h)]. At $z=70$ mm, the time delay between the edge and central wave is 3.2 μs based on the AC model, compared to 1.5 and 1.45 μs based on the equivalent reflector model and pressure wave form measurement [Fig. 6(h)].

It is interesting to note that the pressure distributions predicted by the two models are quite similar, although the locations of the maximum and minimum pressures are different (Fig. 7). Using the AC model, the maximum and minimum pressures occur at 4 and -12 mm, respectively, while the corresponding values are 12 and -7 mm for the equivalent reflector model. There is a discrepancy of the location of maximum peak negative pressure between the measurement and the simulation, and the reason is unknown. When the new reflector insert was used, this discrepancy became much smaller [Fig. 3(c)].

2. Comparison between the model simulation and measured pressure wave forms produced by the original and upgraded HM-3 reflectors

Although in general the simulated wave forms predicted by the equivalent reflector model show a good agreement with the measurement results, especially with respect to the characteristics of the evolution of wave components, there are also some noticeable discrepancies. First, the double-peak structure in the leading compressive wave due to the lateral truncation on the original HM-3 reflector aperture is not predicted because of axisymmetric assumption in the source condition. Second, the -6 dB beam width predicted by the theoretical models (5.2 and 5.4 mm by the AC and the equivalent reflector model, respectively, see Fig. 7) is much smaller than the measured results (see Table I). Hence, away from F_2 the peak pressure of the simulated wave form is always lower than the corresponding value in the measured wave form (Figs. 5 and 6). Third, multiple local maxima of p^+ along the lithotripter axis [such as between $z=-40$ and 40 mm in Fig. 3(c)], similar to the pressure distribution in a high-intensity focused ultrasound field,[31] were not predicted by the numerical models. The reason for this discrepancy is not clear.

The equivalent reflector model was also used to predict the pressure wave forms generated by the upgraded reflector both transverse to (Fig. 8) and along (Fig. 9) the lithotripter axis, and the numerical simulations are compared with the measurements. It was found that the time delay (Δt) between the leading LSW (produced by wave reflection from the insert) and the trailing second shock wave (produced by wave reflection from the uncovered bottom surface of the original reflector) increases as these two waves propagate towards the focal point. At $z=-70$ mm, Δt is 2.04 μs while at $z=0$ mm, it increases to 3.26 μs, which is close to the measurement result of 3.42 ± 0.15 μs. In addition, Δt also increases transversely from F_2. At $x=9$ mm, the simulated Δt is 3.98 μs, compared to 4.07 μs from the measurement [see Fig. 8(d)]. At this location the leading LSW splits into two peaks as shown previously for the original reflector [Fig. 5(d)]. At $z=25$ mm the edge wave begins to separate from the central wave of the leading LSW, and subsequently the rise time of shock front becomes larger [Fig. 9(f)]. However, because the diffraction wave was generated much earlier from the uncovered bottom of original HM-3 reflector, the second shock

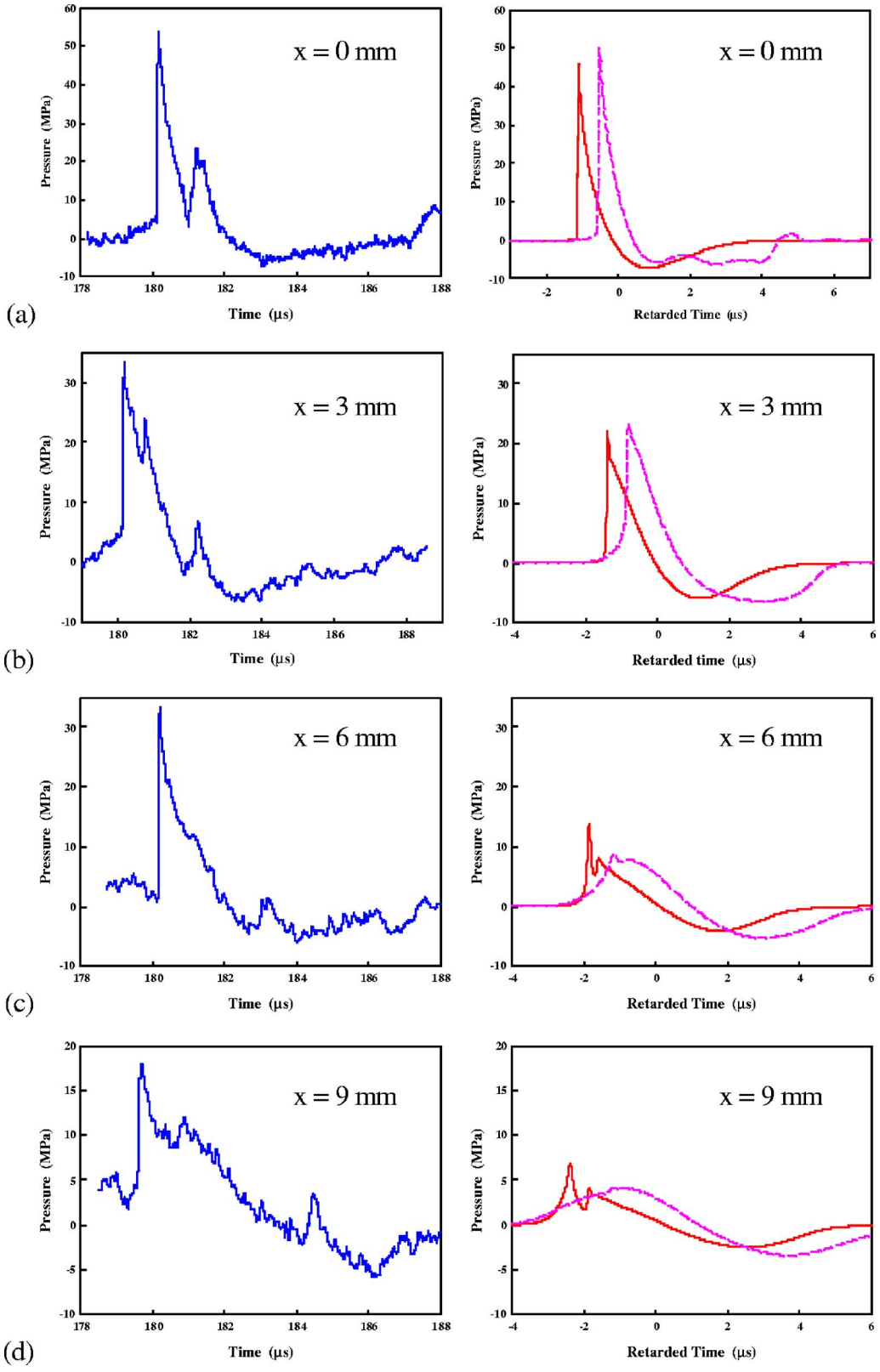


FIG. 5. The pressure wave forms of Dornier HM-3 lithotripter at 20 kV transverse to lithotripter axis in the focal plane at (a) x=0 mm, (b) x=3 mm, (c) x=6 mm, and (d) x=9 mm. Left column: Measured results by fiber optic probe hydrophone, Right column: Simulated results of the equivalent reflector model (solid line) and the AC model (dashed line).

wave maintains a steep shock front even at a long distance from F_2 [see Fig. 9(f)–9(h)]. Overall, the characteristics of different wave components and their evolutions in the lithotripter field could be clearly predicted by the equivalent reflector model and were found to match well with the measured pressure wave forms. Despite these similarities, a clear difference can be observed in the shape of the second pulse. In a large area within the focal region, the second pulse in

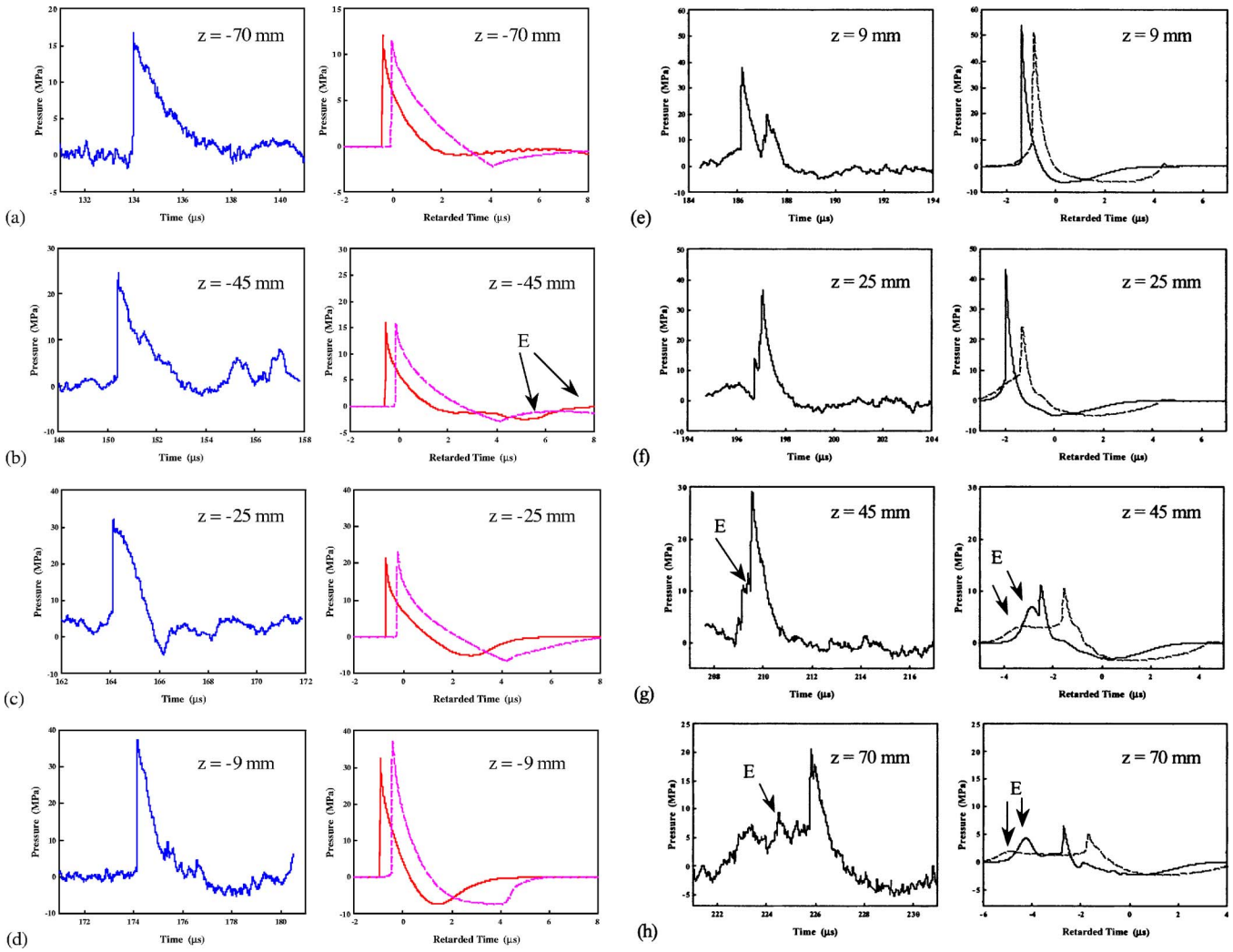


FIG. 6. The pressure wave forms of Dornier HM-3 at 20 kV along the lithotripter axis at (a) $z=-70$ mm, (b) $z=-45$ mm, (c) $z=-25$ mm, (d) $z=-9$ mm, (e) $z=9$ mm, (f) $z=25$ mm, (g) $z=45$ mm, and (h) $z=70$ mm. Left column: Measured results by FOPH. Right column: simulated results of the equivalent reflector model (solid line) and the AC model (dash line). E: the edge wave.

the measured wave form looks more like a triangle wave rather than an exponentially decaying curve for the first pulse.

As shown in Fig. 10, the −6 dB beam size of the second shock wave (244×9.6 mm) was found to be much larger than that of the leading LSW (59×5.6 mm). Hence, at locations either far beyond ($z>40$ mm) or away ($x>10$ mm) from F_2, the amplitude of the second shock wave could be comparable or even larger than that of the leading LSW, which is in agreement with the measurements [Fig. 8(d), Figs. 9(g) and 9(h)]. The broad coverage of the second shock wave ensures the effectiveness of suppressing intraluminal bubble expansion for reduced tissue injury in the whole lithotripter field.[4,13]

D. Effect of reflector geometry on bubble dynamics

The equivalent reflector model can be used to calculate the pressure wave forms produced by the upgraded reflectors of different geometries, and then the results can be input to the Gilmore model to evaluate the ensuing bubble dynamics. Using this approach, two series of numerical simulations have been carried out first at F_2 to evaluate its feasibility. One is to change the semi-major axis of the inner reflector surface (a') and thus varying the interpulse delay time ($\Delta t=2[(a-a')-(c-c')]/c_0$).[12] The other is to change the extension distance from the lower edge of the reflector insert to the first focal plane, d_T.[13] In both cases, c' of the reflector insert remained the same. It was found that when Δt is larger than 2 μs, the leading LSW and the second shock wave are distinctively separated from each other, as shown in Fig. 11(a) for $a'=130.25$ mm with $\Delta t=5.2$ μs. However, when Δt is smaller than 2 μs, the profile of the second shock wave will be affected by the leading LSW during the propagation. For example, the second pulse begins to enlarge its duration and flatten its profile at $a'=133.6$ mm and $\Delta t=1.85$ μs [Fig. 11(b)]. The two waves start to merge at $a'=134.25$ mm and $\Delta t=1.66$ μs [Fig. 11(c)]. At $a'=134.38$ mm, the second pulse disappears and merges with the tensile component of the leading LSW, leading to a sudden decrease of the peak negative pressure (i.e., $p^-=-4.9$ MPa). The measured pressure wave form from one of these reflector inserts, $a'=134.5$ mm, has confirmed the merge of the two shock waves experimentally [dotted line in Fig. 11(d)]. At even

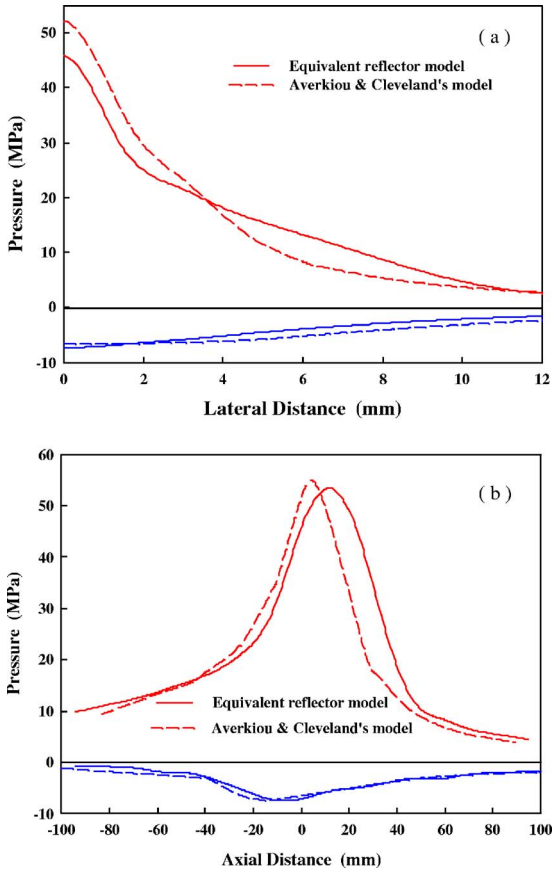


FIG. 7. The simulated peak pressure distributions (a) transverse to and (b) along the lithotripter axis. The solid lines are the results of the equivalent reflector model and the dashed ones are those of the AC model.

shorter Δt, the second shock wave will move even closer to the leading LSW and further enhance the peak positive pressure of the pressure wave form, for example, $p^+=68.4$ MPa for $a'=135.13$ mm. Subsequently, the induced bubble dynamics may be similar to or even stronger than that produced by the original HM-3 reflector, i.e., $R_{\max}=0.615$ and 0.69 mm at 20 and 22 kV, respectively. Furthermore, the strongest suppression effect on bubble expansion was found to occur at $a'=132.8$ mm, which is close to our previous design.[13]

The truncation of the reflector insert begins to affect the maximum bubble expansion when d_T is larger than 1 mm [Fig. 12(b)]. However, if d_T is larger than 4 mm $R_{\max}$ will also increase. With the increase of d_T from −4 to 10 mm the amplitude of the second pulse will increase almost linearly from 10.3 to 15.5 MPa. Meanwhile the first compressive wave is initially consistent within the range of $d_T=-4$ to 4 mm, then decreases from 60.2 to 57.8 at $d_T=10$ mm. However, the variation of the peak negative pressure is much smaller, only 0.2 MPa. In addition, the interpulse delay time appears to decrease from 3.39 to 3.16 μs within the range. At $d_T=-4$ mm p^+ for the leading LSW, second shock wave and Δt are 59.4 MPa, 10.3 MPa, and 3.39 μs, respectively. The corresponding values at $d_T=10$ mm are 57.8 MPa, 15.5 MPa, and 3.16 μs. The minimum $R_{\max}$ occurs at around $d_T=4$ mm, which is similar to the value used in our previous study.[13] Altogether, it is shown that the equivalent reflector model can be a useful tool for design optimization of the HM-3 reflector.

IV. DISCUSSION

In this study, an equivalent reflector model was developed and applied to model the acoustic field of an HM-3 lithotripter with either the original or the upgraded reflector. The simulated wave forms show a good agreement with the measurements by FOPH and illustrate the primary features of the LSW captured by the model, especially the evolution of three wave components (the central wave, the edge wave, and the wake) in the lithotripter field. The delay time between the edge wave and the central wave simulated by the equivalent reflector model matches better with the measured value than that predicted by the AC model. In addition, the LSW simulated by the equivalent reflector model has a longer retarded time. These differences and improvements are due to the consideration of shock wave propagation inside the ellipsoidal reflector. In particular, the equivalent spherical reflector is $2a/(c+14\ \text{mm})=2.16$ times as large as that used in the AC model at the aperture of the HM-3 reflector. In the lithotripter field, such as $z=50$ mm and $x=10$ mm, the normalized σ and ρ in the AC model [Eq. (4)] are 1.39 and 0.129, respectively. In comparison, the corresponding values in the equivalent reflector model are 1.21 and 0.07, respectively, which means that the locations are relatively closer to the lithotripter focus. Therefore, the predicted edge wave is closer to the central wave and the shock front could be kept even far away from the focal point using the equivalent reflector model.

The double-pulse structure from the upgraded reflector can also be predicted in the simulation. It is interesting to find that when Δt is smaller than 1.5 μs, the two shock waves will merge together, which has been confirmed by pressure measurements. The fundamental mechanism of acoustic wave propagation is the compression and extension of the elastic media to transfer the acoustic energy. It is also noticed that the interpulse delay time increases when propagating toward the focal point of lithotripter (Fig. 9). So if two pulses with finite amplitudes are initially close to each other, the wave-wave interaction may cause the media to vibrate in phase, rather than acting as two independent sources. Subsequently, they may not separate even at F_2.[32] Altogether, it is found that the *in situ* pulse superposition technique is not simply adding two individual wave forms together as predicted by the linear model.

Although the equivalent reflector model can be used successfully to simulate the acoustic wave form in the lithotripter field, there are still some limitations in the numerical solution. First, the KZK equation is only valid for a moderately focusing gain (less than 20).[17] Although a simplified method of introducing a coordinate transformation in the frequency domain[33] was used in our calculation to increase the upper limit of the focusing gain, some artificial oscillations and additional peaks were still found far away from the focal point [Figs. 5(d), 6(g), 6(h), 8(c), 8(d), 9(g), and 9(h)]. Second, the KZK equation is actually a 1.5 D model, assuming

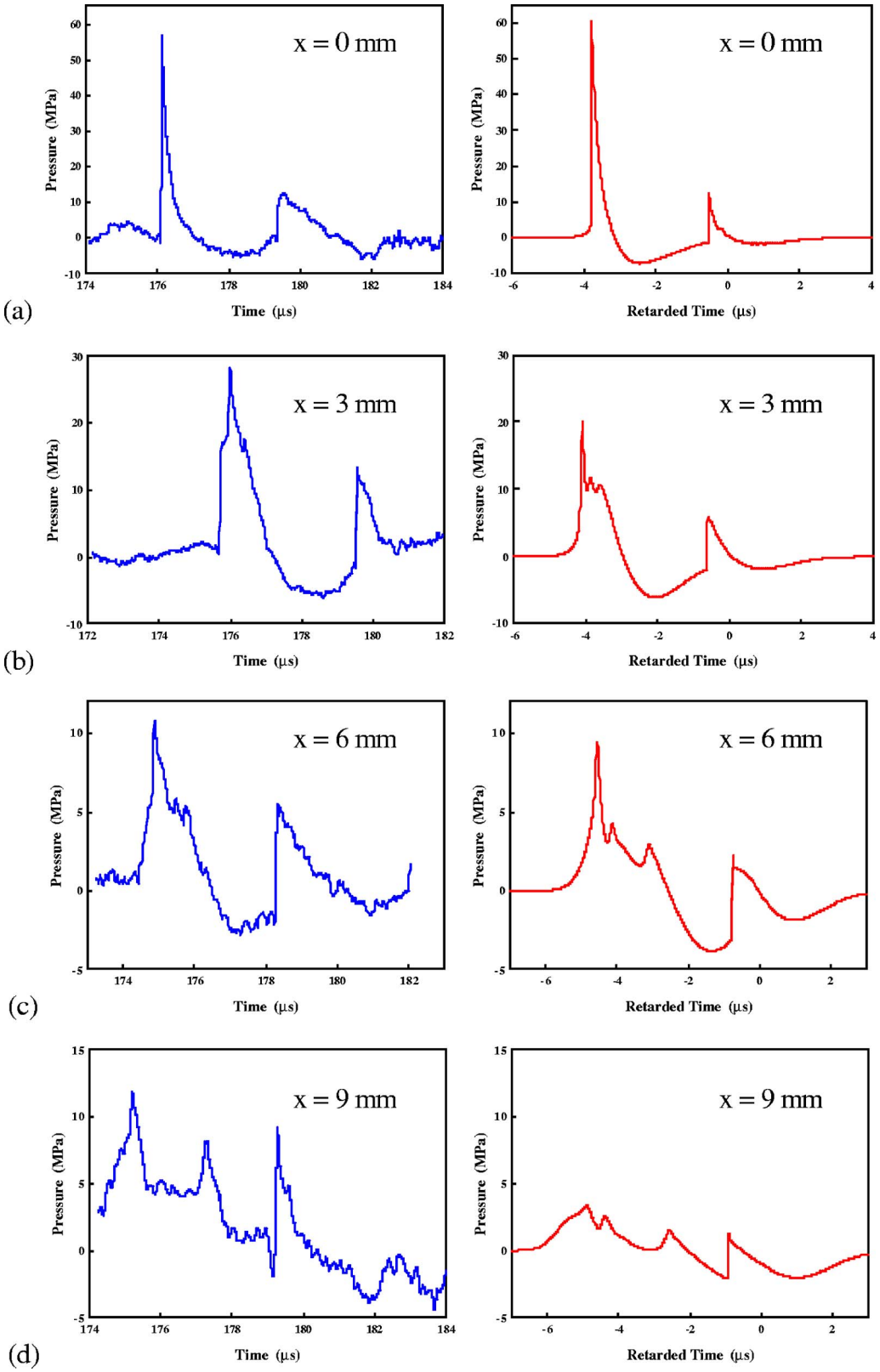


FIG. 8. The pressure wave forms of the upgraded reflector at 22 kV transverse to the lithotripter axis in the focal plane at (a) x=0 mm, (b) x=3 mm, (c) x=6 mm, and (d) x=9 mm. Left column: Measured results by FOPH. Right column: Simulated results of the equivalent reflector model.

axis-symmetric source distribution for initial conditions. Therefore, more complicated reflector structure, such as the lateral truncation of the HM-3 reflector, cannot be considered. Currently three-dimensional modeling in computational fluid dynamics is becoming a valuable method for gaining deeper insight into dynamics of lithotripters, such as using a wave propagation approach and finite volume method for hyperbolic partial differential equations[20] and using a Weighted Essentially Non-Oscillatory fifth-order shock capturing scheme for accurate representation of shock wave

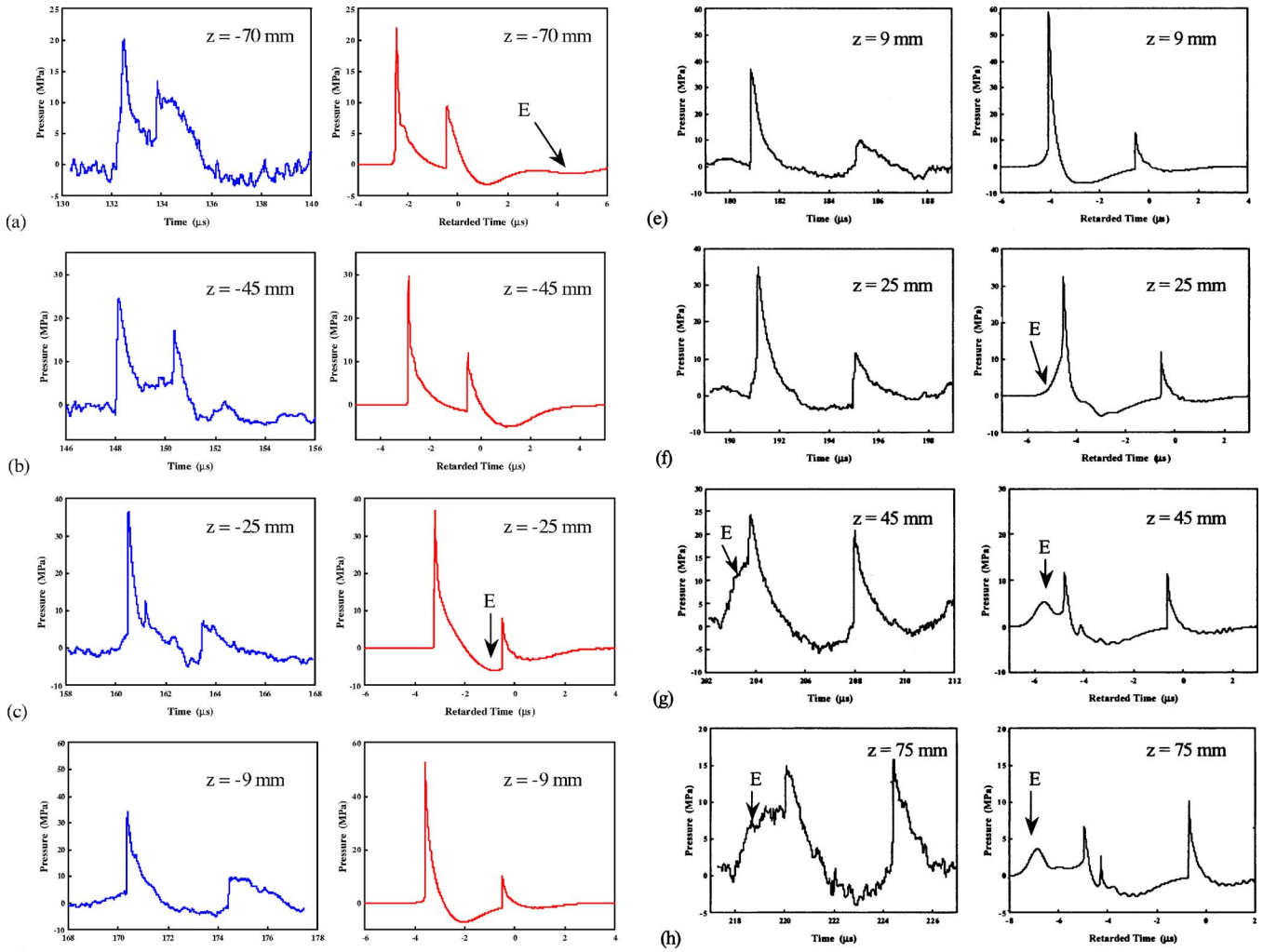


FIG. 9. The pressure wave forms of the upgraded reflector at 22 kV along the lithotripter axis at (a) $z=-70$ mm, (b) $z=-45$ mm, (c) $z=-25$ mm, (d) $z=-9$ mm, (e) $z=9$ mm, (f) $z=25$ mm, (g) $z=45$ mm, and (h) $z=70$ mm. Left column: Measured results by FOPH. Right column: Simulated results of the equivalent reflector model. E: the edge wave.

propagation.[18,19] Even assuming reflector with two-dimensional symmetry and running on terascale machine, the computation time is still very long.[21] Overall, further development and improvements in the theoretical model and numerical simulation are still needed.

The measured pressure field of the HM-3 lithotripter revealed a much larger −6 dB beam size than the theoretical predictions. Truncation on the lateral side of the HM-3 reflector leads to asymmetric pressure distribution, enlarged beam width, and appearance of distinct side lobes. Using a new single-unit reflector insert, the −6 dB beam size becomes smaller and pressure distribution in the focal plane appears axisymmetric. In addition, because of the spark-induced erosion, gap between the electrode tips widens with increased number of shocks delivered. Consequently the ensuing spark discharges may not always occur exactly at F_1. Using ray tracing method, it has been shown that the departure of only 1 mm laterally from F_1 could cause a significant shift of ~10 mm at F_2.[34] The electromagnetic lithotripters, which have similar F number as HM-3 but much more stable shock wave source, have much smaller beam sizes than HM-3.[28] Therefore, the unstable location of the spark discharge may be another important contributing factor for the larger beam size of the HM-3 lithotripter.

Coupling the pressure wave form predicted by the equivalent reflector model with the Gilmore model can be used to estimate the corresponding bubble dynamics. The reflector geometry changes the peak pressure, interpulse delay time, and the subsequent bubble dynamics, simultaneously. The bubble dynamics are sensitive to the pressure wave form. In this preliminary study, the optimal a' and d_T were found in a small range ($a'=132-133.2$ mm and $d_T=3-5$ mm) for achieving the maximum suppression effect on bubble expansion at F_2. Although there are limitations in this feasibility study, such as a' and d_T were evaluated independently and no relaxation phenomenon was included in the wave propagation model,[24] the equivalent reflector model was found to be a useful tool for predicting the LSW wave form generated by different reflector geometries. It should also be noticed that in this preliminary study only the bubble dynamics at F_2 in response to the LSW produced by different geometries of reflector insert were investigated. In clinical lithotripsy, shock wave-induced renal injury occurs in an ex-

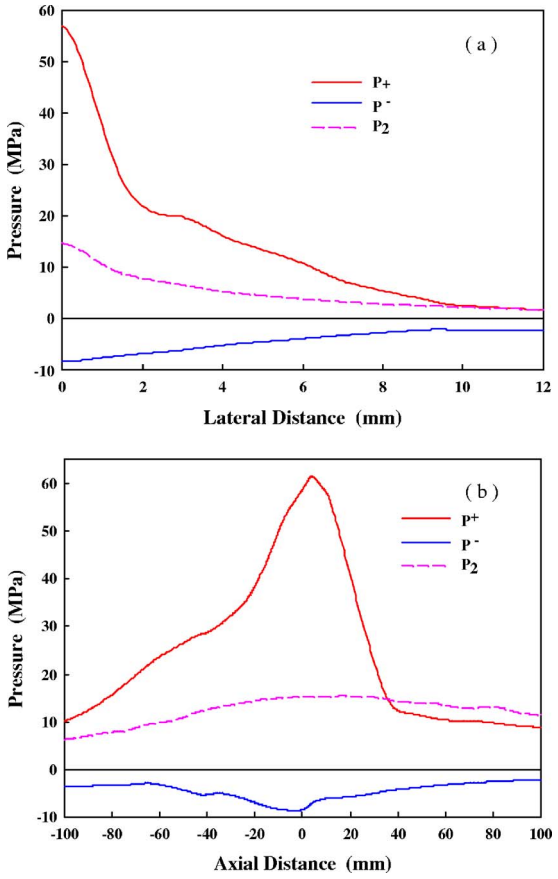


FIG. 10. The simulated peak pressure distributions (a) transverse to and (b) along the lithotripter axis by the equivalent reflector model.

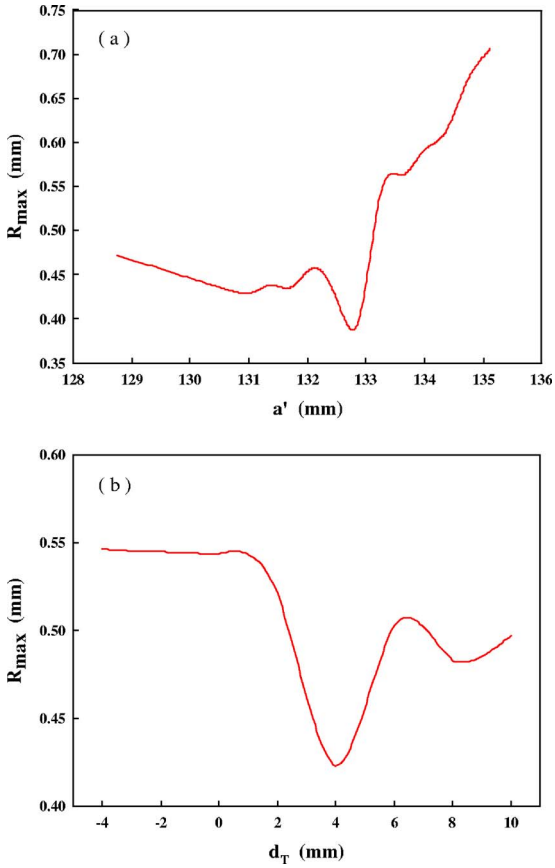


FIG. 12. The maximum bubble size, R_{max}, in accordance with the simulated pressure wave form with varying (a) semi-major axis a' ($c'=111.5$ mm and the lower edge of the reflector insert is 4 mm above the first focal plane) and (b) the extension of the lower edge of the reflector insert to the first focal plane ($a'=132.5$ mm and $c'=111.5$ mm) calculated by the Gilmore model.

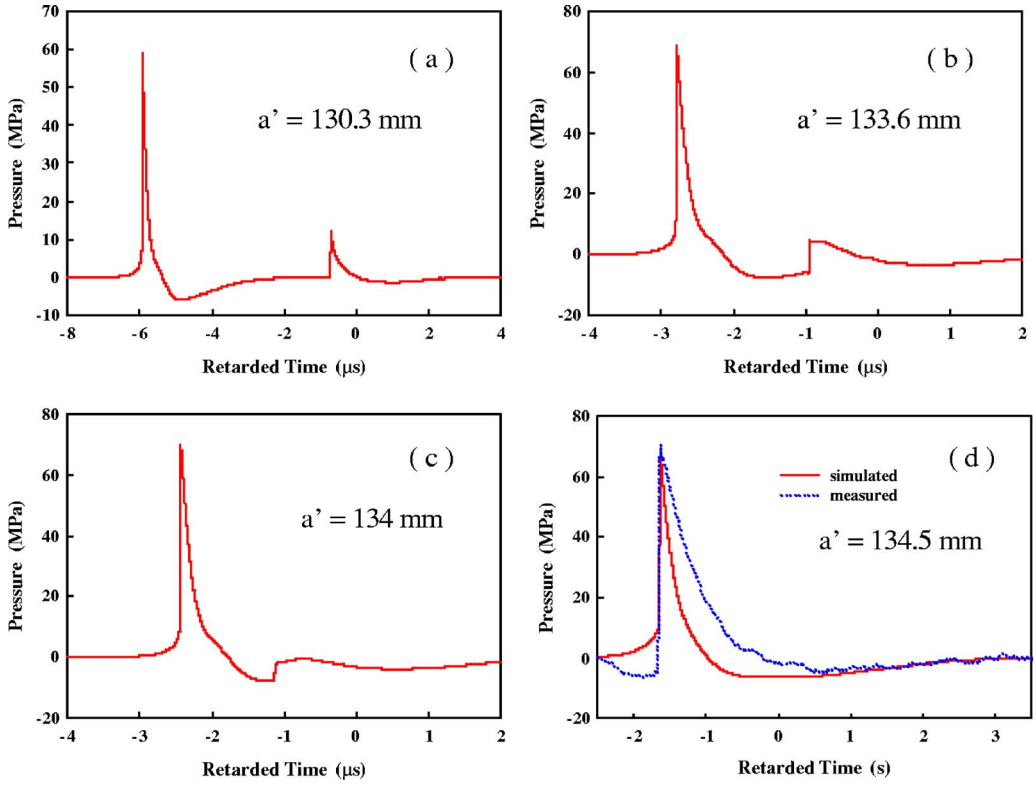


FIG. 11. The simulated wave forms at F_2 produced by different geometries of reflector insert with half focal length $c'=111.5$ mm and semi-major axis, a', of (a) 130.3, (b) 133.6, (c) 134, and (d) 134.5 mm. A measured wave form is shown as dotted curve.

tended volume surrounding the focal point.[5] To fully optimize the reflector geometry, the bubble dynamics in the whole lithotripter field needs to be evaluated. Although theoretically feasible, such a numerical study is beyond the scope of current study. Overall, using numerical models such as the one described in this study, better design of the reflector geometry may be determined expeditiously to improve the performance and safety of clinical lithotripters.

ACKNOWLEDGMENTS

This work is supported in part by NIH through Grant No. RO1-DK52985. The authors are grateful to Dr. Robin O. Cleveland for discussion of the KZK model and algorithm. The authors also acknowledge Dr. Andrew Szeri and Jonathan Iloreta for reading the manuscript and providing valuable comments.

[1]C. Chaussy and G. J. Fuchs, "Current state and future developments of noninvasive treatment of human urinary stones with extracorporeal shock wave lithotripsy," J. Urol. (Baltimore) **141**, 782–792 (1989).
[2]K. Kerbl, J. Rehman, J. Landman, D. Lee, C. Sundaram, and R. V. Clayman, "Current management of urolithiasis: Progress or regress?," J. Endourol **16**, 281–288 (2002).
[3]J. E. Lingman, "Extracorporeal shock wave lithotripsy—Development, instrumentation, and current status," Urol. Clin. North Am. **24**, 185–211 (1997).
[4]Y. F. Zhou, F. H. Cocks, G. M. Preminger, and P. Zhong, "Innovations in shock wave lithotripsy technology: Updates in experimental studies," J. Urol. (Baltimore) **172**, 1892–1898 (2004).
[5]M. Delius, "Medical applications and bioeffects of extracorporeal shock waves," Shock Waves **4**, 55–72 (1994).
[6]A. P. Evan and J. A. McAteer, "Q-effects of shock wave lithotripsy," in *Kidney Stones, Medical and Surgerical Management*, edited by F. L. Coe *et al.* (Lippincott-Raven, Philadephia, 1996), pp. 549–570.
[7]K. Sarica, A. Balat, A. Erbagci, M. Cekmen, M. Yurekli, and F. Yagci, "Effects of shock wave lithotripsy on plasma and urinary levels of nitrite and adrenomedullin," Urol. Res. **31**, 347–351 (2003).
[8]S. F. Graber, H. Danuser, W. W. Hochreiter, and U. E. Studer, "A prospective randomized trial comparing 2 lithotriptors for stone disintegration and induced renal trauma," J. Urol. (Baltimore) **169**, 54–57 (2003).
[9]R. Gerber, U. E. Studer, and H. Danuser, "Is newer always better? A comparative study of 3 lithotriptor generations," J. Urol. (Baltimore) **173**, 2013–2016 (2005).
[10]X. F. Xi and P. Zhong, "Improvement of stone fragmentation during shock wave lithotripsy using a combined EH/PEAA shock wave generator - In vitro experiments," Ultrasound Med. Biol. **26**, 457–467 (2000).
[11]D. L. Sokolov, M. R. Bailey, and L. A. Crum, "Use of dual-pulse lithotripter to generate a localized and intensified cavitation field," J. Acoust. Soc. Am. **110**, 1685–1695 (2001).
[12]P. Zhong and Y. F. Zhou, "Suppression of large intraluminal bubble expansion in shock wave lithotripsy without compromising stone comminution: Methodology and in vitro experiments," J. Acoust. Soc. Am. **110**, 3282–3291 (2001).
[13]Y. F. Zhou and P. Zhong, "Suppression of large intraluminal bubble expansion in shock wave lithotripsy without compromising stone comminution: Refinement of reflector geometry," J. Acoust. Soc. Am. **113**, 586–597 (2003).
[14]M. F. Hamilton, "Transient axial solution for the reflection of a spherical wave from a concave ellipsoidal mirror," J. Acoust. Soc. Am. **106**, 102–112 (1993).
[15]A. J. Coleman, M. J. Choi, and J. E. Saunders, "Theoretical predictions of the acoustic pressure generated by a shock wave lithotripter," Ultrasound Med. Biol. **17**, 245–255 (1991).
[16]T. Christopher, "Modeling the Dornier HM-3 lithotripter," J. Acoust. Soc. Am. **96**, 3088–3095 (1994).
[17]M. A. Averkiou and R. O. Cleveland, "Modeling of an electrohydraulic lithotripter with the KZK equation," J. Acoust. Soc. Am. **106**, 102–112 (1999).
[18]M. Tanguay and T. Colonius, "Numerical simulation of bubble cavitation flow in shock wave lithotripsy," in Fourth International Symposium on Cavitation, CAV2001, Pasadena, 2001.
[19]M. Tanguay and T. Colonius, "Progress in modeling and simulation of shock wave lithotripsy (SWL)," in Fifth International Symposium on Cavitation, CAV2003, Osaka, Japan, 2003.
[20]A. J. Szeri, "Numerical modeling of shock wave focusing and bubble dynamics in an elastic tube" (private communication).
[21]J. I. Iloreta, A. J. Szeri, Y. F. Zhou, and P. Zhong, "Wave propagation and shock formation in a shock wave lithotripter," Phys. Fluids (submitted).
[22]P. Zhong, Y. F. Zhou, and S. L. Zhu, "Dynamics of bubble oscillation in constrained media and mechanisms of vessel rupture in SWL," Ultrasound Med. Biol. **27**, 119–134 (2001).
[23]Y. S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[24]R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, "Time-domain modeling of finite-amplitude sound in relaxing fluids," J. Acoust. Soc. Am. **99**, 3312–3318 (1996).
[25]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[26]S. L. Zhu and P. Zhong, "Shock wave-inertial microbubble interaction: A theoretical study based on the Gilmore formulation for bubble dynamics," J. Acoust. Soc. Am. **106**, 3024–3033 (1999).
[27]R. H. Cole, *Underwater Explosions* (Princeton University Press, Princeton, 1948).
[28]A. J. Coleman and J. E. Saunders, "A survey of the acoustic output of commercial extracorporeal shock wave lithotripters," J. Acoust. Soc. Am. **15**, 213–227 (1989).
[29]R. O. Cle veland, D. A. Lifshitz, B. A. Connors, A. P. Evan, L. R. Willis, and L. A. Crum, "In vivo pressure measurements of lithotripsy shock waves in pigs," Ultrasound Med. Biol. **24**, 293–306 (1998).
[30]M. R. Bailey, D. T. Blackstock, R. O. Cleveland, and L. A. Crum, "Comparison of electrohydraulic lithotripters with rigid and pressure-release ellipsoidal reflectors. II. Cavitation fields," J. Acoust. Soc. Am. **106**, 1149–1160 (1999).
[31]A. C. Baker, "Nonlinear pressure fields due to focused circular apertures," J. Acoust. Soc. Am. **91**, 713–717 (1992).
[32]J. Lighthill, *Waves in fluids.* (Cambridge University Press, Cambridge, 1980).
[33]T. S. Hart and M. F. Hamilton, "Nonlinear effects in focused sound beams," J. Acoust. Soc. Am. **84**, 1488–1496 (1988).
[34]B. Sturtevant, "Shock wave physics of lithotriptors," in *Smith's Textbook of Endourology*, edited by A. D. Smith *et al.*, (Quality Medical, St. Louis, 1996). pp. 529–552.

Resonant solutions in wave phase conjugation induced by a limited magnetoacoustic conjugator

Alain Merlen[a] and Qi Zhang
Joint European Laboratory in Magneto-acoustics (LEMAC), Laboratoire de Mécanique de Lille, UMR CNRS 8107, USTL, Cité Scientifique, bât. M3, 59655 Villeneuve d'Ascq Cédex, France

(Received 13 September 2005; revised 15 March 2006; accepted 27 March 2006)

This paper presents, in one dimension, the general analytical solution of the acoustic phase conjugation in an active medium in contact with passive media of arbitrary impedance. The homogeneous case where no impedance jumps exist at the edge of the active zone is obtained as a particular case. This homogeneous case was the only one treated explicitly in the literature but mostly in the frame of Brillouin scattering. In contrast to this previous work, the present theory is based on a preliminary straightforward analysis using a dual-time-scale method and provides very practical results like the threshold of the supercritical modes, the rate of amplification, and its link with the stress repartition in the conjugator. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2197799]

PACS number(s): 43.25.Dc, 43.25.Lj [MFH] Pages: 3637–3648

I. INTRODUCTION

The problem of wave phase conjugation (WPC) has long attracted attention because almost all manifestations of the phenomenon are of interest for applications. Phase conjugation is the reversal of the wave vectors in the spectrum of a propagating wave, resulting in a time inverted propagation. This leads sometimes to the denomination "time reversal." The first domain in which WPC was investigated is optics, where the first kind of generation of the conjugate wave can be viewed as a dynamic holographic process. In the four waves mixing (FWM) technique, a grating is created in a third-order nonlinear optical medium by two counter propagating pump planar waves and a weak incident wave is coupled with them by the third-order nonlinearity. A steady solution of this interaction implies the creation of a conjugate wave which can be amplified regarding the incident one. This amplification is steady and depends on the properties of the system and particularly the intensity of the pump waves. In stimulated Brillouin scattering the pump wave excites an electrostrictive effect in a transparent dielectric medium which generates a traveling acoustic wave that modulates the refractive index through the elasto-optic effect. This is equivalent to a dynamic phase grating that diffracts, in the opposite direction, a phase conjugate replica of an incident wave. In all these processes, the pump, incident, and conjugate waves have the same (or almost) frequency. This acousto-optical research started around 1965,[1,2] and gave rise very quickly to theoretical development in order to find the solutions of the coupled equations of the process. The paper of Bobroff and Haus[3] is probably the most elaborate since it solves the case of an infinite system with a finite homogeneous active zone for a wide band incident wave. Bobroff and Haus used a Laplace transform in time and Fourier transform in space, leading to a very complicated formalism but allowing a wide discussion on the different transient solutions. They exhibited the existence of absolutely unstable situations, where the solutions grow exponentially in time. Unfortunately the paper focused more on Brillouin scattering and provides very few comments on these growing solutions. In acoustics, such a possibility was demonstrated experimentally in the same period in ferromagnets[4] and in the early 1970s in $LiNbO_3$ crystals.[5,6] The physical process was not holography but parametric resonance between a spatially uniform electric field (pump) of frequency Ω and an acoustic wave of frequency $\Omega/2$. Unfortunately, the exponential growth was quickly counterbalanced by saturation processes. An intuitive explanation of the interest of a pump modulation of the sound velocity at double frequency of the incident wave can be found in Ref. 7 but, in contrast to the present paper, for stable situations (in the sense of Bobroff and Haus).

Systematic studies of applications of phase conjugation methods were initiated in the early 1980s by Bunkin and coauthors[8] when the progress in experimental techniques in acoustics made these studies possible. The first nonlinear media used to realize a holographic mechanism in acoustics were liquids but the effect was limited by the advective properties of the liquids generating saw tooth waves. In liquids with bubbles[9] the phenomenon could be observed with an efficiency of about 1%.[10] Parametric WPC really developed when the acoustic pumping had been abandoned for the modulation of parameters in the medium. This had as a consequence the choice of some solid active media in which the sound velocity can be modulated by an external field. Among them some magnetostrictive ferrites or rare-earth compounds are of special interest because of their giant magnetostiction, the external field is therefore a magnetic field. A good review of the WPC development can be found in Ref. 11. Of course this presentation of WPC would not be complete without mentioning the technique of the multichannel transmitting receiving antenna arrays[12] which is extremely efficient in operating frequency under 10 MHz but where time reversal is obtained by a numerical process and not by a physical

[a]Electronic mail: alain.merlen@univ-lillel.fr

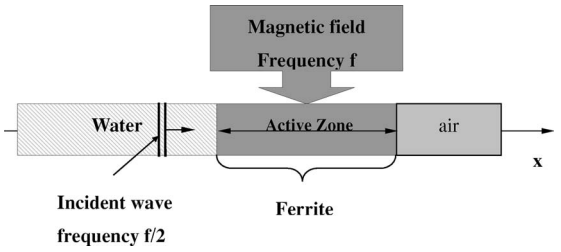


FIG. 1. Principle of magnetoacoustics. The active zone is an elastic solid where the sound velocity can be modulated in real time by a magnetic field. Phase conjugation happens when the pumping frequency is double that of one mode that is present in the spectrum of the incident wave.

effect. This technique has the great advantage of ignoring restrictions on the shape of the signal but its technical complexity and cost limit its applications to laboratory environments.

From the point of view of the applications, the great interest of parametric WPC was demonstrated recently in acoustic imaging[13] but the potential applications are numerous also in nondestructive testing, and for higher energy, in nonintrusive surgery, hyperthermia, or other medical applications.

The present paper provides analytical solutions for parametric WPC focusing on the giant amplification regime characterized by the exponential growth in time. When the energy provided by the pump exceeds the losses due to the forward and conjugate waves traveling out of the active zone, the amplification begins. A threshold of energy has to be surpassed to reach this situation known as the supercritical mode of parametric WPC. The problem is treated with a multiscale method that can be compared in a more formal way to the slow varying envelope approximation (SVEA) often used in optical WPC (Ref. 14) for steady solutions (holographic WPC)[15] or unsteady ones.[16] Here, approximations are introduced in a less heuristic way than in this article and provides both slow time and space variations. In practice, a magnetoacoustic wave conjugator is generally a cylinder of magnetostrictive material partially submitted to a magnetic field (active zone). The incident wave travels approximately along the axis of the conjugator, therefore the most simple theoretical model is the one-dimensional situation presented in Fig. 1 which corresponds to the "paraxial" approximation in optics. Numerical simulation[17,18] performed with this model gave qualitative results in agreement with experiments.[19] In acoustics, in contrast to optics, the saturation process leading to the saturation, does not come from the constitutive law of the medium since the relative variation of the sound velocity with the intensity of the wave is negligible compared with what happens to the refraction index and losses in optics. The nonlinear processes leading to saturation have been shown to be the feedback of elastic energy to electric energy.[18] This happens only when the stress becomes very high and explains why the exponential growth can be observed during a long time before saturation. Therefore, the linear pumping theory is far more useful in acoustics than in optics. The basic assumption of this theory reduces the pumping to a small harmonic oscillation around a constant value of the sound velocity of the medium. In acoustics, the long duration of the exponential growth and consequently, of the validity of the linear pumping assumption, confers practical interest to the analytical solutions of the present model problem. The solutions are derived in a very practical formulation which could been generalized to heterogeneous cases where the active zone is directly in contact with other media. These analytical solutions explain many properties of the phase conjugators (critical threshold, variation of the stress spatial repartition inside the sample) and give theoretical expressions linking these properties to the given parameters. Moreover they can allow the optimization of the impedance ratio between the solid and the fluid in order to extract the maximum acoustical energy possible for a given pumping. This is a key point for medical applications like hyperthermia of cancer cells. Finally, despite the simplification due to the one-dimensional (1D) hypothesis, the conclusions obtained here are very similar to the experimental and numerical ones. By their analytical specificity they help give a deeper and quantitative understanding of the phenomenon, answering questions which were not solved hitherto. The paper successively presents the general solution, the homogeneous, and the semihomogeneous. The choice of the model equation and the way solutions are obtained by the dual-time-scale method are also discussed and clarified.

II. THEORETICAL ANALYSIS

A. The model equation for the velocity

The magnetostrictive material is supposed to be linearly elastic and isotropic. The physical problem is then ruled by the Navier equation. Under the assumption of a unidirectional stress and displacement along the x axis, the only nonzero stress coefficient σ_{xx} can be replaced by

$$\theta = -\frac{\sigma_{xx}}{\rho_0 c}$$

with ρ_0 the density of the solid and c the instantaneous longitudinal sound speed that depends on time. Therefore Navier equation simplifies in

$$\frac{\partial v}{\partial t} + c\frac{\partial \theta}{\partial x} = 0, \tag{1}$$

where v is the displacement velocity along the x axis. With the same variables, the elastic constitutive law reduces to

$$\theta = -c\frac{\partial u}{\partial x}, \tag{2}$$

where u is the displacement, it is related to v by $v = \partial u/\partial t$.

The time derivative of Eq. (2) gives the evolution of θ

$$\frac{\partial \theta}{\partial t} + c\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial x}\frac{\partial c}{\partial t}$$

and taking into account Eq. (1)

$$\frac{\partial \theta}{\partial t} + c\frac{\partial v}{\partial x} = \frac{\theta}{c}\frac{\partial c}{\partial t}. \tag{3}$$

In the frame of the linear pumping assumption, the active properties of the medium can be summarized by a time variation of c such as

$$c^2 = c_0^2[1 + m\cos(\Omega t + \varphi)],$$

where m is a small parameter ($m \ll 1$) referred to as the "modulation depth" and c_0, the constant sound velocity of the material when no magnetic field is applied. The phase φ is a constant that allows an arbitrary starting value for the pumping. It is clear that systems (1) and (3) can be replaced by

$$\frac{\partial \theta}{\partial t} + c\frac{\partial v}{\partial x} = -m\Omega\frac{\theta}{2}\sin(\Omega t + \varphi), \quad \frac{\partial v}{\partial t} + c\frac{\partial \theta}{\partial x} = 0. \tag{4}$$

Introducing new variables $w_1 = v + \theta, w_2 = v - \theta$ in Eq. (4) yields

$$\begin{aligned} &\frac{\partial w_1}{\partial t} + c\frac{\partial w_1}{\partial x} = -m\frac{\Omega}{4}(w_1 - w_2)\sin(\Omega t + \varphi), \\ &\frac{\partial w_2}{\partial t} - c\frac{\partial w_2}{\partial x} = m\frac{\Omega}{4}(w_1 - w_2)\sin(\Omega t + \varphi). \end{aligned} \tag{5}$$

In this system, the modulation of the sound velocity is taken into account both in the nonlinear source term on the right-hand side and also in the advective term on the left-hand side.

Problem (5) with its initial conditions has the symbolic form

$$\begin{aligned} &\frac{\partial U}{\partial t} + \frac{\partial F(U)}{\partial x} = S, \quad \forall x \in \mathbb{R}, t \geqslant 0, \\ &U(x,0) = U_0(x), \qquad \forall x \in \mathbb{R}, \end{aligned} \tag{6}$$

where

$$F(U) = \begin{pmatrix} cw_1 \\ -cw_2 \end{pmatrix}, \quad U = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix},$$

and

$$S = \begin{pmatrix} -m\frac{\Omega}{4}(w_1 - w_2)\sin(\Omega t + \varphi) \\ m\frac{\Omega}{4}(w_1 - w_2)\sin(\Omega t + \varphi) \end{pmatrix}.$$

In this conservative form, the hyperbolic system (6) can be solved numerically by an explicit finite volume method based on second-order Godunov type methods.[17,20,21]

Inserting $c = c_0[1 + m\cos(\Omega t + \varphi)/2]$ in Eq. (4) yields

$$\frac{\partial \theta}{\partial t} + c_0\frac{\partial v}{\partial x} = -m\Omega\frac{\theta}{2}\sin(\Omega t + \varphi) - c_0\frac{m}{2}\cos(\Omega t + \varphi)\frac{\partial v}{\partial x}, \tag{7}$$

$$\frac{\partial v}{\partial t} + c_0\frac{\partial \theta}{\partial x} = -c_0\frac{m}{2}\cos(\Omega t + \varphi)\frac{\partial \theta}{\partial x}. \tag{8}$$

The elimination of θ in the left-hand side leads to

$$\frac{\partial^2 v}{\partial t^2} - c_0^2\frac{\partial^2 v}{\partial x^2} = c_0^2\frac{m}{2}\cos(\Omega t + \varphi)\frac{\partial^2 v}{\partial x^2} - c_0\frac{m}{2} \times \cos(\Omega t + \varphi)\frac{\partial^2 \theta}{\partial x\,\partial t}.$$

Using Eq. (7), i.e., $\partial^2\theta/\partial x\,\partial t = -c_0\partial^2 v/\partial x^2 + O(m)$ gives the model equation of WPC for v in the active zone

$$\frac{\partial^2 v}{\partial t^2} - c_0^2\frac{\partial^2 v}{\partial x^2} = mc_0^2\cos(\Omega t + \varphi)\frac{\partial^2 v}{\partial x^2}. \tag{9}$$

Here the active zone is defined by $|x| \leqslant L/2$. Outside the active zone the modulation depth is switched to 0 and the equation becomes the linear wave equation.

The surrounding fluid is treated exactly the same way except that c becomes c_f and θ is defined by $\theta = p/(\rho_0 c_f)$ where p is the acoustic pressure and c_f the constant sound speed in the fluid. The modulation depth in the fluid is zero.

B. General resonant solutions

Equation (9) in the active zone is simply the wave equation with a small perturbation on the right-hand side. Regular asymptotic solutions can be obtained by a development in powers of m, but it is easy to show that in the specific case where $\Omega = 2\omega$ and $L \to \infty$, a dimensionless form of Eq. (9) is

$$\ddot{V} + [1 - m\cos(2\tau_0 + \varphi)]V = 0,$$

where $\tau_0 = \omega t, v = V(\tau_0)\sin(k_0 x)$, $k_0 = \omega/c_0$, and $\ddot{V} = d^2V/d\tau_0^2$. This is obviously a Mathieu equation well known to exhibit solution terms of the form: $e^{m\tau_0/4}\cos\tau_0$ or $e^{m\tau_0/4}\sin\tau_0$.[22] In dimensional form this corresponds to an exponential growth $e^{\Gamma t}$ with $\Gamma = m\Omega/8$. This gives the maximum growth rate possible. In terms of physical interpretation, this means that if the active zone is infinite, and the pumping has the frequency Ω, the mode $\omega = \Omega/2$ of any wave inside the sample is exponentially amplified. Bobroff and Haus[3] provided results for an homogeneous medium including a limited active zone. Nevertheless the Laplace transform technique used in this paper is not very manipulable and the reader who requires practical results may have difficulty finding them. This is particularly true for the value of Γ that depends, in this case, on the size of the active zone. More explicit formulas on supercritical modes were given for microwave phase conjugation in finite crystals[5] but without explicit derivation. In Ref. 5, losses of energy were explicitly introduced in the equations and partial reflections were assumed at the edges of the crystal.

Hereafter a very straightforward resolution is proposed that is limited to resonant solutions in a nondissipative finite active zone in contact with arbitrary media on both sides. Nonresonant disturbances are disregarded but can be treated later as regular perturbations of the present solutions.

The resolution begins by considering the fact that the system can be seen as an oscillator of varying frequency in the range: $k_0 c_0(1 - m/2)$ and $k_0 c_0(1 + m/2)$ where k_0 is the wave number of the incident wave. The resonance conditions

are fulfilled, i.e., $k_0=k/2$, with $k=\Omega/c_0$. In the range of excitation, the wave numbers of the modes can be written

$$\frac{k}{2}\left(1\pm\frac{\alpha}{2}\right),\quad \text{for } 0\leqslant\alpha\leqslant m,$$

where α is the frequency shift amplitude relative to k_0.

The general form of an excited mode is

$$v=A(t,m)\sin\left[\frac{k}{2}\left(1\pm\frac{\alpha}{2}\right)x+\xi\right], \tag{10}$$

where ξ is an arbitrary constant. For clarity, in the further development, the minus sign is chosen in the previous expression. With this choice, Eq. (9) becomes

$$\begin{aligned}&\frac{d^2A}{dt^2}\sin\left[\frac{k}{2}\left(1-\frac{\alpha}{2}\right)x+\xi\right]+\frac{c_0^2k^2}{4}\left(1-\frac{\alpha}{2}\right)^2A\\&\quad\times\sin\left[\frac{k}{2}\left(1-\frac{\alpha}{2}\right)x+\xi\right]\\&\quad=c_0^2m\cos(\Omega t+\varphi)\left[\frac{k^2}{4}\left(1-\frac{\alpha}{2}\right)^2A\sin\left(\frac{k}{2}\left(1-\frac{\alpha}{2}\right)x+\xi\right)\right].\end{aligned}$$

Since α and m are small parameters, a development to the first order in α and m yields

$$\frac{d^2A}{dt^2}+\frac{c_0^2k^2}{4}(1-\alpha)A=m\frac{k^2c_0^2}{4}A\cos(\Omega t+\varphi).$$

There are two characteristic time scales in this expression: $4/(kc_0)$ and a "slower one" $4/(mkc_0)$. A dual-time-scale method, renormalizes the time scales by $\tau=t, T=mt$ and allows the solution to be written as a regular development of the form: $A(t,m)=A_0(\tau,T)+mA_1(\tau,T)$. The equation for A becomes

$$\begin{aligned}&\frac{d^2A_0}{\partial\tau^2}+\frac{c_0^2k^2}{4}A_0-\alpha\frac{c_0^2k^2}{4}A_0+m\left[\frac{\partial^2A_1}{\partial\tau^2}+2\frac{\partial^2A_0}{\partial\tau\,\partial T}+\frac{c_0^2k^2}{4}A_1\right]\\&\quad=m\frac{c_0^2k^2}{4}A_0\cos(\Omega t+\varphi).\end{aligned}$$

The primary order leads to

$$\frac{\partial^2A_0}{\partial\tau^2}+\frac{c_0^2k^2}{4}A_0=0,$$

the general solution of which can be written $A_0=a(T)\cos[c_0k\tau/2+\psi(T)+\varphi/2]$.

The first order in m (and α) is

$$\begin{aligned}\frac{\partial^2A_1}{\partial\tau^2}+\frac{c_0^2k^2}{4}A_1&=\frac{c_0^2k^2}{4}A_0\cos(\Omega t+\varphi)-2\frac{\partial^2A_0}{\partial\tau\,\partial T}\\&\quad+\frac{c_0^2k^2\alpha}{4m}A_0.\end{aligned}$$

Taking into account the expression of A_0, developing and neglecting the nonsecular terms (of pulsation $3\Omega/2$), this equation reduces to

$$\begin{aligned}\frac{\partial^2A_1}{\partial\tau^2}+\frac{\Omega^2}{4}A_1&=\frac{\Omega^2}{8}a(T)\cos\left[\frac{\Omega}{2}\tau+\psi(T)+\frac{\varphi}{2}\right]\cos[2\psi(T)]\\&\quad+\frac{\Omega^2}{8}a(T)\sin\left[\frac{\Omega}{2}\tau+\psi(T)+\frac{\varphi}{2}\right]\sin[2\psi(T)]\\&\quad+\Omega\frac{da(T)}{dT}\sin\left[\frac{\Omega}{2}\tau+\psi(T)+\frac{\varphi}{2}\right]\\&\quad+\Omega a(T)\frac{d\psi(T)}{dT}\cos\left[\frac{\Omega}{2}\tau+\psi(T)+\frac{\varphi}{2}\right]\\&\quad+\frac{\Omega^2}{4}\frac{\alpha}{m}a(T)\cos\left[\frac{\Omega}{2}\tau+\psi(T)+\frac{\varphi}{2}\right]\\&\quad+\text{nonsecular terms}.\end{aligned}$$

The development of A is valid if the secular terms are zero, i.e.,

$$\frac{da(T)}{dT}=-a(T)\frac{\Omega}{8}\sin[2\psi(T)] \tag{11}$$

$$-\frac{d\psi(T)}{\frac{1}{2}\cos[2\psi(T)]+\frac{\alpha}{m}}=\frac{1}{4}\Omega dT. \tag{12}$$

Equation (12) has a periodic solution if $\alpha/m>1/2$

$$\tan\left(\frac{\sqrt{\frac{\alpha^2}{m^2}-\frac{1}{4}}}{2}\Omega(T-T_0)\right)=-\frac{\sqrt{\frac{\alpha^2}{m^2}-\frac{1}{4}}\sin[2\psi(T)]}{\frac{1}{2}+\frac{\alpha}{m}\cos[2\psi(T)]},$$

where T_0 is the integration constant. In this case, $\psi(T)$ is a periodic function of frequency $\Omega\sqrt{\frac{\alpha^2}{m^2}-\frac{1}{4}}/(4\pi)$. Elimination of T for ψ in Eq. (11) and integration lead to

$$a=a_0\sqrt{\frac{1}{2}\cos(2\psi)+\frac{\alpha}{m}},$$

where a_0 is a constant. This shows that a is also a periodic function with the same frequency. It is obvious that the change of the minus sign by a plus in Eq. (10) gives the same result with $\psi^+=\pi-\psi^-$ where subscripts correspond to the choice of the sign in Eq. (10).

More interestingly is the case $\alpha/m<1/2$, for the solution of Eq. (12) is

$$-\frac{\Omega}{2}\sqrt{\frac{1}{4}-\frac{\alpha^2}{m^2}}(T-T_0)=\tanh^{-1}\frac{\sqrt{\frac{1}{4}-\frac{\alpha^2}{m^2}}\sin[2\psi(T)]}{\frac{1}{2}+\frac{\alpha}{m}\cos[2\psi(T)]}.$$

For large enough T, this leads to

$$\frac{\sqrt{\frac{1}{4}-\frac{\alpha^2}{m^2}}\sin[2\psi(T)]}{\frac{1}{2}+\frac{\alpha}{m}\cos[2\psi(T)]} = -\tanh\frac{\Omega}{2}\sqrt{\frac{1}{4}-\frac{\alpha^2}{m^2}} \times (T-T_0) \to -1.$$

After a transient time, the asymptotic solution of this equation is $\cos[2\psi(T)]=-2\alpha/m$, and $\sin[2\psi(T)]=-\sqrt{1-4\alpha^2/m^2}$. Hence for $m \geqslant 2\alpha \geqslant 0$, ψ tends toward a constant limit

$$\psi_0 = \pm\frac{1}{2}\arccos\left(-\frac{2\alpha}{m}\right). \tag{13}$$

Replacing $\sin[2\psi(T)]$ in Eq. (11) and for large values of T yields

$$\frac{da(T)}{a(T)} = \frac{\Omega}{8}\sqrt{1-\left(\frac{2\alpha}{m}\right)^2}dT$$

the solution of which, after the initial transient, is

$$a(T) = a_0 e^{\frac{\Omega}{8}(T-T_0)\sqrt{1-\frac{(2\alpha)^2}{m^2}}} = a_0 e^{\frac{\Omega}{8}(t-t_0)\sqrt{m^2-(2\alpha)^2}}.$$

Here $T_0=mt_0$ corresponds to the end of the initial transient time, where $\psi \to \psi_0$, instead of the beginning of the pumping. Returning to the variable t, the general solution is

$$A(t,m) = A_0[1+O(m)] = a_0 e^{\frac{\Omega}{8}(t-t_0)\sqrt{m^2-(2\alpha)^2}}\cos\left(\frac{\Omega}{2}t+\psi_0+\frac{\varphi}{2}\right)[1+O(m)]. \tag{14}$$

The change of α to $-\alpha$ leaves the results unchanged. This shows that exponentially growing solutions are possible if the boundary conditions of the problem select a mode such as $2\alpha<m$, the growth rate of which $\Gamma=\Omega\sqrt{m^2-(2\alpha)^2}/8$ is independent of the sign of the frequency shift. Finally, after the initial transient, a general solution for v is composed by elements of the form

$$E_0 = a_0 e^{\Gamma t}\cos\left[\frac{\Omega}{2}t \pm \psi_0 + \varphi\right]\sin\left(\frac{k}{2}\left[1\pm\frac{\alpha}{2}\right]x+\xi\right), \tag{15}$$

where the constants a_0, ξ, φ, and undetermined signs are fixed by the symmetry and boundary conditions of the problem. Equation (7) or (8) shows that, neglecting terms of order m, exponentially growing solutions of type (15) exist also for θ.

It is therefore interesting to consider the model equation for θ, obtained by elimination of v between Eqs. (7) and (8)

$$\frac{d^2\theta}{dt^2}-c_0^2\frac{\partial^2\theta}{\partial x^2} = mc_0^2\cos(\Omega t+\varphi)\frac{\partial^2\theta}{\partial x^2}+\frac{m}{2}\Omega^2\cos(\Omega t+\varphi)\theta.$$

Taking solutions of type (15) for one value of α and neglecting terms of order m yields

$$\frac{\Omega^2}{4}\theta = -c_0^2\frac{\partial^2\theta}{\partial x^2},$$

and finally

$$\frac{\partial^2\theta}{\partial t^2}-c_0^2\frac{\partial^2\theta}{\partial x^2} = -mc_0^2\cos(\Omega t+\varphi)\frac{\partial^2\theta}{\partial x^2},$$

which shows that the resonant modes for θ fulfill the same model equation as v.

It must be underlined that, as stated previously, the weaker the shift α the stronger the amplification. Therefore, only the most amplified mode, with a selected value α_s of the frequency shift amplitude, can be observed in the numerical results.[17] Hereafter the physical selection process of α_s is revealed explicitly by analytical solutions.

III. PHYSICAL RESONANT SOLUTIONS

Resonant solutions for v and θ are composed of a linear combination of elements E_0. It is therefore also the case for w_1 and w_2. Since the boundary condition of the resonant solutions are easier to write analytically for w_1 and w_2, this representation of the wave system is used in the following section. Moreover, any linear combination of elements E_0 can be used as a basis of the space of solutions. Expressions like

$$W_0 = \mathrm{C}e^{\Gamma t}\sin\frac{k\alpha}{4}(x+\xi)\sin\left(\frac{k}{2}x\pm\frac{\Omega}{2}t\pm\varphi_0\right) \tag{16}$$

are also general solutions of the resonant problem and are very useful for solving situations with finite active zones.

A. General case

Here the active zone of length L is in contact on the left and on the right with media of different acoustic impedance (Fig. 1). This general case was unsolved explicitly in the literature. Let $\tau_l=(\rho_l c_l)/(\rho_0 c_0)$ be the impedance ratio between the left medium and the active zone and $\tau_r=(\rho_r c_r)/(\rho_0 c_0)$ the same for the right edge of the active zone. In the above expressions, subscripts l and r correspond, respectively, to the left and right media.

Starting with elementary waves W_0, w_1 and w_2, inside the active zone, are given by

$$w_1 = -\mathrm{C}e^{\Gamma t}\left[\sin\left(\frac{k\alpha}{4}x+\xi_1\right)\sin\left(-\frac{k}{2}x+\frac{\Omega}{2}t+\varphi_1\right)+mf_1(x,t)\right] \tag{17}$$

and

$$w_2 = \mathrm{C}e^{\Gamma t}\left[\sin\left(\frac{k\alpha}{4}x+\xi_2\right)\sin\left(\frac{k}{2}x+\frac{\Omega}{2}t+\varphi_2\right) - mf_1(x_0-x,t)\right], \tag{18}$$

where ξ_1 and ξ_2 are unknowns to be determined by the boundary conditions at $x=\pm L/2$. The function f_1 represents the terms of order m in the solution and x_0 is a constant. It can be seen that Eqs. (17) and (18) respect the invariability of system (5) for the change of variables $x \to x_0-x, w_1 \to w_2$ under suitable conditions on $\xi_1, \xi_2, \varphi_1, \varphi_2$.

At $x=-L/2$, only the conjugate wave can radiate toward the left, hence, outside of the active zone, $w_{1l}=v_l-\theta_l=0$ and therefore $w_{2l}=v_l-\theta_l=2v_l=-2\theta_l$. The other conditions are the

stress continuity at $x=-L/2, \theta(-L/2)=\tau_l\theta_l(-L/2)$ and the velocity continuity $v(-L/2)=v_l(-L/2)$. These conditions can be expressed only with the variables w_1 and w_2, since

$$w_2\left(-\frac{L}{2},t\right)=v_l\left(-\frac{L}{2},t\right)-\tau_l\theta_l\left(-\frac{L}{2},t\right)=\frac{(1+\tau_l)}{2}w_{2l}\left(-\frac{L}{2},t\right)$$

and

$$w_1\left(-\frac{L}{2},t\right)=v_l\left(-\frac{L}{2},t\right)+\tau_l\theta_l\left(-\frac{L}{2},t\right)=\frac{(1-\tau_l)}{2}w_{2l}\left(-\frac{L}{2},t\right).$$

These lead to

$$w_1\left(-\frac{L}{2},t\right)=\frac{(1-\tau_l)}{(1+\tau_l)}w_2\left(-\frac{L}{2},t\right) \tag{19}$$

and, using Eqs. (17) and (18) and neglecting terms of order m

$$-\sin\left(-\frac{k\alpha L}{8}+\xi_1\right)\sin\left(\frac{kL}{4}+\frac{\Omega}{2}t+\varphi_1\right)$$
$$=\frac{1-\tau_l}{1+\tau_l}\sin\left(-\frac{k\alpha L}{4}+\xi_2\right)\sin\left(-\frac{kL}{4}+\frac{\Omega}{2}t+\varphi_2\right)\ \forall\, t.$$

Developing this gives for all t

$$\sin\left(\frac{kL}{4}+\frac{\Omega t}{2}+\varphi_1\right)\left[\sin\left(-\frac{k\alpha L}{8}+\xi_1\right)+\frac{1-\tau_l}{1+\tau_l}\times\sin\left(-\frac{k\alpha L}{8}+\xi_2\right)\cos\left(-\frac{kL}{2}+\varphi_2-\varphi_1\right)\right]$$
$$+\cos\left(\frac{kL}{4}+\frac{\Omega t}{2}+\varphi_1\right)\left[\frac{1-\tau_l}{1+\tau_l}\sin\left(-\frac{k\alpha L}{8}+\xi_2\right)\times\sin\left(-\frac{kL}{2}+\varphi_2-\varphi_1\right)\right]=0.$$

Explicitly, this yields

$$\sin\left(-\frac{kL}{2}+\varphi_2-\varphi_1\right)=O(m) \tag{20}$$

and

$$\sin\left(-\frac{k\alpha L}{8}+\xi_1\right)\pm\frac{1-\tau_l}{1+\tau_l}\sin\left(-\frac{k\alpha L}{8}+\xi_2\right)=O(m). \tag{21}$$

The same can be done at $x=L/2$ where the boundary condition $w_{2r}=0$ and the continuity of v and stresses leads to

$$w_1\left(\frac{L}{2},t\right)=\frac{(1+\tau_r)}{(1-\tau_r)}w_2\left(\frac{L}{2},t\right) \tag{22}$$

and finally to

$$\sin\left(\frac{kL}{2}+\varphi_2-\varphi_1\right)=O(m) \tag{23}$$

and

$$\sin\left(\frac{k\alpha L}{8}+\xi_1\right)\pm\frac{1+\tau_r}{1-\tau_r}\sin\left(\frac{k\alpha L}{8}+\xi_2\right)=O(m). \tag{24}$$

Equations (20) and (23) are compatible only if: $kL=O(m)\times(\mathrm{mod}\,\pi)$ or in terms of wave length of the conjugate wave λ if $L/\lambda-p/4=O(m)$ where p is an integer. This condition is not very strict and is always fulfilled in practice. Equations (21) and (24) provide relations between ξ_1, ξ_2, and α. One more equation is needed to solve the problem. This closure equation comes from the fact that Eqs. (17) and (18) are not solutions of the problem for any arbitrary set of values of these three parameters. Taking Eq. (5) and developing the advective term around c_0, the governing equations become

$$\frac{\partial w_1}{\partial t}+c_0\frac{\partial w_1}{\partial x}=-m\frac{\Omega}{4}(w_1-w_2)\sin(\Omega t+\varphi)-\frac{m}{2}c_0\frac{\partial w_1}{\partial x}\cos(\Omega t+\varphi),$$

$$\frac{\partial w_2}{\partial t}-c_0\frac{\partial w_2}{\partial x}=m\frac{\Omega}{4}(w_1-w_2)\sin(\Omega t+\varphi)+\frac{m}{2}c_0\frac{\partial w_2}{\partial x}\cos(\Omega t+\varphi).$$

Taking, for example, the equation of w_1 and introducing Eqs. (17) and (18) yields

$$\left[\Gamma\sin\left(\frac{k\alpha x}{4}+\xi_1\right)+\frac{\Omega\alpha}{4}\cos\left(\frac{k\alpha x}{4}+\xi_1\right)\right]\times\sin\left(-\frac{kx}{2}+\frac{\Omega t}{2}+\varphi_1\right)+m\left(\frac{\partial}{\partial t}+c_0\frac{\partial}{\partial x}\right)f_1$$
$$=\frac{m\Omega}{8}\left[2\sin\left(\frac{k\alpha x}{4}+\xi_1\right)\cos\left(-\frac{kx}{2}+\frac{3\Omega t}{2}+\varphi_1+\varphi\right)+\sin\left(\frac{k\alpha x}{4}+\xi_2\right)\cos\left(\frac{kx}{2}+\frac{3\Omega t}{2}+\varphi_2+\varphi\right)-\sin\left(\frac{k\alpha x}{4}+\xi_2\right)\cos\left(-\frac{kx}{2}+\frac{\Omega t}{2}-\varphi_2+\varphi\right)\right]. \tag{25}$$

All these terms are of order $m\Omega$ and this relation holds for all t and x. An explicit expression of f_1 is not needed. It is enough to know that the term $(\partial/\partial t+c_0\partial/\partial x)f_1$ does not contain functions $g(\varsigma)$, with $\varsigma=(-kx/2+\Omega t/2+\varphi_1)$, since any of such functions is a solution of the operator $\partial/\partial t+c_0\partial/\partial x$. Moreover any function of this kind, slowly modulated in space by functions of $k\alpha x/4$, are rejected to order $m\alpha\Omega$. Hence, neglecting terms of order m, projection of Eq. (25) on $\sin(-kx/2+\Omega t/2+\varphi_1)$ reduces to

$$\Gamma\sin\left(\frac{k\alpha x}{4}+\xi_1\right)+\frac{\Omega\alpha}{4}\cos\left(\frac{k\alpha x}{4}+\xi_1\right)=\frac{m\Omega}{8}\sin\left(\frac{k\alpha x}{4}+\xi_2\right)\sin(\varphi-\varphi_1-\varphi_2)\ \ \forall\, x. \tag{26}$$

Projection on $\cos(-kx/2+\Omega t/2+\varphi_1)$ gives

$$\cos(\varphi-\varphi_1-\varphi_2)=O(m)$$

that immediately leads to

$$\varphi_2 = \varphi_1 + \varphi + \frac{\pi}{2} + O(m) \ (\mathrm{mod}\ \pi). \tag{27}$$

Taking into account Eqs. (20) and (23), this yields

$$\varphi_1 = \frac{\varphi}{2} - \frac{kL}{4} + O(m) \tag{28}$$

and

$$\varphi_2 = \frac{\varphi}{2} + \frac{kL}{4} + O(m). \tag{29}$$

From Eq. (26), this result yields

$$\cos(\gamma)\sin\left(\frac{k\alpha x}{4} + \xi_1\right) + \sin(\gamma)\cos\left(\frac{k\alpha x}{4} + \xi_1\right) = \pm \sin\left(\frac{k\alpha x}{4} + \xi_2\right) \forall\, x, \tag{30}$$

where, according to the results of Sec. II, $\cos(\gamma) = 8\Gamma/(m\Omega)$ and $\sin(\gamma) = 2\alpha/m$. For fulfilling this relation for all x, a positive sign has to be chosen on the right-hand side, hence

$$\sin\left(\frac{k\alpha x}{4} + \xi_1 + \gamma\right) = \sin\left(\frac{k\alpha x}{4} + \xi_2\right) \tag{31}$$

and finally

$$\gamma = \xi_2 - \xi_1 + O(m) \quad (\mathrm{mod}\ 2\pi). \tag{32}$$

The same process, applied to the equation for w_2, leads to exactly the same results. Relations (32) in Eq. (21) and Eq. (24) lead to

$$\tan\xi_1 = \frac{\sin\left(\frac{k\alpha L}{8}\right) + \frac{1-\tau_l}{1+\tau_l}\sin\left(-\frac{k\alpha L}{8} + \gamma\right)}{\cos\left(\frac{k\alpha L}{8}\right) - \frac{1-\tau_l}{1+\tau_l}\cos\left(-\frac{k\alpha L}{8} + \gamma\right)}$$

and

$$\tan\xi_1 = \frac{-\sin\left(\frac{k\alpha L}{8}\right) + \frac{1+\tau_r}{1-\tau_r}\sin\left(\frac{k\alpha L}{8} + \gamma\right)}{\cos\left(\frac{k\alpha L}{8}\right) - \frac{1+\tau_r}{1-\tau_r}\cos\left(\frac{k\alpha L}{8} + \gamma\right)}.$$

Elimination of ξ_1 leads to

$$\left[\tan\left(\frac{\gamma}{2}\right)\right]^2 - (\tau_l + \tau_r)\cot\left(\frac{k\alpha L}{4}\right)\tan\left(\frac{\gamma}{2}\right) - \tau_l\tau_r = 0 \tag{33}$$

the solution of which is

$$\tan\frac{\gamma}{2} = \frac{1}{2}\left[(\tau_l + \tau_r)\cot\left(\frac{k\alpha L}{4}\right) \pm \sqrt{\left[(\tau_l + \tau_r)\cot\left(\frac{k\alpha L}{4}\right)\right]^2 + 4\tau_l\tau_r}\right].$$

Finally, according to the definition of γ, the value $\alpha = \alpha_s$ selected by the system must be the lowest positive solution of

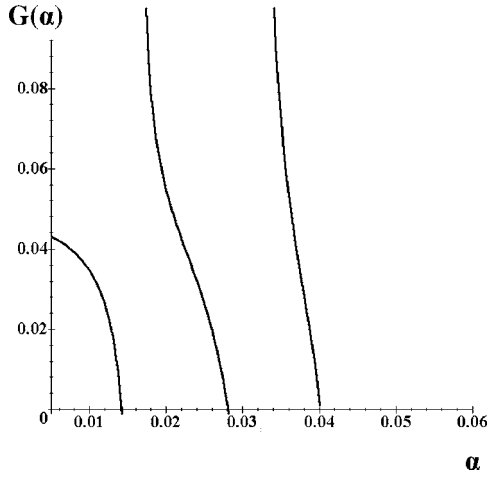


FIG. 2. Function $G(\alpha)$. It can be seen that the solutions of $G(\alpha)=0$ can be multiplied. Nevertheless since the rate of exponential amplification is a decreasing function of α, only the smallest solution is relevant.

$$G(\alpha) = \frac{\sqrt{m^2 - (2\alpha)^2}}{2\alpha} - \cot[2\ \arctan\ G_0(\alpha)] = 0 \tag{34}$$

with

$$G_0(\alpha) = \frac{(\tau_l + \tau_r)\cot\left(\frac{k\alpha L}{4}\right) \pm \sqrt{\left[(\tau_l + \tau_r)\cot\left(\frac{k\alpha L}{4}\right)\right]^2 + 4\tau_l\tau_r}}{2}. \tag{35}$$

Of course Eq. (34) can have multiple solutions for each set of parameters (Fig. 2) but, since the lowest α_s corresponds to the highest value of Γ, only this most amplified mode is observable.

Another expression can be derived from Eq. (33) by expressing $\tan(\gamma/2)$ as a function of $\tan\gamma$, and leads to

$$\tan\left(\frac{k\alpha L}{4}\right) = \frac{2\alpha(\tau_r + \tau_l)}{\pm m(1 - \tau_r\tau_l) - (1 + \tau_r\tau_l)\sqrt{m^2 - (2\alpha)^2}}. \tag{36}$$

This is a general form of the formula given by Thompson and Quate[5] with a "+" sign instead of a "±". The reader can verify this equivalence using the following change in notation:

$$\text{Coupling coefficient } \kappa = \frac{mk}{8}$$

$$\text{Reflection coefficient } r = \frac{1-\tau}{1+\tau}.$$

Table I gives the solution of Eq. (34) for $m = 2.8 \times 10^{-2}$, $\Omega = 2\omega = 4\pi \times 10^7\ \mathrm{s}^{-1}$, $L = 2.5$ cm, and $c_0 = 4000$ m/s which are the classical values obtained by experiments[23] for different values of τ_l and τ_r. The smaller α_s is the higher the amplification rate Γ.

Figure 3 shows the comparison between the computation and the value of α_s predicted by Eq. (34) taking the case $\tau_l = 0.3$ and $\tau_r = 0.1$ as a generic example. The figure shows the

TABLE I. General case: α_s versus τ_l and τ_r for generic materials.

Left material	Active material	Right material	τ_l	τ_r	α_s
Air	Ferrite	Air	≈ 0	≈ 0	≈ 0
Water	Ferrite	Air	0.0412	≈ 0	0.002 329
Water	Ferrite	Water	0.0412	0.0412	0.003 341
Glass	Ferrite	Air	0.4073	≈ 0	0.005 592
Glass	Ferrite	Water	0.4073	0.0412	0.006 225
Glass	Ferrite	Glass	0.4073	0.4073	0.008 753
Steel	Ferrite	Air	1.2996	≈ 0	0.006 970
Steel	Ferrite	Water	1.2996	0.0412	0.007 576
Steel	Ferrite	Steel	1.2996	1.2996	0.010 558

stress just at the left edge of the active zone where the conjugate wave leaves the active zone. The simulation starts with the following initial condition, used for all the computations presented in this paper:

$$\begin{cases} w_1 = w_2 = 2 \times 10^{-3} \sin\left[\dfrac{2\pi}{\lambda}(x - x_L)\right] \text{ for } x_L < x < x_R \\ w_1 = w_2 = 0 \text{ elsewhere.} \end{cases} \tag{37}$$

Abscissa x_L and x_R are chosen such that $(x_R - x_L) = 3\lambda$ where λ is the wave length: $\lambda = 2\pi\omega/c_0$. It is a pulse of three waves of angular frequency $\omega = \Omega/2$ released in the middle of the active zone. This pulse shares in two waves propagating in opposite directions. The part of this wave propagating toward the left (w_2) is visible on the figure at approximately 3 μs. This initial condition is just a starter for the parametric resonance and has been chosen because its spectrum is dominated by frequency ω. The symmetrical propagation of both initial pulses is not necessary for the process but it allows one to illustrate the accuracy of the numerical computation

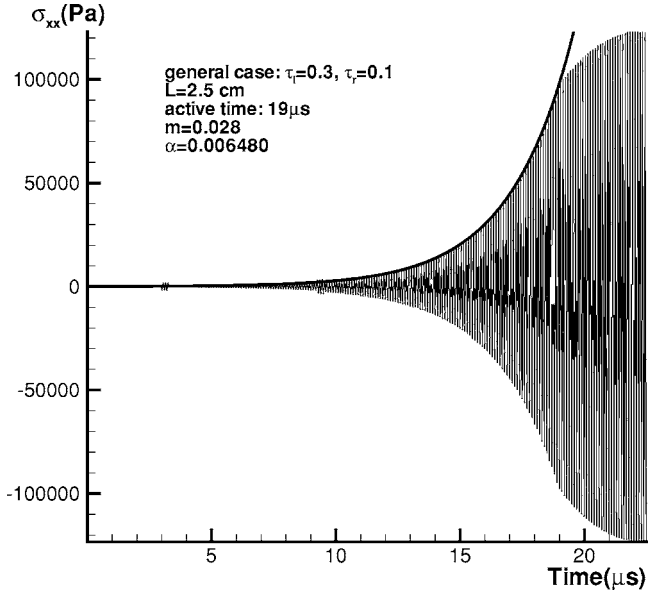


FIG. 3. Time evolution of the stress at the left edge of the active zone in the general case. The pumping duration is 19 μs. $c_0 = 4 \times 10^3$ m/s, $\Omega = 4\pi \times 10^7$, $\rho_0 = 9000$ kg/m^3. — theory [$\mathrm{Ce}^{\Gamma t} \sin(-k\alpha L/8 + \xi_2)$], — computation [$w_2(-L/2, t)$ by Godunov scheme[20]].

of the waves. The reflection of the initial forward pulse (w_1) on the right edge of the active zone is visible, slightly amplified by the parametric process around 9 μs. The time between both pulses is exactly the time needed to cross the active zone with the sound velocity c_0 showing that the numerical scheme does not introduce any numerical dispersion. With such symmetrical initial conditions, the naming of the direct and conjugate wave is purely a matter of convention since the initial signal does not come from a determinate direction. In practice, the initial condition is the instantaneous situation created by a traveling wave originating, for example, from the left w_1,[17] thus the conjugate wave is clearly the counter propagating wave w_2.

The waves w_1 and w_2 created inside the active zone by the parametric resonance with components of angular frequency ω in the spectrum of the initial pulse, are symmetrical until they reach the edges of the active zone. Then, they partially reflect and propagate back to the opposite edges and once they reach them, the wave repartition inside the sample adopts its definitive shape. The only evolution remaining inside the active zone is a spatially homogeneous exponential growth. The direct (w_1) and conjugate wave (w_2), perfectly in phase with the components ω of the initial pulse, are partially evacuated through the edges. This process lasts until the end of the pumping. Afterward, waves w_1 and w_2 become classical spatially modulated progressive waves. They evacuate the active zone at the local sound speed with more or less reflections on the edges according to the impedance ratios.

At exactly 19 μs the theoretical exponential curve $\mathrm{Ce}^{\Gamma t}$ and the computation diverges since the pumping stops and the conjugate waves begin to evacuate the active zone. The shape of the signal is then the same as the spatial repartition in the active zone.

The simulation which presented in Fig. 3 and the following have been performed with the same second order Godunov scheme with slope limiter as in Refs. 20 and 21 in which the method has shown its capacity for elastic waves and WPC simulations. The main characteristic of the scheme is its very low numerical dissipation which avoids the nonphysical damping of the high frequency oscillation ω for a sufficiently refined grid. In these 1D simulations, there are around 100 points for one wave length λ which well captures this component of the spectrum. This resolution is also sufficient for capturing the small, first-order nonresonant $3\Omega/2$ component of the spectrum neglected in this theory but, of course, present in the simulation.

B. Homogeneous medium with finite active zone

Applied to the homogeneous situation, the present theory allows an elegant comparison of analytical results with simulations, previous theories, and experiments. In this case the active zone of length L is in contact on both sides with the same medium such that $\tau_l = \tau_r = 1$. The resonant solution must fulfill the condition $w_{1l}(-L/2, t) = w_1(-L/2, t) = 0 \ \forall t$ and $w_{2r}(L/2, t) = w_2(L/2, t) = 0 \ \forall t$. Equation (16) gives the general form of the direct wave w_1 in the active zone

$$w_1 = \mathrm{C}e^{\Gamma t} \sin \frac{k\alpha}{4}\left(x + \frac{L}{2}\right)\sin\left(\frac{k}{2}x - \frac{\Omega}{2}t - \varphi\right)$$

and

$$w_2 = \mathrm{C}e^{\Gamma t} \sin \frac{k\alpha}{4}\left(x - \frac{L}{2}\right)\sin\left(\frac{k}{2}x + \frac{\Omega}{2}t + \varphi\right)$$

for w_2.

Equation (32) leads to $\gamma = -k\alpha L/4$ and function G has the far simpler form

$$G_h(\alpha) = \frac{\sqrt{m^2 - (2\alpha)^2}}{2\alpha} + \cot\left(\frac{k\alpha L}{4}\right) = 0. \tag{38}$$

This equation can receive a specific interpretation here. In fact, in the homogeneous case, Eq. (31) becomes trivial and the right-hand side of Eq. (30) becomes 0 at $x = L/2$. This allows one to show that the value of α_s is selected by the system because of the continuity of the waves at the boundaries. Taking w_1, for example, for $x > L/2$ the direct wave is a solution of the unperturbed linear advection equation

$$\left(\frac{\partial}{\partial t} + c_0\frac{\partial}{\partial x}\right)w_1 = 0.$$

Since $w_2 = 0$ for $x > L/2$, the following relations hold:

$$w_1 = 2v = 2\theta = F(x - c_0 t).$$

The continuity of v and θ and therefore of w_1 at $x = L/2$, allows one to write the solution of w_1^+ for $x > L/2$, as

$$w_1^+ = \mathrm{C}e^{\Gamma L/2c_0}e^{\Gamma(c_0 t - x)/c_0} \sin \frac{k\alpha L}{4} \times \sin\left[\frac{k}{2}(x - c_0 t - \varphi)(1 + O(m))\right].$$

Since, for all t, $w_1^- = w_1^+$ at $x = L/2$ for all t, the equality

$$\left.\frac{\partial w_1^-}{\partial t}\right|_{L/2} = \left.\frac{\partial w_1^+}{\partial t}\right|_{L/2}$$

also holds for all t and therefore, from the advection equation, yields

$$\left.\frac{\partial w_1^-}{\partial x}\right|_{L/2} = \left.\frac{\partial w_1^+}{\partial x}\right|_{L/2} + m\frac{\Omega}{4c_0}w_1|_{L/2}\sin(\Omega t + \varphi).$$

This reduces to Eq. (30) at $x = L/2$, written in a condensed way

$$\alpha \cot\left(\frac{\pi}{2}\frac{\alpha}{\alpha_0}\right) = -\frac{\sqrt{m^2 - (2\alpha)^2}}{2}, \tag{39}$$

where

$$\alpha_0 = \frac{2\pi c_0}{\Omega L}. \tag{40}$$

Equation (39) is simply the continuity of the slope of the amplitude of the direct wave through the edge of the active zone at $x = L/2$. It can be obtained directly from

TABLE II. Comparison between numerical and theoretical values of α_s.

m	α_s computation	α_s theory
0.016	0.008 000	0.008 000
0.028	0.011 396	0.011 248
0.041	0.012 698	0.012 622

$$\left.\frac{\partial A_{w_1^+}}{\partial x}\right|_{L/2} = \left.\frac{\partial A_{w_1^-}}{\partial x}\right|_{L/2}$$

with

$$A_{w_1^-} = Ae^{\frac{\Omega t}{8}(t)\sqrt{m^2-(2\alpha)^2}} \sin \frac{k\alpha}{4}\left(x + \frac{L}{2}\right).$$

and

$$A_{w_1^+} = \mathrm{C}e^{\Gamma L/2c_0}e^{\Gamma(c_0 t - x)/c_0}\sin \frac{k\alpha L}{4}.$$

This property was mentioned in Ref. 23 on the basis of experiments and an equation similar to Eq. (39) was then deduced from both the continuity of the slope and of the amplitude itself. Here the theory confirms this experimental deduction showing that this continuity of the slope is a direct consequence of the continuity of stress and displacements. Nevertheless, the experiments were conducted with a finite active zone inside a finite sample in contact with water on one side and air on the other. However the gap between the edges of the active zone and the edges of the sample was far higher than the wavelength of the conjugate wave. This can explain why these experiments gave rise to a homogeneous pattern. This continuity of the amplitude slope was confirmed by numerical simulations in Ref. 17 but always with a large distance between the end of the active zone and the edge of the sample. Present theory shows that this "slope continuity" is a property of homogeneous cases and is not strictly fulfilled in more general cases. In nonhomogeneous cases, like in Fig. 3 a break of slope can be observed on the envelope of the stress signal as a function of time at the end of the pumping ($t = 19\ \mu$s), when the conjugate wave begins to evacuate the active zone.

The solution of Eq. (39) provides the value of α_s for each m. Table II gives this value for the same cases as in Ref. 17 which are in the range of the experimental values obtained with sample of ferrite of 2.5 cm length.

Formula (39) also easily confirms the value given by Bobroff and Haus[3] for m_0, the threshold of amplification in the homogeneous case. Since for $\alpha_s = \alpha_0$ it is obvious that $m = 2\alpha_0 = m_0 = 2\pi c_0/(\omega L)$. For the case $\Gamma = 0$, this is the marginal solution where the system oscillates at a constant amplitude. For $m < m_0$ real solutions of Eq. (39) do not exist, and this corresponds to the subcritical mode.

Formula (39) can be interpreted in terms of the wavelength Λ of the envelope of the spatial repartition of the waves inside the active zone, by writing

$$\frac{k\alpha}{4} = \frac{2\pi}{\Lambda}. \tag{41}$$

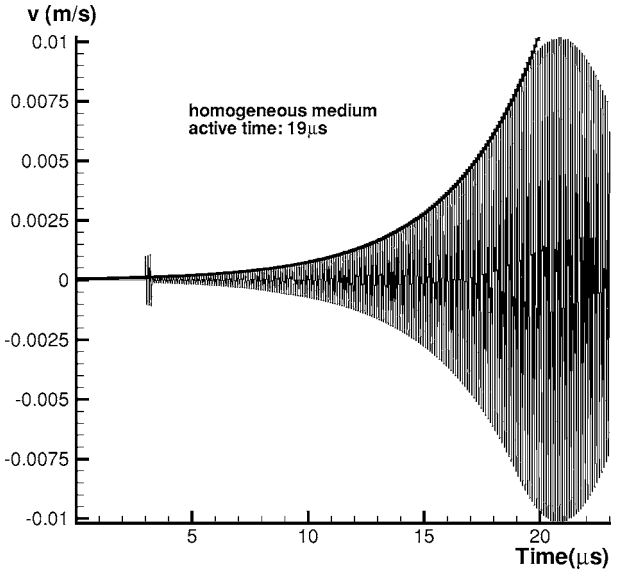


FIG. 4. Time evolution of the displacement velocity at the left edge of the active zone in the homogeneous case. $m=0.028$, $c_0=4\times10^3$ m/s, $\Omega=4\pi\times10^7$, $\rho_0=9000$ kg/m^3. The active zone is located from 0.5 to 3.0 cm. — theory, — computation.

At the threshold $m=m_0$ and $\alpha_s=\alpha_0$, the wave length Λ is just $4L$ and the envelope of w_1 and w_2 is exactly the quarter of a sinusoidal spatial period. For $\alpha_s>\alpha_0$, the value of Λ decreases as α_s increases. This means that the maximum amplitude moves progressively toward the center of the active zone. Figures 4–6 show the total agreement between the present theory and numerical solutions, computed from Eq. (5) with the same technique as in Ref. 21. In Fig. 4, the initial pulse is visible but not the second since there is no reflection on the right edge of the active zone because of the homogeneity $\tau_r=1$. The shape of the signal at the end of the pumping can be compared to the spatial distribution of w_2 (Fig. 6). Comparison between Figs. 5 and 6 shows that the maximum value is the same for w_1 and w_2 as stated in the theory to fulfill the invariance property of systems (17) and (18).

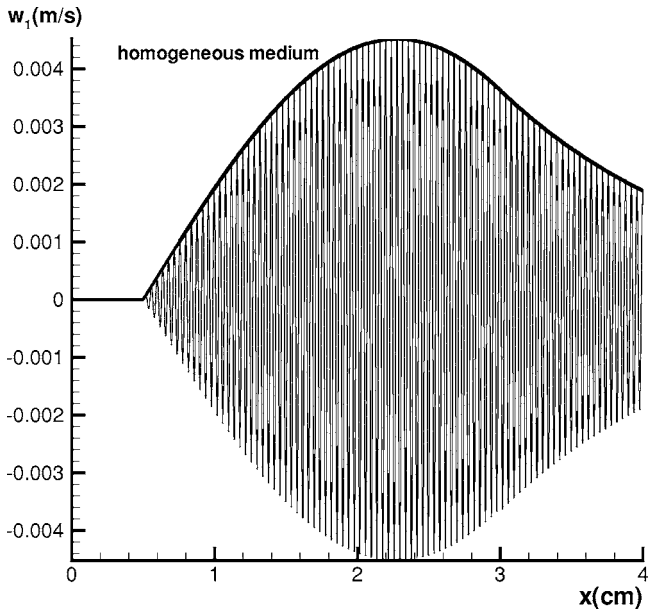


FIG. 5. Spatial distribution of the direct wave w_1, at $t=13.5$ μs in the homogeneous case and for the same conditions as in Fig. 4. The wave does not radiate toward the left of the active zone. — theory $\{Ce^{\Gamma t}\sin[k\alpha(x+L/2)/4]\}$, — computation.

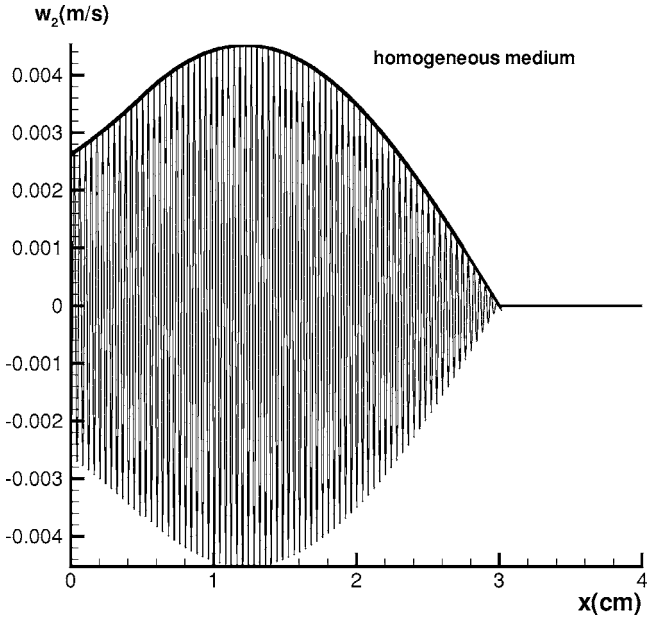


FIG. 6. Spatial distribution of the conjugate wave w_2, at $t=13.5$ μs, in the homogeneous case and for the same conditions as in Fig. 4. The wave does not radiate toward the right of the active zone. — theory $\{Ce^{\Gamma t}\sin[k\alpha(x-L/2)/4]\}$, — computation.

Since Eq. (39) has an asymptotic value for $\alpha_s\to2\alpha_0$ where $m\to\infty$, all physical situations verify $\alpha_s<2\alpha_0$ and $\Lambda<2L$. Therefore, if m is high enough, $\Lambda\to2L$. Hence, in the case of high energy pumping, the maximum amplitude of w_1 and w_2 tends toward the middle of the active zone and a node appears at the edges. Consequently, the signal becomes small on the edge of the active zone and very little energy is radiated outside (Fig. 7) before the end of the pumping. This implies that, for strong excitation (m large), energy accumulates in the sample during the pumping and leads to heat production in the material itself. The energetic efficiency of the process falls very quickly as the excitation increases. Consequently the use of ultrasonic phase conjugation for high energy applications like hyperthermia must be carefully studied through an optimization procedure. Of course Fig. 7 is not realistic but is only an extrapolation of the linear pumping theory for high energy transfers in order to confirm the theoretical predictions on the shape of the spatial repartition under intense pumping. In reality, nonlinear processes take place before the amplification reach such high values (saturation by feedback to the electric circuit,[18] dissipation, plasticity, etc.). Because of these phenomena, very little energy (relative to the pump energy) can evacuate the active zone in the conjugate wave w_2 even after the end of the pumping.

The homogeneous case allows a very straightforward analysis and gives the major tendencies of the phenomenon despite the fact that the model above is purely one dimensional. Of course it still leaves open many possibilities for optimization either by exploiting the geometry of the sample

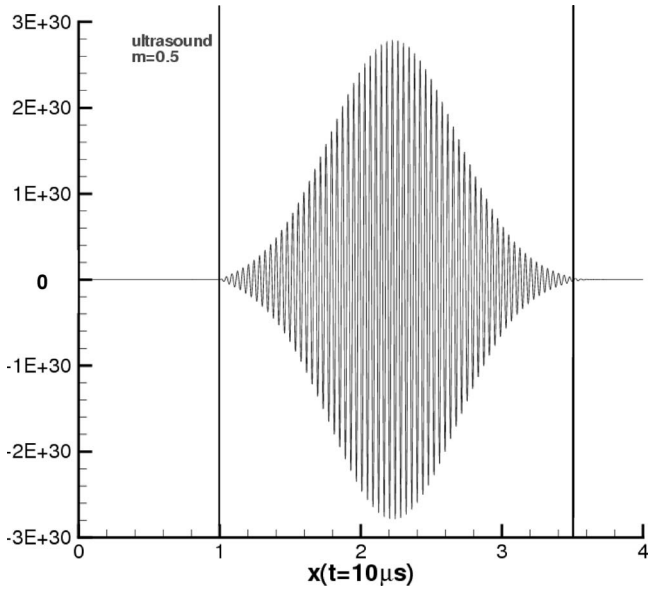


FIG. 7. Spatial distribution of the direct wave w_1 in the homogeneous case for the same condition as in Fig. 5 but with $m=0.5$. After 19 μs, the values are extremely high because of the exponential growth. This simulation does not pretend to represent a realistic situation but is only an extrapolation of the linear pumping theory to high energy transfers. The theoretical predictions are confirmed: There is a node of amplitude near the exit of the active zone and the energy efficiency of the wave emission remains very low during the pumping. The same figure can be drawn for the conjugate wave.

to three dimensions or by modifying the boundary condition of the problem, i.e., by varying the impedance ratio at the edge of the active zone. Unlike Bobroff solutions, the present dual-time-scale method allows more progress in this direction. As an illustration of that, the following section treats the case of an impedance jump at one edge of the active zone.

C. Semihomogeneous medium

Here the active zone of length L is in contact on the left with a medium of different acoustic impedance. On the right of the active zone, the medium remains homogeneous. This case which was not solved in the literature can be treated by letting $\tau_r=1$ in the general expression. As in the homogeneous case, $\xi_2=-k\alpha L/8$ but ξ_1 can be written

$$\xi_1 = \frac{k\alpha L}{8} + n\pi + (-1)^n \arcsin\left(\frac{\tau_l - 1}{\tau_l + 1}\sin\frac{k\alpha L}{4}\right) \tag{42}$$

with $n \in \mathbb{N}$. Equation (34) becomes

TABLE III. α_s for the semihomogeneous case.

Left material	Active and right material	τ_l	α_s
Air	Ferrite	0	0.006 724
Water	Ferrite	0.0412	0.007 333
Glass	Ferrite	0.4073	0.009 935
Steel	Ferrite	1.2996	0.010 897

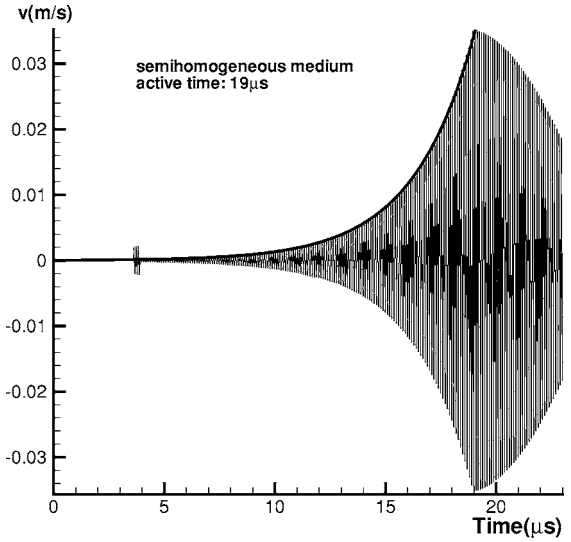


FIG. 8. Time evolution of the displacement velocity at the left edge of the active zone in the semihomogeneous case with $\tau_l=0.042$. $m=0.028$, $c_0=4\times10^3$ m/s, $\Omega=4\pi\times10^7$, $\rho_0=9000$ kg/m^3. The active zone is located from 0.5 to 3.0 cm. — theory $[\mathrm{Ce}^{\Gamma t}\sin(-k\alpha L/4)]$, — computation.

$$G_{sh}(\alpha) = \alpha\cot\left[\frac{k\alpha L}{4} \pm \arcsin\left(\frac{1-\tau_l}{1+\tau_l}\sin\frac{k\alpha L}{4}\right)\right] + \frac{\sqrt{m^2-(2\alpha)^2}}{2} = 0. \tag{43}$$

This form is more practical than the general one for specifically studying the influence of the impedance ratio, at one edge, on the energy transmission between the sample and the external medium. This is not the aim of the present article but as examples, Table III gives a few values of α for a realistic set of parameters for $\Omega=4\pi\times10^7$, $L=2.5$ cm. Figures 8–10 show a total agreement between the theory and

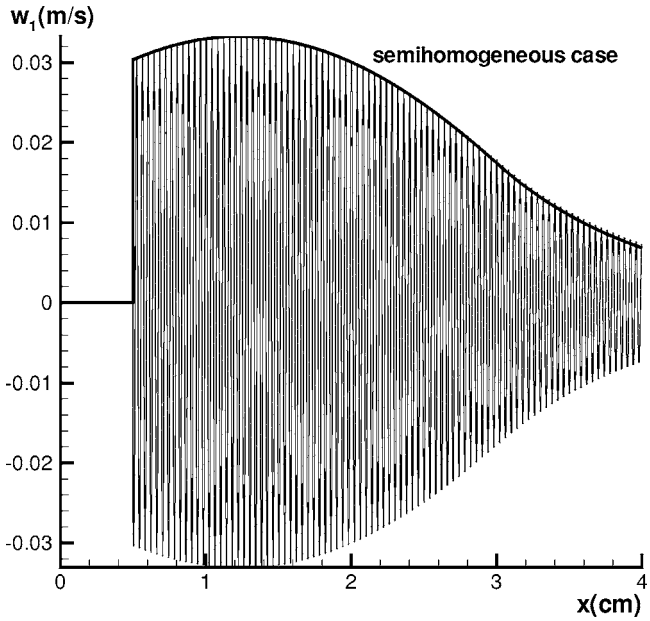


FIG. 9. Spatial distribution of the direct wave w_1, at the end of the pumping, in the semihomogeneous case and for the same conditions as in Fig. 8. The wave does not radiate toward the left of the active zone. — theory $\{-\mathrm{Ce}^{\Gamma t}\sin[k\alpha x/4+\xi_1]\}$, — computation.

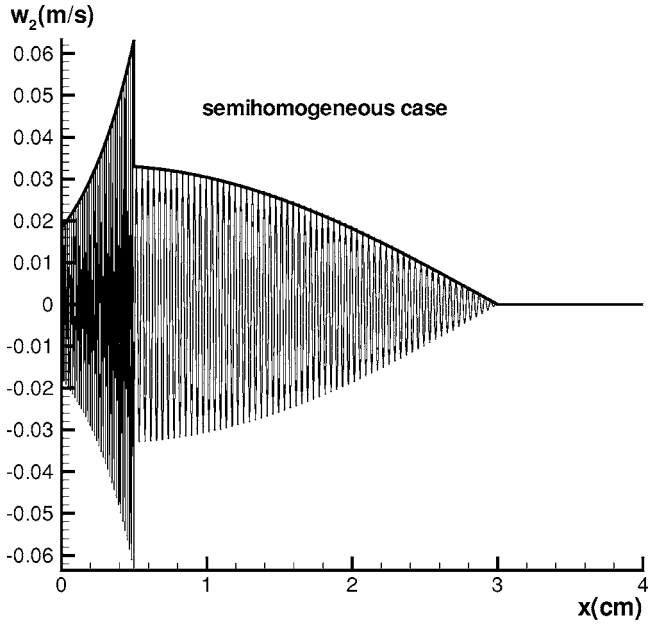


FIG. 10. Spatial distribution of the conjugate wave w_2, at the end of the pumping, in the homogeneous case and for the same conditions as in Fig. 4. The wave does not radiate toward the right of the active zone. — theory $\{Ce^{\Gamma t}\sin[k\alpha(x-L/2)/4]\}$, — computation.

the computation. Again the initial pulse is visible (Fig. 8) but no second pulse because the homogeneity at the right edge avoids any reflection. The shape of the signal after the end of the pumping can be compared with the spatial distribution of w_2 in the active zone (Fig. 10). It can be observed that this spatial repartition is strongly changed with respect to the homogeneous case (Fig. 6). This shows how much the boundary conditions at the edges can affect the phenomenon. The impedance jump at $x=0.5$ cm results in a jump of the value of w_2, with the consequence of a discontinuity in θ and in its amplitude. This impedance jump is also a consequence of the modification in wavelength due to the change of sound velocity and of the stress continuity. Comparison between Figs. 5 and 9 shows that the impedance jump also modifies the direct waves through the condition $w_{1l}(-L/2,t)=0$ which gives $w_1(-L/2,t)\neq 0$ when $\tau_l \neq 0$.

IV. CONCLUSION

The present analytical results provide a very accurate description of the phenomena happening inside a magnetoacoustic phase conjugator. The one-dimensional assumption allows the derivation of the solutions with standard mathematical tools, once the general forms have been identified by a dual-time-scale technique. Of course, this hypothesis limits the discussion to situations where most of the incident rays cross the active zone from edge to edge without reflecting on the lateral boundaries of the conjugator, but this is the case for most of the practical cases where the conjugator is far from the emitting source. This is why the present results can be used for the optimization of real processes. Moreover this theory is an unavoidable basis for any generalization among which the most important is the case of a nonhomogeneous pumping (i.e., m depending on x). One case of particular interest is the partially active conjugator with an active zone not directly in contact with the external media. Finally, it must be underlined that WPC is not specific to magnetoacoustics and that many phenomena are ruled by the same equations, hence, the present theory can be adapted, for example, to shallow water equations, optics or microwaves.

[1] N. M. Kroll, "Excitation of hypersonic vibrations by means of photoelastic coupling of high-intensity light waves to elastic waves," J. Appl. Phys. **36**(1), 34–43 (1965).
[2] D. L. Bobroff, "Coupled-modes analysis of the phonon-photon parametric backward-wave oscillator," J. Appl. Phys. **36**(5), 1760–1769 (1965).
[3] D. L. Bobroff and H. A. Haus, "Impulse response of active coupled wave systems," J. Appl. Phys. **38**, 390–403 (1967).
[4] B. A. Auld and H. Mathews, "Parametric traveling-wave acoustic amplification in ferromagnets," J. Appl. Phys. **36**(11), 3599–3605 (1965).
[5] R. B. Thompson and C. F. Quate, "Nonlinear interaction of microwave electric fields and sound in $LiNbO_3$," J. Appl. Phys. **42**(3), 907–919 (1971).
[6] A. A. Chaban, "On one nonlinear effect in piezoelectric semiconductors," Sov. Phys. Solid State **9**, 3334–3335 (1967).
[7] H. J. Simpson, P. L. Marston, "Parametric Layers, Four-wave mixing and wave front reversal," *Non Linear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), pp. 399–420.
[8] F. V. Bunkin, D. V. Vlasov, and Yu. A. Kravtsov, "On problem of sound wave phase conjugation with amplification of conjugate wave," Sov. J. Quantum Electron. **11**, 687–688 (1981).
[9] F. V. Bunkin, D. V. Vlasov, E. A. Zabolotskaya, and Yu. A. Kravtsov, "Wave-front inversion of sound beams and other nonlinear effects in a medium containing bubbles," Sov. Phys. Acoust. **28**, 249 (1982).
[10] L. M. Kustov, V. E. Nazarov, and A. M. Sutin, "Nonlinear sound scattering by a bubble layer," Sov. Phys. Acoust. **32**(6), 500–503 (1986).
[11] A. P. Brysev, L. M. Krutyianskii, and V. L. Preobrazhenskii, "Wave phase conjugation of ultrasonic beams," Sov. Phys. Usp. **41**(8), 793–805 (1998).
[12] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J. L. Thomas, and F. Wu, "Time-reversed acoustics," Rep. Prog. Phys. **63**, 1933–1995 (2000).
[13] A. Brysev, L. Krutyansky, P. Pernod, and V. Preobrazhensky, "Acoustic microscope based on magneto-elastic wave phase conjugator," Appl. Phys. Lett. **76**(21), 3133–3135 (2000).
[14] A. E. Siegman, *Lasers* (University Science Books, Mill Valley, CA, 1986), pp. 944–947.
[15] M. W. Bowers and R. W. Boyd, "Phase locking via Brillouin-enhanced four-wave-mixing phase conjugation," IEEE J. Quantum Electron. **34**(4), 634–644 (1998).
[16] M. Tsang and D. Psaltis, "Spectral phase conjugation with cross-phase modulation compensation," Opt. Express **12**(10), 2207–2219 (2004).
[17] S. Ben Khelil, A. Merlen, V. Preobrazhensky, and P. Pernod, "Numerical simulation of acoustic wave phase conjugation in active media," J. Acoust. Soc. Am. **109**(1), 75–83 (2001).
[18] A. Merlen, V. L. Preobrazhenskii, and P. Pernod, "Supercritical parametric phase conjugation of ultrasound. Numerical simulation of non linear and non stationary modes," J. Acoust. Soc. Am. **112**, 2656–2665 (2002).
[19] K. Yamamoto, P. Pernod, and V. Preobrazhensky, "Visualization of phase conjugate ultrasound waves passed through inhomogeneous layer," Ultrasonics **42**, 1049–1052 (2004).
[20] P. Voinovich, A. Merlen, E. Timofeev, and K. Takayama, "A Godunov-type finite-volume scheme for unified solid-liquid elastodynamics on arbitrary two-dimensional grids," Shock Waves **13**, 221–230 (2003).
[21] P. Voinovich and A. Merlen, "2-D numerical simulation of acoustic wave phase conjugation in active medium," J. Acoust. Soc. Am. **118**(6), 3491–3498 (2005).
[22] F. Verhulst, "*Nonlinear Differential Equations and Dynamical Systems*" (Universitext Springer-Verlag, Berlin, 1996), pp. 152–155.
[23] P. Pernod and V. Preobrazhensky, "Parametric phase conjugation of a wide-band acoustic pulse in supercritical mode," Appl. Phys. Lett. **76**(3), 387–389 (2000).

Identification of multipole noise sources in low Mach number jets near the peak frequency

Takao Suzuki
Graduate School of Engineering, University of Fukui, Fukui, Japan

(Received 27 June 2005; revised 30 March 2006; accepted 4 April 2006)

A beam-forming algorithm is developed to identify noise source structures in low Mach number jets near the peak frequency. Pressure data from a conical phased microphone array are Fourier decomposed in the azimuthal direction, and the dominant eigenvectors of a cross-spectral matrix are extracted for each azimuthal mode; subsequently, a robust adaptive beam-forming technique is applied to a cross-spectral matrix reconstructed from the primary eigenmodes. The coherent sound signals are then decomposed into multipoles in spherical coordinates, and the primary mode and its axial position are identified for the axisymmetric and first two azimuthal modes. The proposed beam-forming algorithm is applied to an unheated round jet over a range of Mach numbers and frequencies. The result indicates that at low Mach numbers, the second zenithal mode (i.e., quadrupole) appears to be dominant for each azimuthal mode, and the source position slightly shifts downstream with increasing azimuthal mode number. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2200070]

PACS number(s): 43.28.Ra, 43.28.Tc, 43.60.Fg [DRD] Pages: 3649–3659

I. INTRODUCTION

To understand noise generation mechanisms, a variety of features for subsonic jets have been investigated for many years. Specifically, far-field sound radiation patterns have been studied over a wide range of flow conditions, and various models for jet noise sources have been proposed to predict far-field sound. It has been implied by several studies that noise near the peak frequency, particularly coherent noise, is associated with large-scale coherent structures and generated near the end of the potential core.[1–3] In early experimental studies, approaches using an acoustic mirror[4] or multiple microphones[5–7] have been introduced to characterize noise source structures. However, it is difficult to extract useful information other than the approximate source position from these experimental results, because no decisive source model is available for subsonic jets.

On the other hand, phased microphone array techniques, i.e., beam-forming, which had been used for radars[8,9] and sonars,[10] have become increasing popular in aeroacoustics. The applications include not only stationary and compact sources, such as airframe noise,[11,12] but also noncompact ones, such as fan noise[13,14] and jet instability waves.[15] Several researchers have also applied advanced beam-forming algorithms to distributed sources, such as, trailing edge noise[16] and jet noise.[13,17,18] Yet, the source models used in these studies were mostly based on monopole sources, and there is still controversy about the interpretation, especially of jet noise source maps. Although Ghosh *et al.*[19] analyzed the sound radiation patterns as a series of multipoles, they focused on the over all directivity. Thus, the radiation patterns of coherent sound signals and their source positions remain unexploited.

The objective of this study is to develop a beam-forming algorithm that can identify subsonic jet-noise structures and source positions. This study focuses on primary coherent sound signals near the peak frequency ($St \equiv fD/U_{\text{jet}} \approx 0.35-0.80$, where f, D, and U_{jet} denote the source frequency, the jet diameter, and the jet velocity, respectively), which are most likely related to the large-scale flow structures. Using a conical-phased microphone array in the far field, sound signals are Fourier decomposed in the azimuthal direction, and the dominant eigenvectors of a cross-spectral matrix are extracted for each azimuthal mode; subsequently, a robust adaptive beam-forming technique[20] is applied to a cross-spectral matrix, reconstructed from these dominant eigenmodes. Using multipoles in spherical coordinates as reference solutions, the primary mode and its axial position can be identified. This beam-forming algorithm is then applied to an unheated subsonic round jet ($M_\infty \equiv U_{\text{jet}}/a_\infty = 0.35-0.70$, where a_∞ denotes the ambient speed of sound), and the axisymmetric and first two azimuthal modes are investigated.

The outline of the rest of the paper is as follows: The conventional and adaptive beam-forming algorithms are first reviewed, and multipoles in spherical coordinates are summarized. These components are then integrated, and a beam-forming algorithm developed for this study is introduced. In the following section, the experimental facilities including the microphone array are described. The proposed algorithm is validated based on model simulations, and the results for subsonic round jets are discussed.

II. BEAM-FORMING ALGORITHMS

A. Conventional beam-forming

The idea of the conventional beam-forming is based on a least-square minimization between measured data and a reference solution. Suppose a monopole source at an angular frequency of ω is located at a fixed position, $\mathbf{x}_s$. Its pressure amplitude at a point $\mathbf{x}$ can then be represented as

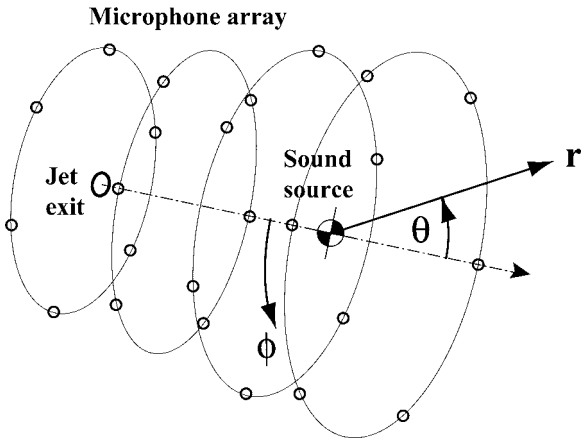


FIG. 1. Coordinate system for spherical multipoles. For the actual microphone distribution, refer to Fig. 4.

$$\frac{a_s(\omega)}{4\pi|\mathbf{x}-\mathbf{x}_s|}e^{ik|\mathbf{x}-\mathbf{x}_s|} \equiv a_s(\omega)\hat{p}(\mathbf{x}_s;\mathbf{x}), \tag{1}$$

where k ($\equiv\omega/a_\infty$) denotes the wave number, a_s is the complex source amplitude including an arbitrary phase shift, and $\hat{p}$ the reference solution, which is considered to be a monopole in this example (a factor of 1/2 may be needed unless the imaginary part of the pressure signals is compensated).

On the other hand, when the pressure data are available at N_{mic} microphones, we express the pressure amplitude at the lth microphone as $\hat{q}_l(\omega)$ and define the following cost function:

$$J(\omega,a_s,\mathbf{x}_s) \equiv \sum_{l=1}^{N_{\mathrm{mic}}} |a_s(\omega)\hat{p}(\mathbf{x}_s;\mathbf{x}_l) - \hat{q}_l(\omega)|^2, \tag{2}$$

where $\mathbf{x}_l$ denotes the lth microphone position. By differentiating Eq. (2)with respect to a_s, we find that minimizing Eq. (2)is equivalent to maximizing the following output:[15]

$$\tilde{P}_{\mathrm{conv}}(\omega,\mathbf{x}_s) \equiv |\mathbf{p}|^2 P_{\mathrm{conv}}(\omega,\mathbf{x}_s) \equiv \frac{\mathbf{p}^\dagger}{|\mathbf{p}|}\cdot\mathbf{q}\mathbf{q}^\dagger\cdot\frac{\mathbf{p}}{|\mathbf{p}|}, \tag{3}$$

where $\mathbf{q}\equiv\hat{q}_l(\omega)$ and $\mathbf{p}\equiv\hat{p}(\mathbf{x}_s;\mathbf{x}_l)$, and † denotes the complex conjugate transpose. Thus, the problem yields phase matching between the cross-spectral matrix, $\mathbf{q}\mathbf{q}^\dagger$, and the steering vector, $\mathbf{p}/|\mathbf{p}|$. At the position, $\mathbf{x}_s$, where the cost function is minimized, $\tilde{P}_{\mathrm{conv}}(\omega,\mathbf{x}_s)$ becomes maximum and indicates the source position. We normalize the output so that $P_{\mathrm{conv}}(\omega,\mathbf{x}_s)$ corresponds to the source intensity, $|a_s|^2$, at the optimal point.

When the pressure signal is composed of N_{src} stationary uncorrelated signals and random noise, it may be written as

$$\mathbf{q} \equiv \sum_{l'=1}^{N_{\mathrm{src}}} \alpha_{l'}\mathbf{v}_{l'} + \mathbf{r}, \tag{4}$$

where the vector corresponding to the l'th signal is normalized as $|\mathbf{v}_{l'}|=1$ for convenience and $\mathbf{r}$ denotes the noise components. To suppress uncorrelated noise between microphones, the cross-spectral matrix is generally constructed by averaging a number of samples; namely,

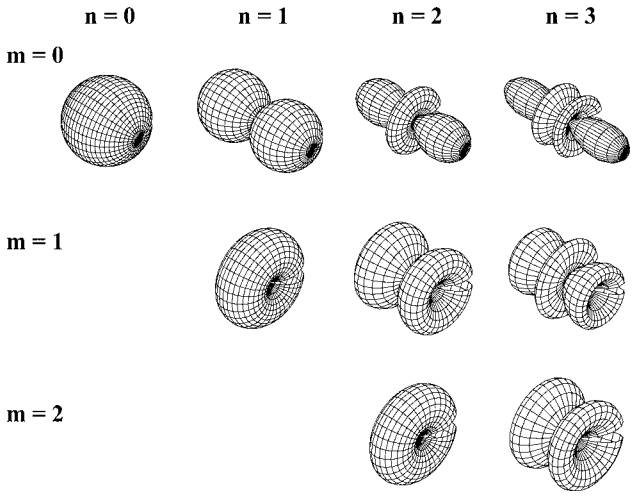


FIG. 2. Radiation patterns of spherical multipoles. Absolute values of amplitude are illustrated as polar plots.

$$\overline{\mathbf{q}\mathbf{q}^\dagger} \approx \sum_{l'=1}^{N_{\mathrm{src}}} \overline{|\alpha_{l'}|^2}\mathbf{v}_{l'}\mathbf{v}_{l'}^\dagger + \sigma^2\mathbf{I}, \tag{5}$$

where the overline denotes the average over samples and $\mathbf{I}$ is the identity matrix. Here, the noise components are assumed identical and uncorrelated between microphones, i.e., $\overline{r_i r_j} = \sigma^2\delta_{ij}$.

Because Eq. (5)is positive definite and Hermitian, it can be diagonalized with a unitary matrix:

$$\overline{\mathbf{q}\mathbf{q}^\dagger} = \sum_{l=1}^{N_{\mathrm{mic}}} \lambda_l\mathbf{u}_l\mathbf{u}_l^\dagger = \mathbf{U}\Lambda\mathbf{U}^\dagger, \tag{6}$$

where $\lambda_l>0$ and $|\mathbf{u}_l|=1$. Thus, we can extract correlated signals regardless of the radiation pattern by calculating an eigenvalue problem of the cross-spectral matrix. In particular, when the primary signal has by far the greatest amplitude (say, α_1), the eigenvector with the greatest eigenvalue (say, λ_1) is nearly identical to the primary signal (i.e., $|\mathbf{u}_1|\approx|\mathbf{v}_1|$). If the signal actually coincides with the reference solution, the primary eigenvalue yields $\lambda_1\approx|a_s|^2|\mathbf{p}|^2$.

The output of the conventional beam-forming, thereby, consists of projections of the reference solution onto the eigenvectors. Reference solutions steered toward incorrect directions or representing radiation patterns different from the actual source (e.g., higher multipoles) often have relatively large projections onto the primary eigenvector and provide sidelobes of the order of $O(|a_s|^2)$. These make source maps misleading when multipoles radiate the sound or multiple sources coexist.

B. Adaptive beam-forming

An adaptive beam-forming technique of the minimum variance type is one of the effective methods to suppress sidelobes.[21,22] We wish to find the weight function that minimizes the output defined as $\mathbf{w}^\dagger\overline{\mathbf{q}\mathbf{q}^\dagger}\mathbf{w}$, subject to the unit response, i.e., $\mathbf{w}^\dagger\mathbf{p}=1$. The solution can be calculated as

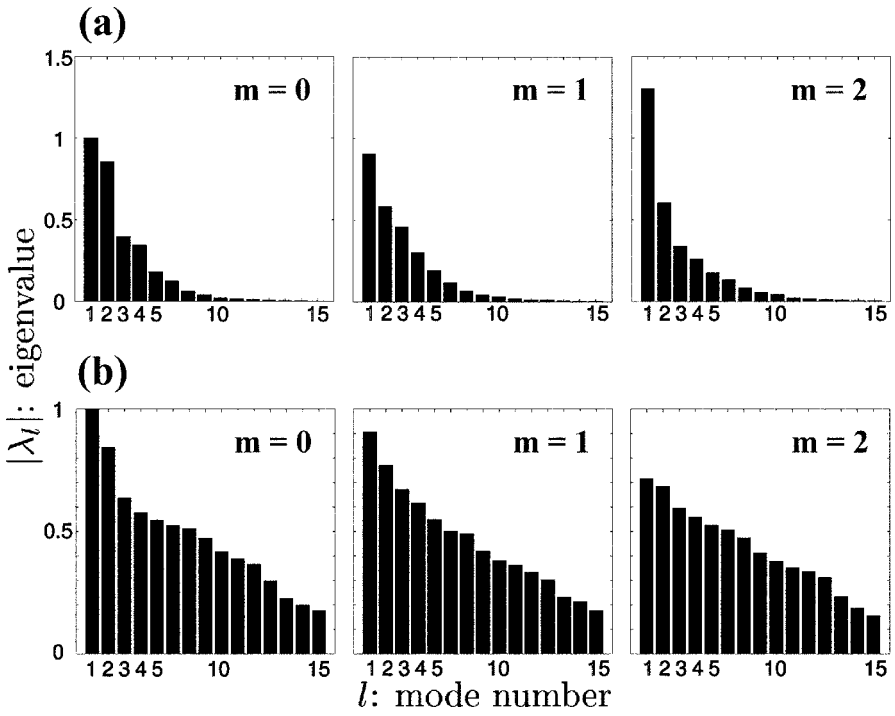


FIG. 3. Eigenvalues of a cross-spectral matrix (M_∞=0.50): (a) St =0.35 and (b) St=1.50. The greatest eigenvalue for m=0 is normalized as unity at each Strouhal number. (Detailed data processing is described in Sec. III C).

$$\mathbf{w} = \frac{(\overline{\mathbf{qq}^\dagger})^{-1}\mathbf{p}}{\mathbf{p}^\dagger(\overline{\mathbf{qq}^\dagger})^{-1}\mathbf{p}}, \tag{7}$$

and the resultant output can be expressed using Eq. (6)as

$$P_{\text{adap}}(\omega,\mathbf{x}_s) = \frac{1}{\mathbf{p}^\dagger(\overline{\mathbf{qq}^\dagger})^{-1}\mathbf{p}} = \frac{1}{\mathbf{p}^\dagger\mathbf{U}\Lambda^{-1}\mathbf{U}^\dagger\mathbf{p}}. \tag{8}$$

If the amplitude of the primary signal is by far the greatest, $P_{\text{adap}}(\omega,\mathbf{x}_s)$ similarly recovers $|a_s|^2$.

We notice that the projections of the reference solution onto the eigenvectors are weighted with the inverse of the eigenvalues in the denominator of Eq. (8). This allows the adaptive beam-forming to identify each source position exclusively, even when uncorrelated multiple sources coexist. Suppose that in the measured pressure given by Eq. (4), all $\mathbf{v}_{l'}$ are linearly independent, and the orders of $\alpha_{l'}$ are comparable (denoted by $\sim\alpha$) but much larger than σ. The first N_{src} eigenvalues are then on the order of α^2 and the rest are of σ^2. When we calculate an output with a reference solution steered toward an incorrect direction or for an incorrect radiation pattern, its vector typically has a component outside the space spanned by the N_{src} signals. As a result, the output yields $O((\sigma/\alpha)^2|a_s|^2)$, and sidelobes generally disappear. This technique is effective even when we detect a single source.

As the signal to noise ratio increases, the eigenvectors associated with the noise components dominate $\mathbf{U}\Lambda^{-1}\mathbf{U}^\dagger$ in Eq. (8); thus, its limit is analogous to a technique, so-called *MUSIC*.[23] One of the problems for adaptive beam-forming methods is the intolerance against inaccuracy of the reference solution, when the signal-to-noise ratio is too high. We can circumvent this difficulty by using a robust adaptive beam-forming technique,[20] which will be explained in Sec. II D.

C. Multipole decomposition

Since the sound signals can be decomposed in the azimuthal direction with the array used in this study, it is convenient to express noise sources as a superposition of multipoles in spherical coordinates. To be precise, we implicitly assume that these modes are uncorrelated with each other. Assuming that the refraction effect is weak at low Mach numbers, we consider the Helmholtz equation:

$$\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right]\hat{p} + k^2r^2\hat{p} = 0. \tag{9}$$

The coordinate system is illustrated in Fig. 1. Assuming that the transformed pressure is variable separable, i.e., $\hat{p} \sim R(r)\Theta(\theta)\Phi(\phi)$, the radial dependence of outgoing waves can be expressed as[24]

$$R(r) = r^{1/2}H^{(1)}_{n+1/2} \to \sqrt{\frac{2}{\pi}}\frac{e^{i(kr-n\pi/2-\pi/2)}}{r} \quad \text{as} \quad r \to \infty, \tag{10}$$

where $H^{(1)}_{n+1/2}$ denotes the Hankel function of the first kind and n corresponds to the zenithal mode number, which takes a non-negative integer. In this study, we only consider the leading terms of the far-field asymptote (i.e., $r \gg \lambda$).

Furthermore, if we decompose the azimuthal mode as $\Phi(\phi)=e^{\pm im\phi}$, the zenithal dependence yields the associated Legendre function:[24]

$$\Theta(\theta) = P^m_n(\cos\theta). \tag{11}$$

Here, P^m_n takes a nonzero solution only when $n \geq |m|$.

We now normalize the coefficient of each mode so that the sound intensity over the unit sphere is 4π and consider up to n=3 mode. The radiation patterns of these multipoles

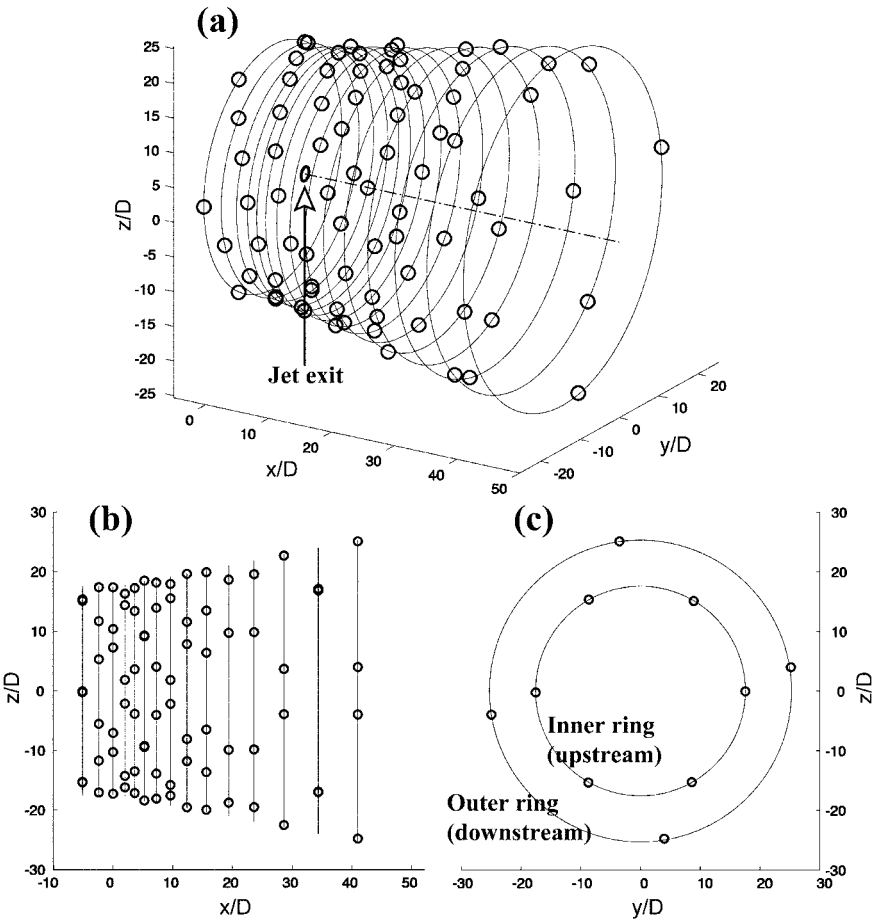


FIG. 4. Diagram of the microphone distribution. The center of the jet exit is taken to be the origin: (a) Three-dimensional view, (b) side view, and (c) front view.

are shown in Fig. 2. These representations are used as the reference solutions for the algorithm developed in this study.

It should be mentioned that these multipoles are mathematically orthogonal to each other; namely, $\int_0^{2\pi}\int_0^{\pi} P_{n'}^{m'}(\cos\theta)e^{-im'\phi}P_n^m(\cos\theta)e^{im\phi}\sin\theta d\theta d\phi \neq 0$, if and only if $n'=n$ and $m'=m$. In fact, when microphones are axisymmetrically distributed, eigenvectors associated with two different azimuthal modes (which are, of course, assumed to be dealiased) are orthogonal to each other. However, in general microphone distributions, the steering vectors corresponding to these multipoles lack this property. For example, vectors for two different zenithal modes calculated based on a conical array are generally nonorthogonal. To take this point into account, a composite cross spectral matrix will be introduced in the next section.

D. Proposed beam-forming algorithm

In contrast to the broadband noise at high frequencies, many studies have concluded that the sound generation mechanism near the peak frequency is associated with large-scale coherent structures in the mixing layer. In fact, experimental results of this study ($M_\infty=0.50$) in Fig. 3 demonstrate that the primary eigenvalues of the cross-spectral matrix at St=0.35 occupy a larger portion of the total intensity compared with those at St=1.50. Therefore, this study focuses on dominant coherent signals for each azimuthal mode and develops an algorithm that can detect their source positions.

First, dominant coherent signals are extracted from a cross-spectral matrix. In the microphone array for the present study, mostly six (at least four) microphones are axisymmetrically distributed on each ring (see Fig. 4); therefore, pressure signals up to $m=\pm 2$ are orthogonal to each other. Taking a Fourier transform in the azimuthal direction, we construct $N_{\text{ring}}\times N_{\text{ring}}$ cross-spectral matrices (N_{ring} denotes the number of microphone rings in the axial direction) and extract the eigenvectors for each azimuthal mode ($m=0$, 1, and 2).

Next, to represent the primary coherent signal, we reconstruct a cross-spectral matrix with a series of eigenvectors as follows:

$$\overline{\mathbf{q}\mathbf{q}^{\dagger}} \approx \left(1-\sum_{l=2}^{N_{\text{ring}}}\beta_l\right)\lambda_1\mathbf{u}_1\mathbf{u}_1^{\dagger}+\beta_2\lambda_2\mathbf{u}_2\mathbf{u}_2^{\dagger}+\cdots, \tag{12}$$

where $0\leq\beta_l\leq 1$. We wish to identify the source corresponding to the eigenvector with the greatest eigenvalue (i.e., λ_1); however, as mentioned in Sec. II A, the first eigenvector does not necessarily coincide with the radiation pattern of the primary source. Hence, the high-order terms are superposed in order to include the dimensions outside the first eigenvector. In this study, only the second term is included for simplicity. The procedure for determining β_2 (which will be denoted as β for the rest of this paper) will be explained at the end of this section.

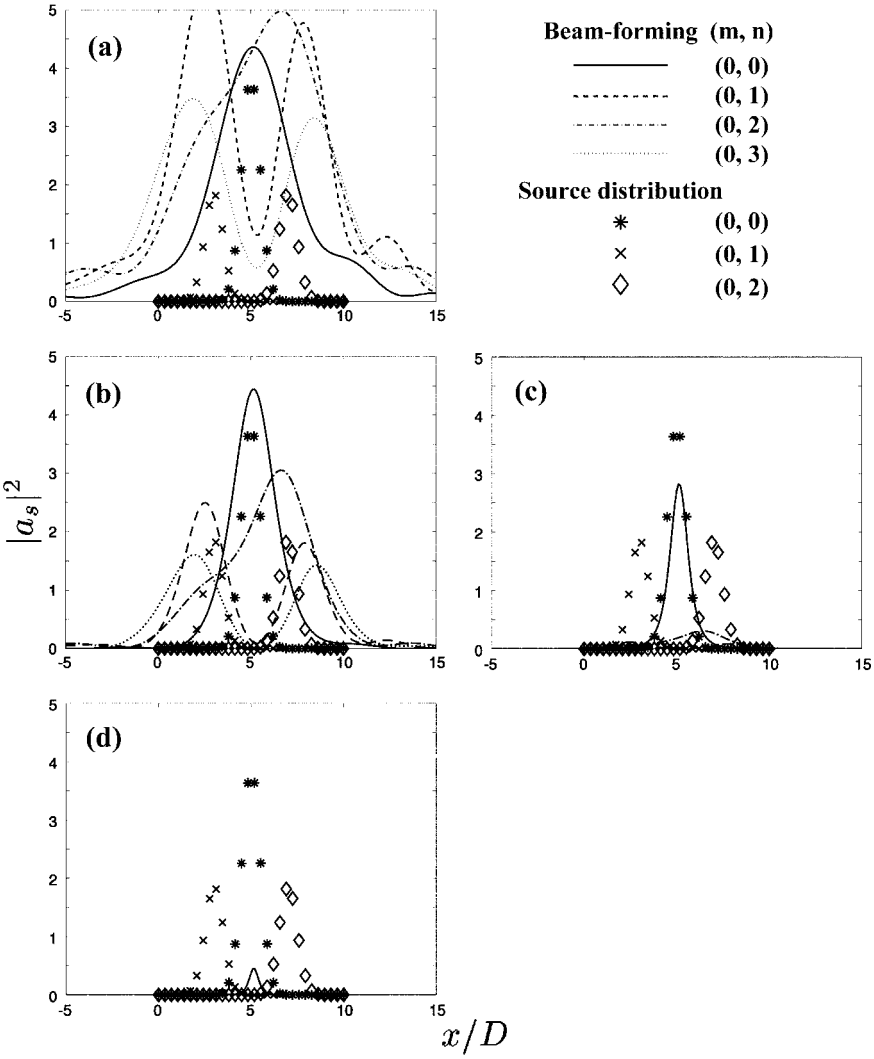


FIG. 5. Source maps of a model problem. The outputs of the beam-forming are compared with the source distributions: (a) conventional beam-forming, (b) proposed beam-forming with $\epsilon = 1.0\max[\lambda_l]$, (c) that with $\epsilon = 0.1\max[\lambda_l]$, and (d) that with $\epsilon = 0.01\max[\lambda_l]$.

The matrix above does not have full rank. To invert it and apply an adaptive beam-forming algorithm, we introduce a technique, called *robust adaptive beam-forming*[20] which uses the minimum variance estimator. Adding an artificial diagonal part on Eq. (12), we set the weight function as

$$\mathbf{w} = \frac{[(1-\beta)\lambda_1\mathbf{u}_1\mathbf{u}_1^\dagger + \beta\lambda_2\mathbf{u}_2\mathbf{u}_2^\dagger + \epsilon\mathbf{I}]^{-1}\mathbf{p}}{\mathbf{p}^\dagger[(1-\beta)\lambda_1\mathbf{u}_1\mathbf{u}_1^\dagger + \beta\lambda_2\mathbf{u}_2\mathbf{u}_2^\dagger + \epsilon\mathbf{I}]^{-1}\mathbf{p}}. \tag{13}$$

Suitable values for ϵ will be examined in Sec. IV. Consequently, the output of the proposed algorithm can be expressed as

$$P_{\text{prop}}(\omega, \mathbf{x}_s) = \frac{1}{\mathbf{p}^\dagger[(1-\beta)\lambda_1\mathbf{u}_1\mathbf{u}_1^\dagger + \beta\lambda_2\mathbf{u}_2\mathbf{u}_2^\dagger + \epsilon\mathbf{I}]^{-1}\mathbf{p}}. \tag{14}$$

For the reference solution $\mathbf{p}$, we use the multipoles in spherical coordinates illustrated in Fig. 2.

The idea of extracting primary eigenmodes is similar to *MUSIC*. However, radiation patterns of the dominant signals are unknown for the jet noise and may be quite different from the spherical multipoles; hence, source maps generated based on *MUSIC* exhibit many spurious peaks. If we apply the adaptive beam-forming algorithm to the whole cross spectral matrix, the output is not focused sufficiently so that the source maps are less informative to identify the source positions, as discussed in Sec. IV B.

For most cases in this study, source maps generated only from the first eigenmode (i.e., $\beta=0$) depict a single distinctive peak. In some source maps, however, peaks of multiple zenithal modes become comparable or an output of a single mode shows multiple peaks. As mentioned in Sec. IV A, it is possible that the second or higher eigenvectors cause the first one to deviate from the primary coherent signal and make the source map misleading. Hence, we cannot expect the exact one-to-one correspondence between noise sources and eigenvectors, especially for jet noise. Assuming that the actual source distribution has only a single distinctive peak for each azimuthal mode, we optimize β if the ratio of the first to the second highest peaks on the source map increases with increasing β. We search the second peak among all the zenithal modes including the one that has the highest peak (as seen later, the optimal β is found to be 0 for most cases). Thus, the source position of the primary coherent signals can be exclusively identified on the resultant source map for each azimuthal mode. A model problem and an actual example regarding non-zero β are presented in Sec. IV.

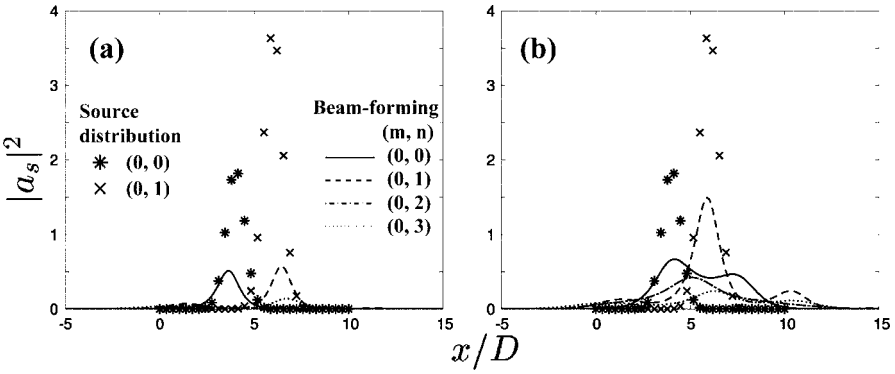


FIG. 6. Comparison of the source maps of a model problem: (a) $\beta=0$ and (b) optimal $\beta(=0.023)$. Notation is the same as Fig. 5.

III. EXPERIMENTAL FACILITIES

A. Jet acoustic rig

Experiments were conducted using the Small Hot Jet Acoustic Rig at NASA Glenn Research Center. A single-stream round jet was mounted at 3.05 m from the ground in an anechoic dome with a 20 m radius. The jet exit was a converging nozzle, and the exit diameter was 5.08 cm (2.0 in.). The total temperature was the same as the room temperature. Among various flow conditions, three Mach numbers ($M_\infty \equiv U_{\text{jet}}/a_\infty = 0.35, 0.50$, and 0.70) are processed in this study, and the corresponding Reynolds numbers are approximately 4,7, and 11×10^5, respectively.

B. Phased microphone array

The phased-microphone array consists of 15 rings, the first ten of which have six microphones, and the latter five of which have four microphones, all axisymmetrically distributed (see Fig. 4). Thus, it carries a total of 80 quarter-inch microphones. The spreading angle of the array is approximately 11°, and the array covers $31.7° < \theta < 106.3°$ in the zenithal angle. The axial ring position is designed so that the zenithal angle is about equally spaced. The distances from the jet exit to the microphones are in the range of $17.3D < r < 48.3D$, while the ambient wavelength at St=0.35 for $M_\infty=0.35$ is $\lambda=8.3D$.

The positions of the microphones were mechanically measured and confirmed using photometry. The error estimated based on these two different measurements is less than $0.1D$, which is sufficiently small compared with wavelengths at frequencies of our interests (the shortest wavelength processed in this study is about $2.5D$).

C. Data processing

The sampling rate was set to be 50 kHz, and the low-pass filter was activated above 20 kHz. In computing Fourier transforms in time, four-period intervals at each target frequency were processed with the Hann window. Cross-spectral matrices were constructed from 500 sample intervals. For example, even if the number of samples is doubled for the baseline case ($M_\infty=0.5$ and St=0.35), the discrepancy of the greatest eigenvalue in the cross-spectral matrix is within 8% for $m=0,1$, and 2. To generate source maps, outputs up to $n=3$ are displayed with a spatial interval of $0.1D$ assuming that sources are located along the jet axis.

IV. RESULTS AND DISCUSSION

A. Validation using model problems

To demonstrate the capabilities of the proposed algorithm, the following simulation is carried out. Along the jet axis, three types of sources, $(m,n)=(0,0)$, $(0,1)$, and $(0,2)$ [i.e., P_0^0, P_1^0, and P_2^0 in Eq. (11)], are distributed according to the Gaussian distribution with a standard deviation of $1D$, centered at $x/D=5$, 3, and 7, respectively. The intensities of the modes $(m,n)=(0,1)$ and $(0,2)$ are set to be one-half of the mode $(0,0)$; hence, the target signal in this case is the mode, $(m,n)=(0,0)$. For each source type, 30 points are equally spaced within $0<x/D<10$ and their phases are set to be uncorrelated. In addition, uncorrelated random noise with a total intensity equivalent to the $(m,n)=(0,0)$ mode is added at each microphone to simulate undesirable disturbances.

Figure 5 compares source maps generated by (a) the conventional beam-forming with (b)–(d) those by the proposed algorithm for various values of ϵ. Parenthetically, the range of ϵ here satisfies the condition for the white noise gain suggested by Ref. 20. Note that β in Eq. (14), which optimizes the ratio of the first to the second peaks in the source map, is found to be zero for all cases in Fig. 5.

While the conventional beam-forming exhibits many spurious peaks, the proposed algorithm can identify the position of the primary source. Yet, when $\epsilon=\max[\lambda_l]$, the lobes appear too wide in Fig. 5(b) although the peaks are still distinguishable. As ϵ decreases, the main lobe becomes narrower and the outputs of the secondary modes are suppressed. With $\epsilon=0.01\max[\lambda_l]$, the total intensity of the sources cannot be recovered. At this level, the order of ϵ is comparable to the diagonal part of the random noise artificially added. Moreover, errors associated with the inaccuracy of the reference solution, such as measurement of the microphone positions or the offset of the source position from the jet axis, are usually amplified with decreasing ϵ.[9,20,21] Thus, $\epsilon=0.1\max[\lambda_l]$ is selected for the rest of the study. This corresponds to the magnitude of the fifth to tenth eigenvalues in the frequency range of our interests for the jet-noise study.

We should emphasize that the widths of lobes are controlled by ϵ. This implies that an output other than the local maxima does not infer the source intensity. In other words, source maps generated by beam-forming algorithms are generally just indicators of the peak source position and its total intensity, not the source distribution.

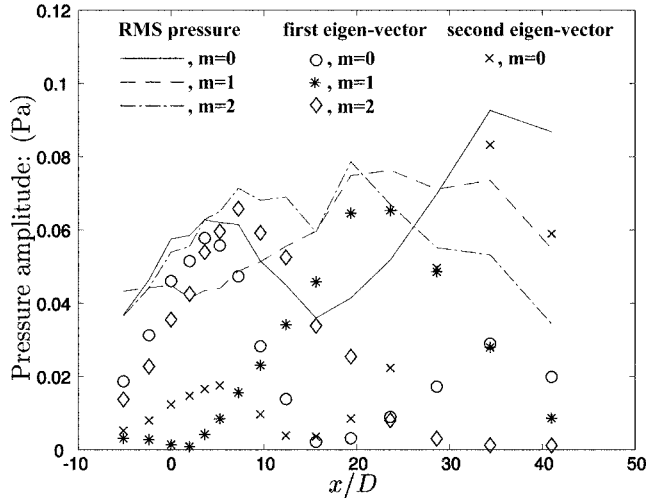


FIG. 7. Root-mean square of total pressure intensity at the microphone rings and dominant eigenvectors of the $N_{\text{ring}} \times N_{\text{ring}}$ cross-spectral matrix (absolute values of the eigenvector times $\sqrt{\lambda}$) for $M_\infty = 0.50$ and St=0.35.

As noted in Sec. II D, β is optimized if the ratio of the highest peak to the second highest one on the source map is an increasing function of β at $\beta = 0$. To demonstrate the necessity of this treatment, we consider the following model problem. We distribute the $(m,n) = (0,0)$ and $(0,1)$ modes centered at $x/D = 4$ and 6, respectively. Here, the intensity for $(0,1)$ is twice as high as that for $(0,0)$; hence, the target is the $(0,1)$ mode in this case. When we apply the robust adaptive beam-forming technique to a cross-spectral matrix reconstructed only from the first eigenmode, the outputs of both modes are comparable, as depicted in Fig. 6(a). In contrast, Fig. 6(b) shows that the peak of the target mode further exceeds the weaker mode when the optimal β is selected in the proposed algorithm.

Another test is examined to estimate the order of aliasing errors. Since six microphones are distributed on most rings, the aliasing error in $m = 2$ contaminated by $m = 4$ is considered most significant. Accordingly, two types of sound sources, $(m,n) = (2,2)$ and $(4,4)$ with the same intensity, are distributed in the same way as the previous model problem, centered at $x/D = 3$ and 7, respectively. The resultant source map (not exhibited) shows that the output at $x/D = 7$, which should be associated with $(m,n) = (4,4)$, is about three orders of magnitude smaller than that at $x/D = 3$. Similarly, a test with $(m,n) = (2,3)$ and $(4,4)$ provides a contrast of the same order. Thus, the aliasing error in the output is considered negligible in this study.

B. Jet-noise source identification

Before analyzing source maps of jet noise, some features of the coherent sound signals are observed. In Fig. 7, the pressure intensities for the axisymmetric and first two azimuthal modes are compared with their primary eigenmodes of the cross-spectral matrices. A flow condition of $M_\infty = 0.50$ at $St = 0.35$ selected here is defined as the baseline case for the rest of the study. The axisymmetric pressure profile depicts two peaks: The first eigenvector dominates the upstream peak, while the second one captures the downstream peak. The sum of these two modes occupies most of the energy [refer to Fig. 3(a) for $m = 0$]. On the other hand, total amplitudes for two azimuthal modes ($m = 1$ and 2) show relatively gentle directivity, but their first eigenvectors exhibit a clear single peak. Thus, the primary eigenmodes are generally more directive than the total amplitude.

We now compare source maps generated by the conventional, adaptive, and proposed beam-forming algorithms in Figs. 8–10, respectively. Both conventional and adaptive beam-forming algorithms generate very gentle peaks (recall that in these two methods, the full cross-spectral matrices are

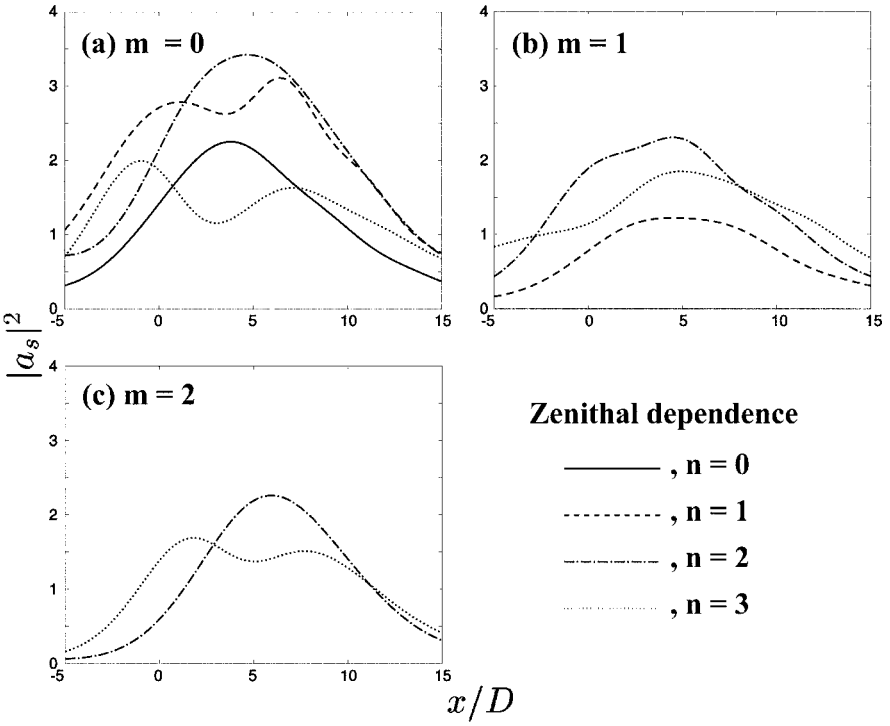


FIG. 8. Source maps generated by the conventional beam-forming algorithm ($M_\infty = 0.50$ and St=0.35).

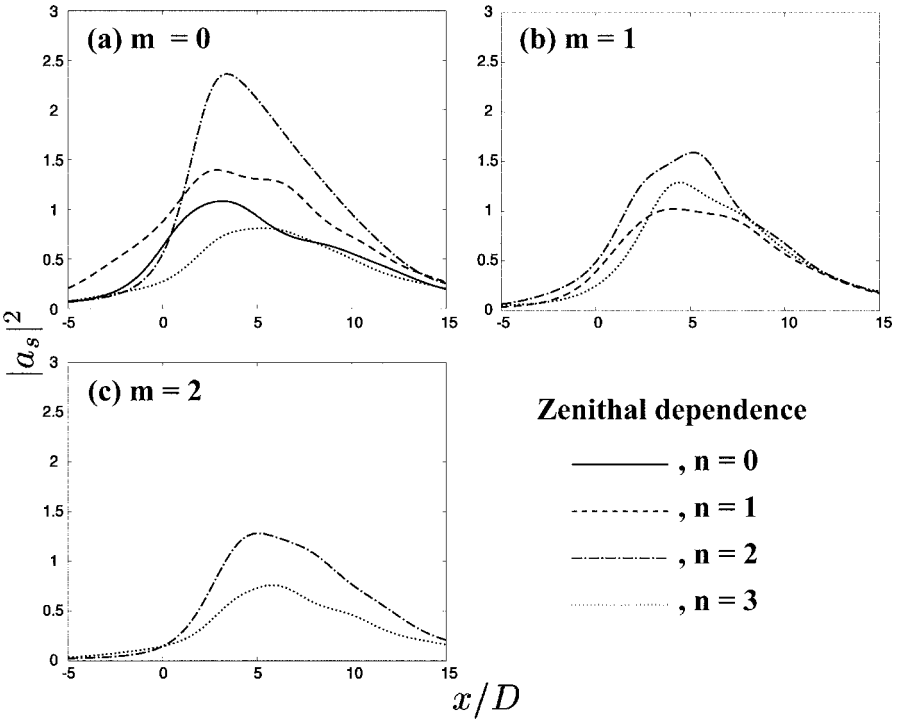


FIG. 9. Source maps generated by the adaptive beam-forming algorithm. Flow conditions are the same as Fig. 8.

processed as opposed to the reconstructed ones for the proposed algorithm). Moreover, in the conventional beam-forming, some outputs show multiple peaks with comparable magnitude, which are likely artifacts. We should remember from Fig. 3 that the first few coherent modes capture a large part of the total intensity. However, the conventional beam-forming tends to amplify outputs of spurious modes as shown in Fig. 5(a). This effect makes the source maps misleading. The regular adaptive beam-forming still indicates as if the source intensity levels of all modes are comparable. In contrast, the map generated by the proposed algorithm depicts a distinctive peak of a single mode for each m. The positions of these peaks approximately coincide with those localized by the conventional beam-forming.

Note that among the three maps of the proposed algorithm, only the axisymmetric case, i.e., Fig. 10(a), resulted in a nonzero value of the optimal β(=0.378). In fact, Fig. 3 shows that the first two eigenvalues are comparable only for $m=0$ (this typically happens when β becomes nonzero). In the original map [i.e., $\beta=0$, see Fig. 11(a)], the apparent peaks of $(m,n)=(0,0)$ and $(0,2)$ are comparable, while Fig. 10(a) exhibits a distinctive peak in the output of $(0,2)$. Since

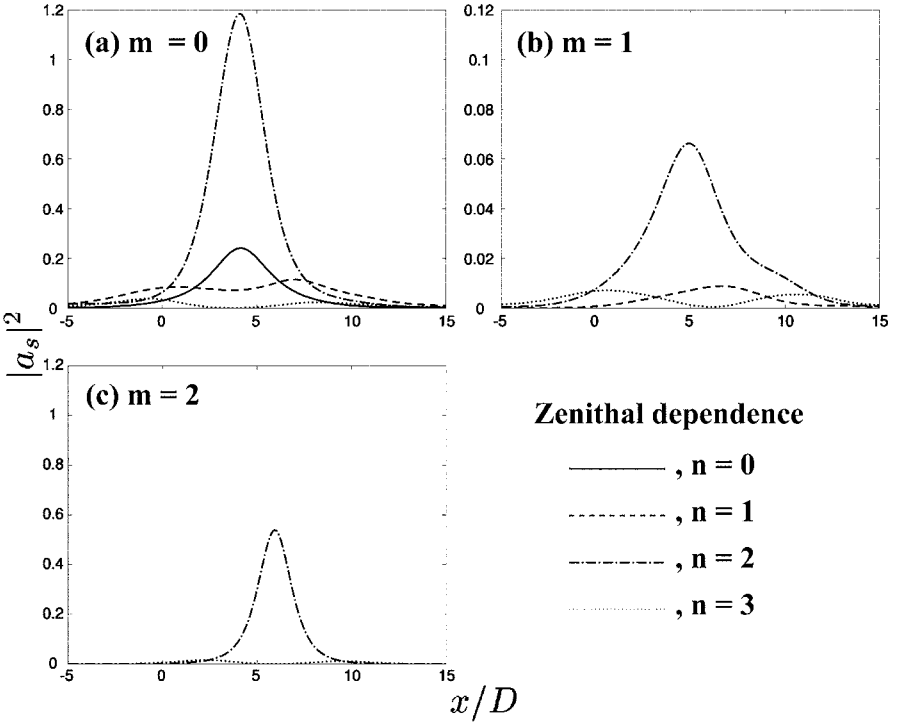


FIG. 10. Source maps generated by the proposed beam-forming algorithm. Flow conditions the same as Fig. 8.

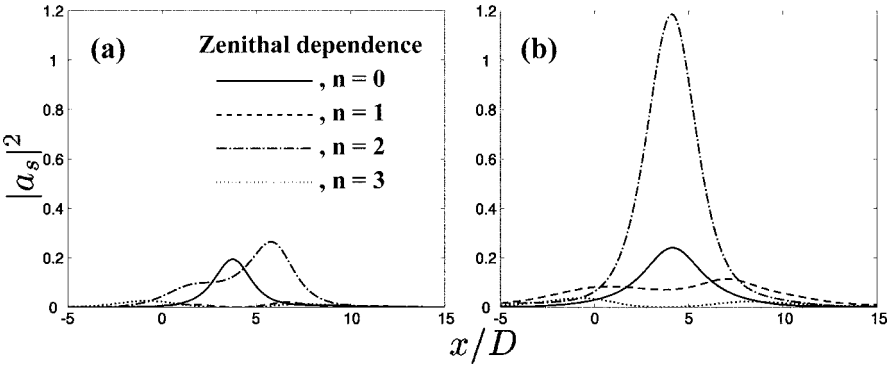


FIG. 11. Comparison of the source maps for $m=0$ at $M_\infty=0.50$ and St $=0.35$: (a) $\beta=0$ and (b) optimal β $(=0.378)$, which is identical to Fig. 10(a).

the features of the coherent signals are very similar between $M_\infty=0.35$ and 0.50, the center of the $(0,2)$ mode detected with the optimal β likely indicates the correct source position.

A drawback of the proposed algorithm is that the magnitude of an output does not indicate real source intensity. Figure 10(b) provides the peak magnitude that is one order smaller than the other two modes. On the other hand, we can obtain more information about the source intensity from the eigenvalues of the cross-spectral matrices. Assuming that the peak position of the coherent signals can be localized by the proposed algorithm, the radiation patterns are reconstructed from the primary eigenmodes, and the noise source structures are studied in the next section.

We notice that for all m, the second zenithal modes (i.e., $n=2$) provide the greatest peak in Fig. 10. These modes are quadrupoles in a free space (see Fig. 2). Recall that the source term of Lighthill's equation[25] is expressed as superposition of quadrupoles. Although Lighthill's analogy does not necessarily suggest that aeroacoustic noise sources consist of quadrupoles, the noise source maps produced in this study may indicate that the directivity of coherent sound signals is close to the quadrupole for each azimuthal mode. We make further observation about the radiation patterns below.

C. Dependence on Mach number and frequency

Finally, the sound radiation patterns of coherent signals are reconstructed. Assuming that the sound is generated at the positions detected by the proposed beam-forming algorithm, the directivity of the dominant sound is inferred by calculating the amplitude at each angle as follows:

$$\sqrt{(1-\beta)\lambda_1|u_1|^2+\beta\lambda_2|u_2|^2}. \tag{15}$$

Subsequently, the distances from the source to the microphones are corrected, and the effects of the Mach number and the frequency are investigated in the following.

Figure 12 depicts the radiation patterns and positions of coherent sound sources at three different Mach numbers. At $M_\infty=0.35$ and 0.50, the radiation patterns and the source positions are relatively similar, and the U_{jet}^4 law[25] (or somewhat lower power) scales the amplitude. The centers of the sound sources shift downstream with increasing azimuthal mode number. One noticeable difference is that the downstream lobe for $m=0$ is enhanced at $M_\infty=0.50$.

It is important to notice that the directivity of each mode is most similar to the quadrupole radiation pattern. Figure 13 actually illustrates the sound radiation pattern of $n=2$ for each azimuthal mode (also refer to Fig. 2). Although the zenithal angle of the array ranges only $31.7° < \theta < 106.3°$, the lobe patterns at lower Mach numbers in Fig. 12 are similar to the free-space quadrupoles. This is consistent with the observation in the preceding section.

As the Mach number is increased, the downstream lobe of $m=0$ is further amplified. In addition, the source centers are found farther downstream. At higher Mach numbers, the amplification rate of the Kelvin-Helmholtz instability is suppressed owing to the compressibility; consequently, the length of the potential core, at the end of which the peak frequency sound is generated, is extended downstream. It should be noted that as the Mach number increases, the magnitudes of secondary peaks on the source maps become comparable to that of the highest peak (not shown). At higher Mach numbers, sound intensity concentrates at shallower angles, and the radiation pattern significantly deviates from free-space multipoles.

We observe the frequency dependence in Fig. 14. As the frequency increases, sources for all three azimuthal modes tend to shift upstream. This trend is consistent with the results from previous studies.[4–7,14,17,18] Linear stability analysis also demonstrates that the neutrally stable point, at which the disturbance is saturated in the streamwise direction, appears more upstream with increasing frequency. The radiation patterns qualitatively remain the same over a frequency, except for $m=1$ at $St=0.80$. In fact, the peak output of this mode on the source map (not shown) is obtained from the $n=1$ mode (i.e., dipole); however, the reason for the switch from the P_1^2 type to the P_1^1 type is not understood.

V. CONCLUSIONS

This study has developed a beam-forming algorithm that can identify the sound radiation patterns and source positions of coherent jet noise. By using a conical microphone array in the far field, the axisymmetric and first two azimuthal modes have been investigated near the peak frequency. The algorithm developed in this study is analogous to the robust adaptive beam-forming technique applied to a cross-spectral matrix reconstructed from the primary eigenvectors. The proposed algorithm has been examined with experimental data of an unheated round jet over a range of Mach numbers and frequencies.

The results of this study indicate that the second zenithal mode, namely a quadrupole, dominates the coherent signals

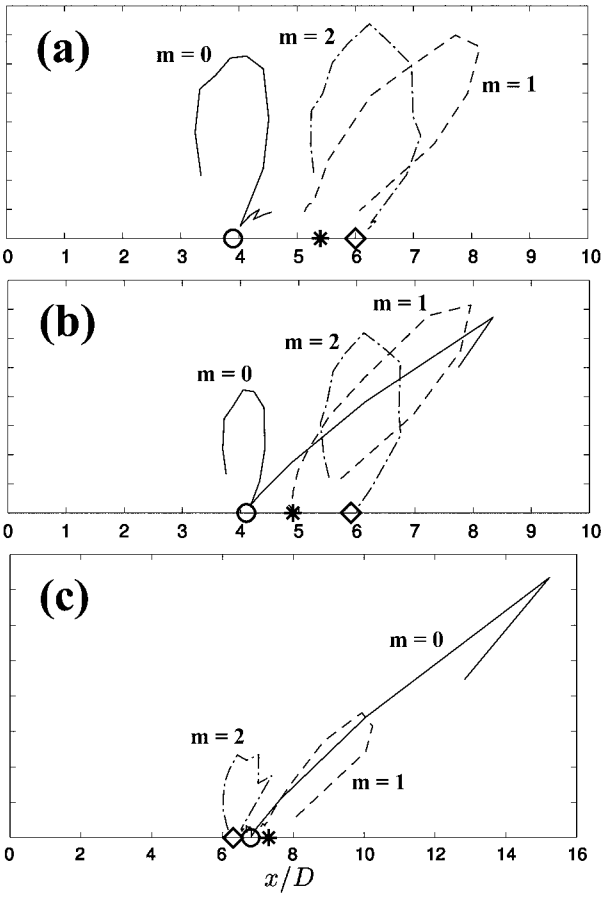


FIG. 12. Sound radiation patterns and source positions of primary coherent signals at different Mach numbers (St=0.35). Pressure amplitude of the composite eigenvector (15)is depicted as a polar plot centered at the peak of the source map (distance from the source is corrected). Amplitude is scaled with U_{jet}^4 (the unit is arbitrary). (a) M_∞=0.35 (b) M_∞=0.50. (c) M_∞=0.70. Source center: ∘, m=0; *, m=1; ◇, m=2. Radiation pattern: —, m=0; – – –, m=1; –·–·–, m=2.

for each azimuthal mode at low Mach numbers, and their centers shift downstream with increasing azimuthal mode. Lighthill's acoustic analogy[25] does not educe that aerodynamic noise sources should consist of quadrupole structures, but several theoretical studies[26,27] have proven that the interaction of two vortex rings emits quadrupole-type sound. From this study, we can deduce that the Kelvin-Helmholtz instabilities in the mixing layer induce large-scale vortex motion and this causes generation of coherent sound near the end of the potential core. Thus, extensive study using phased-microphone array techniques can help understand the mechanisms of noise generation from subsonic jets.

It should be remembered that the proposed algorithm assumes that the noise sources are stationary and the signal-to-noise ratio is reasonably high. These assumptions seem applicable at low frequencies because the coherent signals exhibit strong directivity and occupy a large part of the total intensity. At high frequencies, however, convecting small eddies may generate appreciable sound. In addition, we are no longer able to assume that sources are located on axis because the phase matching becomes more sensitive to the source position. Hence, the proposed algorithm is desirable at low Strouhal numbers.

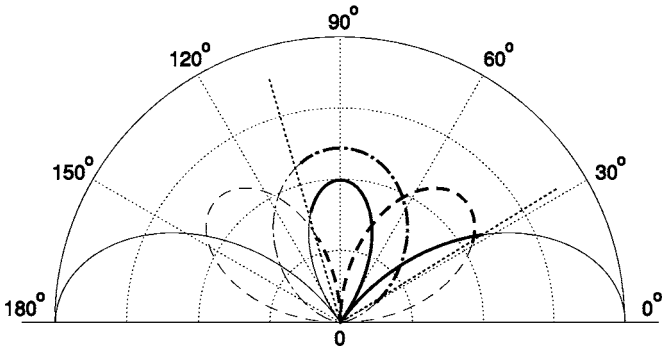


FIG. 13. Sound radiation patterns of quadrupoles. Absolute amplitude is depicted as a polar plot. The zenithal angles covered by the array are denoted by thicker lines. The rest of the notation is the same as Fig. 12.

At higher Mach numbers or higher jet temperature, sound radiation patterns are highly directive; hence, multipoles in a free space are no longer appropriate for reference solutions. From this study, it is not clear whether the highly directive pattern is generated due to refraction or associated with instability waves in a spreading mixing layer. Such a question can be investigated, for example, by using asymptotic Green's functions including the refraction effect for reference solutions. Further development of beamforming algorithms applicable to high Mach numbers or high temperature is important for practical jet noise study.

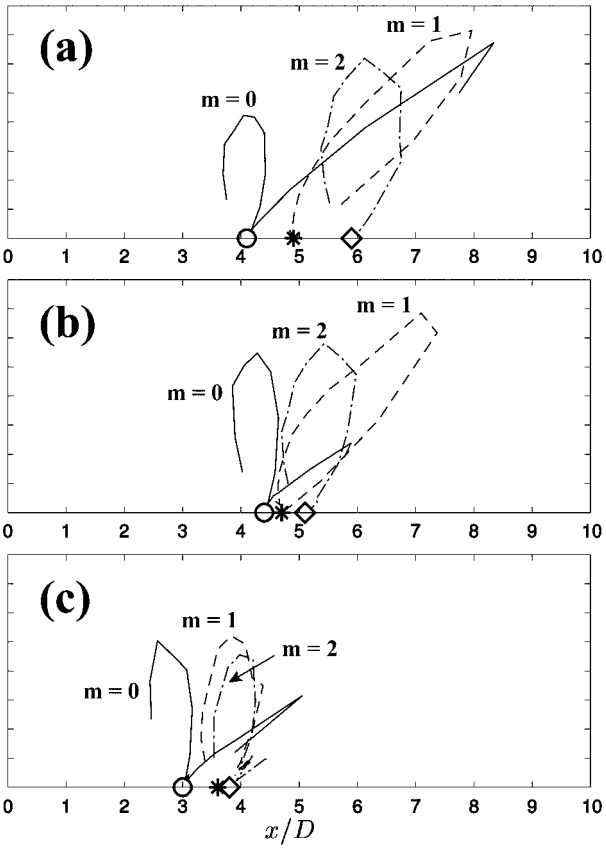


FIG. 14. Sound radiation patterns and source positions of primary coherent signals at different Strouhal numbers (M_∞=0.50): (a) St=0.35, (b)St=0.50, and (c)St=0.80. Notation is the same as Fig. 12.

ACKNOWLEDGMENTS

This research was supported by a grant from the Aero-Acoustics Research Consortium (AARC) of the Ohio Aerospace Institute (OAI). The jet-noise experiments were conducted by Dr. J. Bridges and Dr. S. -S. Lee at NASA Glenn Research Center. And the author is grateful for their generosity in sharing the phased microphone data. The author would also like to thank Professor T. Colonius at the California Institute of Technology and Dr. R. Dougherty at the OptiNav Inc. for useful advice and fruitful discussion.

[1]J. T. C. Liu, "Developing large-scale wavelike eddies and the near jet noise field," J. Fluid Mech. **62**, 437–464 (1974).
[2]C. K. W. Tam and P. J. Morris, "The radiation of sound by the instability waves of a compressible plane turbulent shear layer," J. Fluid Mech. **98**, 349–381 (1980).
[3]D. G. Crighton and P. Huerre, "Shear-layer pressure fluctuations and superdirective acoustic sources," J. Fluid Mech. **220**, 355–386 (1990).
[4]F. R. Grosche, "Distributions of sound source intensities in subsonic and supersonic jets," AGARD Conf. Proc. **131**, 4 (1974).
[5]J. Billingsley and R. Kinns, "The acoustic telescope," J. Sound Vib. **48**, 485–510 (1976).
[6]M. J. Fisher, M. Harper-Bourne, and S. A. L. Glegg, "Jet engine noise source location," J. Sound Vib. **58**, 117–126 (1978).
[7]K. K. Ahuja, K. C. Massey, and M. S. D'Agostino, "A simple technique of locating noise sources of a jet under simulated forward motion," AlAA paper, AIAA 98-2359 (1998).
[8]U. Pillai, *Array Signal Processing* (Springer, New York, 1989), Chap. 2, pp. 8–107.
[9]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice–Hall, Englewood Cliffs, New Jersey, 1993).
[10]A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[11]J.-F. Piet and G. Elias, "Airframe noise source localization using a microphone array," AIAA paper, AIAA 97–1643 (1997).
[12]D. Blacodon and G. Elias, "Level estimation of extended acoustic sources using an array of microphones," AIAA paper, AIAA 2003-3199 (2003).
[13]P. T. Soderman and S. C. Noble, "Directional microphone array for acoustic studies of wind tunnel models," J. Aircr. **12**, 168–173 (1975).
[14]R. P. Dougherty, "Phased array beam-forming for aeroacoustics," AIAA professional development short course (1999).
[15]T. Suzuki and T. Colonius, "Instability waves in a subsonic round jet detected using a near-field phased microphone array," J. Fluid Mech. (to be published).
[16]T. F. Brooks and W. M. Humphreys, "Flap-edge aeroacoustic measurements and predictions," J. Sound Vib. **261**, 31–74 (2003).
[17]T. Suzuki and G. W. Butler, "New beam-forming algorithm for high speed jet flows," AIAA paper, AIAA 2002–2505 (2002).
[18]S. R. Venkatesh, D. R. Polak, and S. Narayanan, "Beam-forming algorithm for distributed source localization and its application to noise," AIAA J. **41**, 1238–1246 (2003).
[19]A. Ghosh, J. Bridges, and F. Hussain, "Instantaneous directivity in jet noise by multipole decomposition," ASME J. Vibr. Acoust. **117**, 172–179 (1995).
[20]H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," IEEE Trans. **ASSP-35**, 1365–1376 (1987).
[21]R. A. Gramann and J. W. Mocio, "Aeroacoustic measurements in wind tunnels using adaptive beam-forming methods," J. Acoust. Soc. Am. **97**, 3694–3701 (1995).
[22]R. P. Dougherty and R. W. Stoker, "Sidelobe suppression for phased array aeroacoustic measurements," AIAA paper, AIAA 98-2242, (1998).
[23]R. O. Schmidt, "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**, 276–280 (1986).
[24]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1970).
[25]M. J. Lighthill, "On sound generated aerodynamically. 1. General theory," Proc. R. Soc. London, Ser. A **211**, 564–587 (1952).
[26]W. Möhring, "On vortex sound at low Mach number," J. Fluid Mech. **85**, 685–691 (1978).
[27]T. Kambe, "Acoustic emissions by vortex motions," J. Fluid Mech. **173**, 643–666 (1986).

First-order and zeroth-order head waves, their sequence, and implications for geoacoustic inversion

Jee Woong Choi and Peter H. Dahl
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

(Received 11 January 2005; accepted 13 March 2006)

The relation between the head wave and an arrival often called a ground wave is analyzed with parabolic equation-based simulations, and an interpretation of such a ground wave as a head wave sequence is presented. For a Pekeris waveguide the envelope of the spectrum of the ground wave arrival corresponds to the spectrum of a single head wave, and spectral peaks correspond to odd multiples of the mode-1 cutoff frequency. This head wave is first order in its ray series classification and its amplitude spectrum goes as $|S(f)|/f$, where $S(f)$ is the source amplitude spectrum. Basic variations from a Pekeris waveguide are considered; isospeed layers or a positive sound speed gradient in the seabed can each give rise to arrivals that are zeroth order in ray series classification and higher amplitude. For a sound speed gradient there is either a low-amplitude interference head wave whose properties are akin to a first-order head wave, or a high-amplitude interference head wave or non-interfering refracted wave whose properties are zeroth order and spectra follow $|S(f)|$. Parametric dependences for distinguishing these arrivals and implications for geoacoustic inversion are discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195110]

PACS number(s): 43.30.Bp, 43.30.Cq, 43.30.Ma [WLS] Pages: 3660–3668

I. INTRODUCTION

In a two-fluid waveguide with water depth h and the water sound speed c lower than the sediment speed c_1 (Pekeris waveguide), there exists a critically refracted wave propagating along the water-sediment interface at the speed of c_1. This is called a head wave, although other terminology, such as lateral wave, is also used. Specifically, the horizontal wavenumber (k_r) representation of the pressure field at frequency f contains a head wave whose contribution is the branch line integral about the branch point at $k_r = 2\pi f/c_1$ in the k_r plane.[1–4] To make clear the subsequent nomenclature, we shall call this a *first-order head wave*, because of its relation to the ray series.[5] For a broadband source, the term *ground wave* is assigned to an arrival (though not exclusively) that is explained by modal dispersion effects, by which for frequencies less than the Airy phase frequency, the group velocity approaches that of the seabed sound speed. (This definition follows that of Brekhovskikh;[6] see also Frisk[2] and Clay and Medwin.[7]) Thus, both the head wave and ground wave represent arrivals propagating at a speed greater than the sound speed in seawater.

Here we present an interpretation of the ground wave (as defined above) as a train of the head waves for a Pekeris waveguide, which we will refer to as a *head wave sequence*. Insofar as the group velocity, modal dispersion, and time spreading of an impulsive source signal in a waveguide can be explained by the guiding of energy caused by reflection at boundaries or refraction by the sound speed profile (geometrical dispersion[8,9]), it is possible to explain the modal ground wave using a ray representation.

The above result is of more theoretical than practical interest. However, it forms the basis of the remainder of this work, which addresses more complex waveguides and is motivated by the use of head waves in geoacoustic inversion. For example, Godin, *et al.*[10] use information from precursor head waves to invert for seabed compressional wave speed (via arrival time versus range and depth) and compressional wave attenuation (via analysis of the spectrum of the precursor head waves). Others[11,12] have utilized head wave arrival time versus range information to invert for seabed compressional wave speed. Looking ahead, we demonstrate that head waves are rather strongly sensitive to the presence of a sound speed gradient in the seabed, and consequently change from first order to zeroth order in their ray series classification. We also demonstrate (via simulation) that geoacoustic inversion using head wave data depends critically on proper classification of the head wave into either first-order or zeroth-order types, the latter being much more energetic. The effects of the sound speed profile in the water column, changes in bathymetry in terms of bottom slope, layering in the seabed, and a sound speed gradient in the seabed on head waves and their sequence are also examined.

The paper is organized as follows. Arguments relating the ground wave to a sequence of head wave arrivals for a Pekeris waveguide are presented in Sec. II; the simulations [based on the parabolic equation (PE)] that illustrate these points are presented in Sec. III. Simulations of a head wave sequence in a more complex shallow water waveguide and important differences from a Pekeris waveguide are discussed in Sec. IV. A summary and discussion are given in Sec. V.

II. THE GROUND WAVE AS A HEAD WAVE SEQUENCE FOR A PEKERIS WAVEGUIDE

For a Pekeris waveguide the cutoff frequency (f_c) of mode m is given by[2,7]

 0001-4966/2006/119(6)/3660/9/$22.50

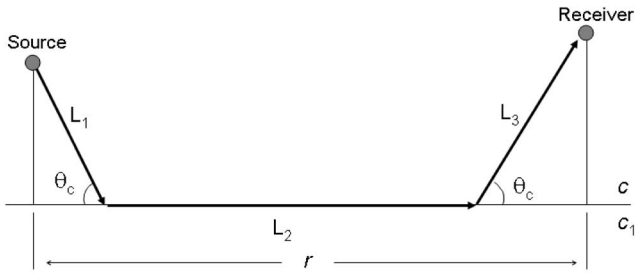


FIG. 1. Ray representation of a first-order head wave from the source to the receiver in the absence of a sea surface boundary. The upper medium has sound speed c, and the lower medium has speed c_1. r is the horizontal range between the source and receiver.

$$f_c = \frac{c(m-0.5)}{2h\sqrt{1-(c/c_1)^2}}. \tag{1}$$

As the frequency tends toward the cutoff frequency from above, the group velocity approaches c_1, and this region in the dispersion relation corresponds to ground wave propagation; sound energy at frequencies less than the cutoff frequency is effectively untrapped and not important at appreciable distances from the source.

The usual description of a head wave existing between two half spaces involves an eigenray connecting source to receiver that is incident on the seabed at the critical angle θ_c after propagation along path L_1 (Fig. 1), then after propagating horizontally along path L_2, it radiates into the upper water medium along path L_3, again at the critical angle. The asymptotic expression for the pressure associated with this head wave is[1,2,6]

$$p_H \sim \frac{i2nS(f)}{km(1-n^2)r^{1/2}L_2^{3/2}} \exp[ik(L_1+L_3)+ik_1L_2], \tag{2}$$

where $S(f)$ is the source spectrum, k and k_1 are wavenumbers in the upper and lower media, respectively, n is the index of refraction (c/c_1), and m is the ratio of lower (ρ_1) and upper (ρ) medium densities. Equation (2) can, for example, be derived from branch-line integration about the branch point in the complex horizontal wavenumber plane originating at k_1. In Eq. (2) the spreading loss for p_H goes as $1/(r^{1/2}L_2^{3/2})$, and the amplitude spectrum of the head wave is proportional to $|S(f)|/f$ (in the absence of attenuation in the lower seabed medium), both of which are consistent with the first-order ray series solution.

Let us next, for simplicity, consider a Pekeris waveguide with both source and receiver located just above the water-sediment interface. Figure 2 shows three head wave arrival types generated in this case. (Other head waves are generated in this case with the same travel time and phase, e.g., Fig. 2, inset.) One propagates along path I (S-R) and arrives at time $t_1=r/c_1$, and another propagates along path III (S-A-B-C-D-R) and arrives at time $t_3=(r-r_1)/c_1+4h/(c\sin\theta_c)$, where r_1 is the horizontal range in which the head wave propagates through the water column. The travel time difference between paths III and I is $4h\sqrt{1-(c/c_1)^2}/c$ and its inverse, the characteristic frequency of the I–III path sequence, is the same as the mode 1 cutoff frequency for the waveguide [Eq. (1), with $m=1$]. The phase of the head wave propagating

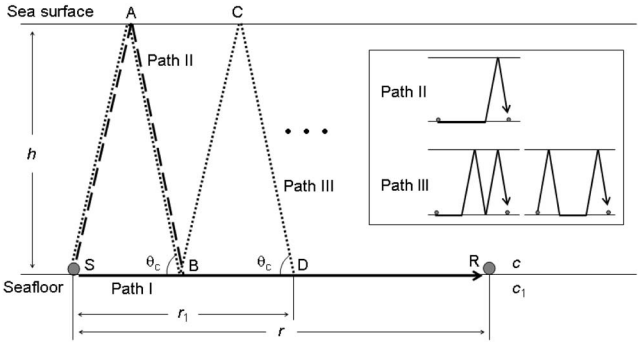


FIG. 2. Ray paths for first-order head waves between a near-bottom source (S) and a near-bottom receiver (R). (Inset) Ray paths for other first-order head waves with equivalent travel time and phase.

along path II (S-A-B-R) differs by π from those in paths I and III owing to the reflection from the sea surface. The amplitude of the head wave for path I decreases as $1/r^2$, but those for paths II and III decrease spherically for their in-water range segments S-A-B and S-A-B-C-D, respectively, then decrease as the square of range for segments B-R and D-R, respectively. As a consequence, the amplitude of the resultant head wave sequence shows gradual growth with time, which is also a characteristic of a ground wave for a Pekeris waveguide that arrives at time r/c_1.[6,7]

A useful view of the contribution of higher-order modes to the head wave sequence is obtained by considering the Fourier transform of a periodic, broadband signal, $x_{HS}(t)$, for which the phase inverts for every other signal. The location of the first impulse in frequency is $1/2T_P$, where T_p is the time interval between broadband signals; the frequency interval is $1/T_P$, and the envelope of amplitude spectrum is governed by the spectrum $[X_H(f)]$ of $x_H(t)$ [Fig. 3(b)]. Here, $x_H(t)$ in Fig. 3(a) is considered a single head wave signal, $x_{HS}(t)$ a head wave sequence (or the ground wave), and each impulse in $X_{HS}(f)$ corresponds to a modal cutoff frequency of a normal mode. That is, peaks in the frequency spectrum of a ground wave arrival (corresponding to one or more modal cutoff frequencies) are governed by the spectrum of a single head wave, owing to the ground wave consisting of a sequence of such head waves.

III. SIMULATION OF A HEAD WAVE SEQUENCE FOR A PEKERIS WAVEGUIDE

A time-dependent simulation of a head wave sequence for a Pekeris waveguide is obtained via Fourier transform of

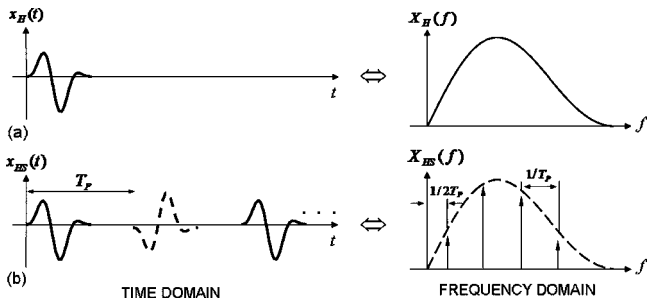


FIG. 3. Fourier transform pairs of (a) a single broadband signal, $x_H(t)$, and (b) a periodic broadband signal, $x_{HS}(t)$, for which phase inverts for every other signal and the space of impulses is $1/T_P$.

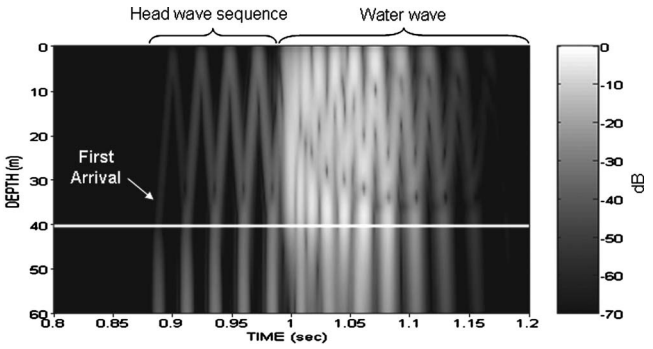


FIG. 4. Simulated acoustic field as a function of depth and time for a Pekeris waveguide with source on the bottom.

a frequency-domain solution, with the latter derived using the RAM (Range-dependent Acoustic Model) PE code.[13] For this, the following geoacoustic parameters were used: $h=40$ m, $c=1508$ m/s, $c_1=1689$ m/s, $\rho=1033$ kg/m^3, and $\rho_1=2\rho$ (for this demonstration seabed attenuation is artificially set to zero). These parameters yield cutoff frequencies for modes 1 through 5 of 21, 63, 105, 147, and 189 Hz, respectively. Figure 4 shows the PE acoustic field level for a source on the bottom, i.e., as in Fig. 2, and range 1.5 km, based on a source spectrum of a Hanning window centered at 150 Hz, with a half-power bandwidth B equal to 110 Hz. (Unless specified otherwise, all broadband simulations utilize this source spectrum.)

The first-arriving contribution in Fig. 4 is associated with path I (Fig. 2) and its amplitude is weak given the $r^{1/2}L_2^{3/2}$ dependence of first-order head waves; subsequent arrivals show increasing amplitude. With a specific receiver depth of 40 m (to again mimic Fig. 2), the PE time series is as shown in Fig. 5. In the expanded amplitude scale of Fig. 5(b), the thick line is the quasisinusoidal ground wave or head wave sequence, and the thin line is the water wave including the Airy phase (not distinguishable). The expanded

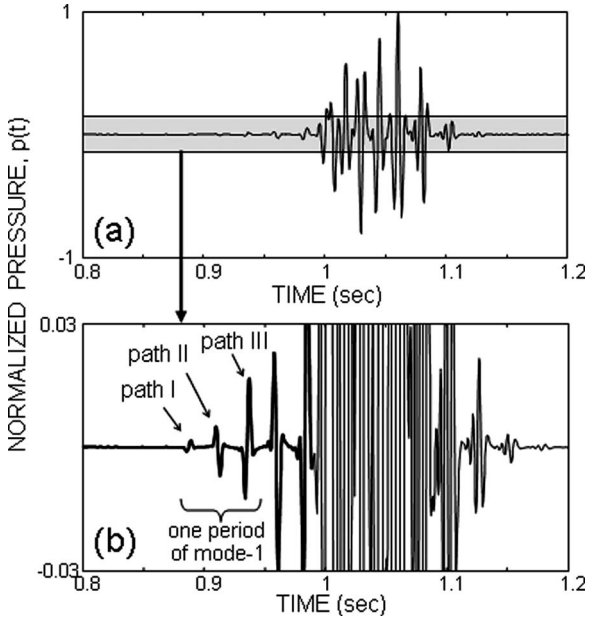


FIG. 5. (a) Time signal simulated for receiver depth of 40 m in Fig. 4 and (b) the same with an expanded vertical scale.

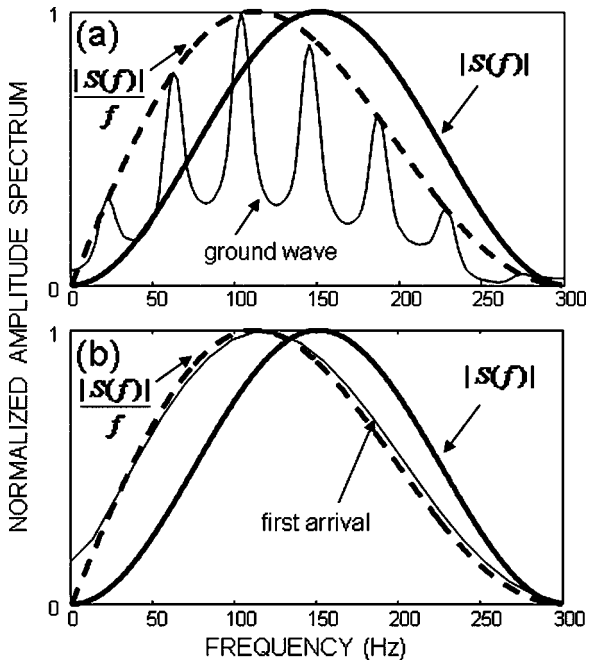


FIG. 6. Frequency spectra of (a) a simulated ground wave as head wave sequence (thin solid line) and (b) a first-arriving head wave (thin solid line). Thick solid line and thick dashed line are the source and theoretical head wave spectra, respectively.

scale reveals the first, second, and third arriving head waves that correspond to ray paths I, II, and III in Fig. 2, respectively, and the phase reversals mentioned previously. One period of the head wave sequence is the time difference between the first- and third-arriving head waves, which corresponds to the mode 1 cutoff frequency.

Figure 6 illustrates the relationship between the source amplitude spectrum, $|S(f)|$, the spectrum of the head wave sequence, and the amplitude spectrum of a first-order head wave, $|S(f)|/f$. First, the envelope of the head wave sequence spectrum [Fig. 6(a)] is well described by $|S(f)|/f$. Second, the spectrum of the first-arriving head wave [Fig. 6(b)] closely matches $|S(f)|/f$, representing consistency between the PE simulated amplitude spectrum and that predicted by the ray series [or Eq. (2)]. That the envelope of the head wave sequence spectrum goes as $|S(f)|/f$ can be better understood in the context of Fig. 3.

Figure 7 shows the comparisons between the simulated head wave sequence (thin line) and bandpass filtered components of the head wave sequence (thick gray line) corresponding to the first five modes. The left side shows the case for which the cutoff frequency for mode 1 is included in the source spectrum (i.e., the case considered previously) and the right side shows the case for which this frequency is not included, which is achieved by shifting $|S(f)|$ to the right by 50 Hz. Each cutoff frequency component is obtained by bandpass filtering using bandwidths of 1–40 Hz, 40–80 Hz, 80–125 Hz, 125–170 Hz, and 170–210 Hz, which corresponds to the approximate modal cutoff frequency ±20 Hz. For the case that includes the mode 1 cutoff frequency, the mode 1 component of the head wave sequence exists, though quite small [Fig. 7(a)], which is consistent with mode 1 energy in the head wave spectrum also being small. The domi-

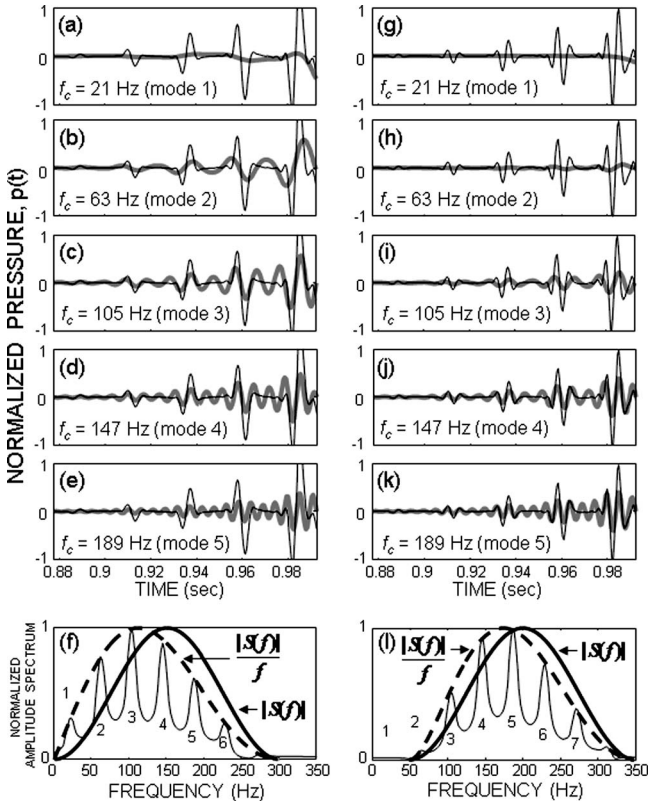


FIG. 7. Comparisons between the simulated ground wave or head wave sequence (thin line) and its narrow band filtered components (thick gray line) for (a), (g) mode 1, (b), (h) mode 2, (c), (i) mode 3, (d), (j) mode 4, and (e), (k) mode 5. The left column is the case for which the mode 1 cutoff frequency is included in the source spectrum [see $|S(f)|$ in (f)] and the right column is the case for which the mode 1 cutoff frequency is excluded [see $|S(f)|$ in (1)]. Thin lines in (f) and (l) are amplitude spectra for the unfiltered ground wave for each case.

nant component in the head wave sequence is associated with mode 3 [Fig. 7(c)], as anticipated by Fig. 7(f). For the case that excludes the mode 1 cutoff frequency (Fig. 7, right side), that component (21 Hz) is absent and the dominant component in the head wave sequence is associated with mode 5 [Fig. 7(k)], as anticipated by Fig. 7(l).

In the proceeding cases both the source and receiver are located on the bottom. Were either to be located within the water column, weighting by the depth-dependent modal eigenfunctions changes the arrival structure of the head wave sequence, although the concepts described thus far still apply. This is illustrated by Fig. 8 for which (a) shows the depth dependence of the modal eigenfunctions from modes 1–6 evaluated at their modal cutoff frequencies. Figures 8(b)–8(d) compare amplitude spectra of head wave sequences obtained from PE simulation for different source/receiver depth combinations, with the relative heights of the peaks within these spectra corresponding to each modal cutoff frequency f_i. The relative peak amplitudes at f_i are set by $|S(f_i)/f_i|W_i$, where W_i is the modal eigenfunction amplitude at the source depth multiplied by that at the receiver depth.

IV. THE HEAD WAVE SEQUENCE IN MORE COMPLEX SHALLOW WATER WAVEGUIDES

The Pekeris waveguide is a useful construct to analyze shallow water acoustic propagation. More complex features common to shallow water conditions will, however, alter the properties of head waves and their sequence. For example, consider a negative sound speed profile in the water column, associated with a summer thermocline, and maintain the isospeed condition in the seabed (here, any resulting head waves are still of the first-order kind). If the seabed sound speed is less than the maximum at the sea surface, the head wave sequence largely vanishes as the (first-order) amplitudes are small compared with the (zeroth-order) amplitudes of the water wave arriving before or nearly simultaneously. For the case in which the seabed sound speed exceeds the maximum water column speed at the sea surface, the ensuing refraction effects produce only minor changes in the ground wave arrival structure (head wave sequence) from that of a Pekeris waveguide. This point is illustrated by Fig. 9 show-

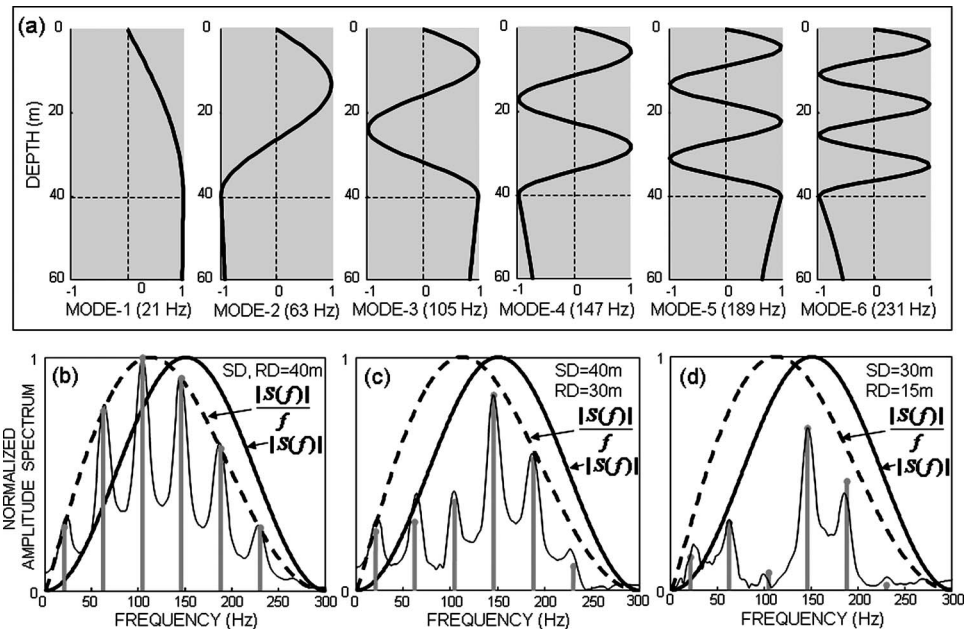


FIG. 8. (a) Depth-dependent modal eigenfunctions for modes 1–6 evaluated at their modal cutoff frequencies. (b)–(d) Amplitude spectra of the ground wave for different source/receiver depth combinations (thin, solid line), compared with $|S(f)|$ and $|S(f)|/f$. Vertical gray lines in each case are $[|S(f_i)|/f_i]W_i$, where W_i is the modal eigenfunction amplitude at the source depth multiplied by that at the receiver depth.

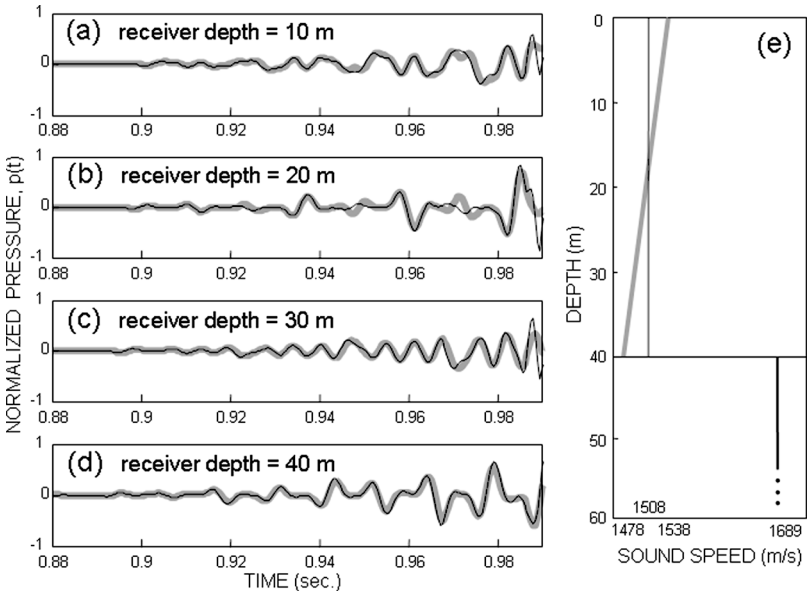


FIG. 9. Ground wave arrivals for a Pekeris waveguide (thin black line) and for a waveguide in which the water column has a negative sound speed profile (thick gray line). The source depth is 20 m and the receiver depths are (a) 10 m, (b) 20 m, (c) 30 m, and (d) 40 m. (e) Sound speed profiles used for these simulations.

ing PE simulated ground wave arrivals for a Pekeris waveguide compared with those for a waveguide for which the water column sound speed profile is negative, but all other waveguide properties are same.

In a waveguide with a sloping bottom, the frequency properties of the head wave sequence are significantly altered as the difference in water depth between the source and receiver locations increases and the index of refraction (c/c_1) decreases, along with increasing bottom slope (ε). These effects can be described approximately with the following:

$$\Delta T = \frac{2|D_s - D_r|(c_1^2 - c^2) - 2|\varepsilon|\sqrt{1-(c/c_1)^2}(D_s + D_r)cc_1}{cc_1^2\sqrt{1-(c/c_1)^2}}, \tag{3}$$

where ΔT is the time difference between the two paths, each identified as "Path II" (see Fig. 2 plus an inset), which becomes zero in a Pekeris waveguide, and D_s and D_r are the water depths at source and receiver locations. We define a parameter $\Delta T f_{\mathrm{HW}}$, where f_{HW}, is the dominant head wave frequency. An exhaustive exploration of this parameter space is beyond the scope of this paper; however, from PE simulation studies over a limited range of slopes, water depths, and sound speeds, we find that the influence of the bottom slope is negligible for $\Delta T f_{\mathrm{HW}} < 0.1$ (i.e., the resultant spectrum is similar to that for a Pekeris waveguide based on the average of D_s and D_r). This result can be understood from the viewpoint of the energy spectrum of the superposition of a pulse and one delayed by ΔT; the resultant spectrum will have properties akin to in-phase superposition, provided $2\pi \Delta T f \ll 1$.

Layering in the sediment will give rise to head waves propagating along interfaces. Figure 10 shows the PE time series (corresponding to range 1.5 km) simulated for a surficial sediment layer of thickness 30 m and sound speed 1689 m/s, overlying a half space with sound speed 1800 m/s; the sediment densities in both the surficial layer and the half space are set to twice the water density. The water depth is 40 m, and the source and receiver depths are 40 and 38 m, respectively. For ease of identification, the first few head wave arrivals are numbered in Fig. 10(b), which correspond to the path numbers in Fig. 10(d). First-order head waves (e.g., arrivals 1, 2, and 3) generated by the two

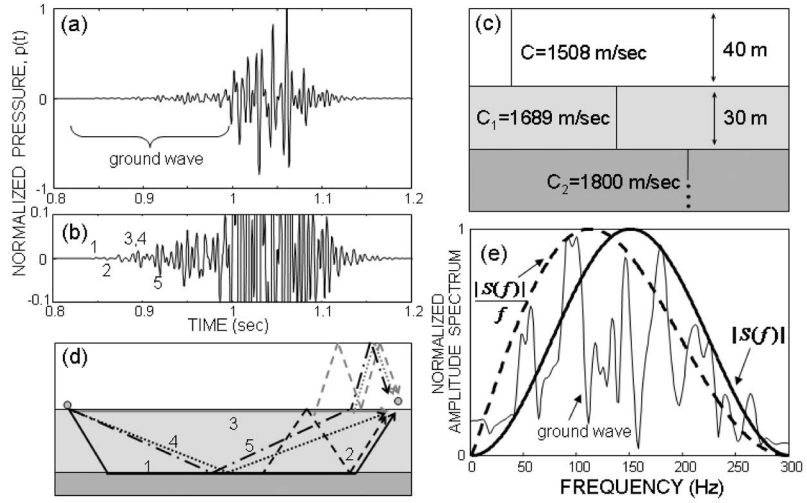


FIG. 10. (a) PE simulated time series for two sediment layers and for source 40 m and receiver 38 m depth; (b) the same with an expanded vertical scale showing the ground wave; (c) layered waveguide consisting of a surficial sediment layer with a thickness of 30 m overlying a half space; (d) schematic ray representations of the head waves and the reflected waves propagating along two interfaces formed by layering of sediment; and (e) the amplitude spectrum of the ground wave (thin solid line).

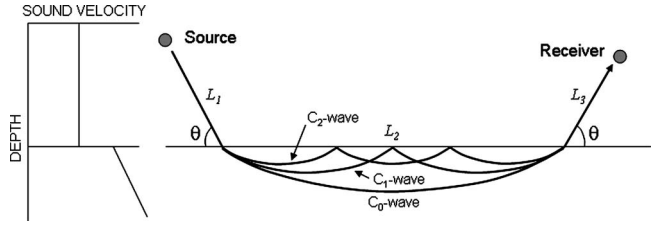


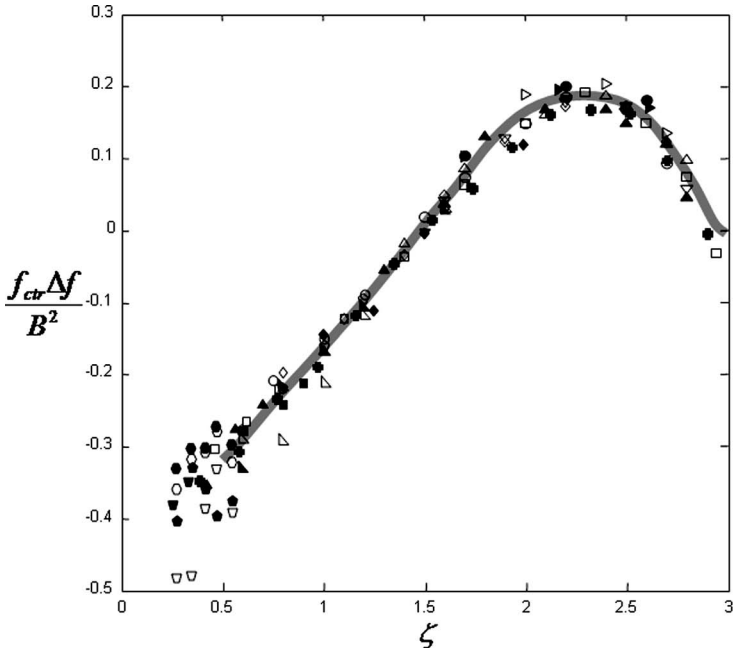
FIG. 11. The refracted waves for a positive sound speed gradient in the seabed. Note that angle θ varies with C_n waves, and is greater than θ_c.

interfaces exist in the overall ground wave, in addition to reflected waves (e.g., arrivals 4 and 5) that in a ray series classification are zeroth order and thus have a considerably higher amplitude. Whereas the ground wave for a Pekeris waveguide is a periodic (or quasisinusoidal) head wave sequence (Fig. 5), the ground wave with sediment layering is no longer a periodic head wave sequence because of the combination of first-order head waves and zeroth-order reflected waves and their interference effects. Because of this combination, the amplitude spectrum [Fig. 10(e)] follows neither the first-order characteristic of $|S(f)|/f$ nor the zeroth-order characteristic of $|S(f)|$.

A sound speed gradient in the sediment gives rise to very different effects via-à-vis head waves and their sequence from that discussed thus far; nevertheless, some of the aforementioned ideas apply. Here we consider the case of a positive gradient a starting at c_{1s} that is greater than the sound speed (c) in the water medium, and incorporate and expand on some derivations given by Červený and Ravindra.[1] Replacing the first-order head wave as the earliest arrival is the C_0 wave (also called a refracted wave or a diving wave) with a grazing angle that exceeds the critical angle and undergoes upward refraction with a single turning point, with subsequent arrivals (C_n waves, $n=1,2,3,\ldots$) being multiple refracting-seabed-reflecting (Fig. 11). Provided these arrivals are noninterfering (time separable), the C_n waves ($n=0,1,2,\ldots$) are zeroth order in their ray series classification and have amplitude spectra that follow $|S(f)|$ rather than $|S(f)|/f$.

The arrival time difference between the C_0 wave and the next arrival with three turning points (C_1 wave), ΔT_{0_1}, is estimated by $a^2L_2^3/32c_{1s}^3$ (see Červený and Ravindra[1]); when $\Delta T_{0\_1} < \delta t$, where δt is the duration of the first arrival, these waves interfere, producing an *interference head wave*. Here, the nondimensional parameter $\zeta = L_2(a/c_{1s})^{2/3}/\lambda_1^{1/3}$ applies,[1] where λ_1 is the wavelength in surficial sediment [that we evaluate at the center frequency of $S(f)$]; for $\zeta < 1$ the interference head wave assumes the spectral characteristics and amplitude akin to first-order head waves, i.e., the amplitude spectrum goes as $|S(f)|/f$. For $\zeta \geqslant 1$, the amplitude, like the C_0 wave, greatly exceeds that of a first-order head wave, and the spectral characteristics also follow $|S(f)|$, but to a degree that depends on ζ. The primary difference from $|S(f)|$ will be a small shift in the dominant frequency of the first-arriving interference head wave. Červený and Ravindra[1] briefly remark on this, showing graphically a maximum positive shift occurring at $\zeta \approx 2$ for spectra corresponding to increasing L_2, thereby increasing ζ. Using a PE simulation, this curious property is duplicated (Fig. 12) for a range of gradients, center frequencies f_{ctr}, and half-power bandwidths B of the source spectrum $S(f)$ for $f_{\mathrm{ctr}}/B \geqslant 1$. When the frequency shift estimates are expressed in terms of the parameter $f_{\mathrm{ctr}}\,\Delta f/B^2$, we find that for $\zeta > 0.5$ the data collapse to a single curve whose maximum also occurs in the vicinity of $\zeta \approx 2$. For $\zeta < 0.5$ it is no longer a simple function of $f_{\mathrm{ctr}}\,\Delta f/B^2$, whereas for ζ identically zero the spectral behavior is $|S(f)|/f$ (i.e., exact first-order case).

These concepts for the case of a sound speed gradient in the sediment are illustrated using PE simulations (Fig. 13), for which again the source and receiver are placed on the bottom to highlight spectral properties and avoid the mode-related issues discussed in the context of Fig. 8. In Figs. 13(a) and 13(b) the geoacoustic parameters are the same as



Symbol	a (s^{-1})	f_{ctr}(Hz)	B(Hz)
○	0.7	150	90
◇	0.7	200	90
◆	0.7	150	110
▲	0.3	150	90
□	0.3	200	90
▽	1.0	150	90
△	1.0	200	90
●	1.0	385	190
▶	1.0	385	230
▷	0.7	385	190
■	1.6	150	90
▼	0.16	200	90
◺	0.5	150	110
◣	0.5	150	90
⬠	0.5	150	80
⬟	0.1	150	110
⬠	0.1	150	90
⬟	0.1	150	80
▽	0.1	150	125
▼	0.03	385	263
⬟	0.3	385	190

FIG. 12. The frequency shift Δf in the amplitude spectrum of a PE-simulated interference head wave expressed as the parameter $f_{\mathrm{ctr}}\,\Delta f/B^2$ and plotted as a function of ζ. Values of f_{ctr}, B, and a used in the simulations are listed in the box at right, with a source-to-receiver range varying 300–2000 m.

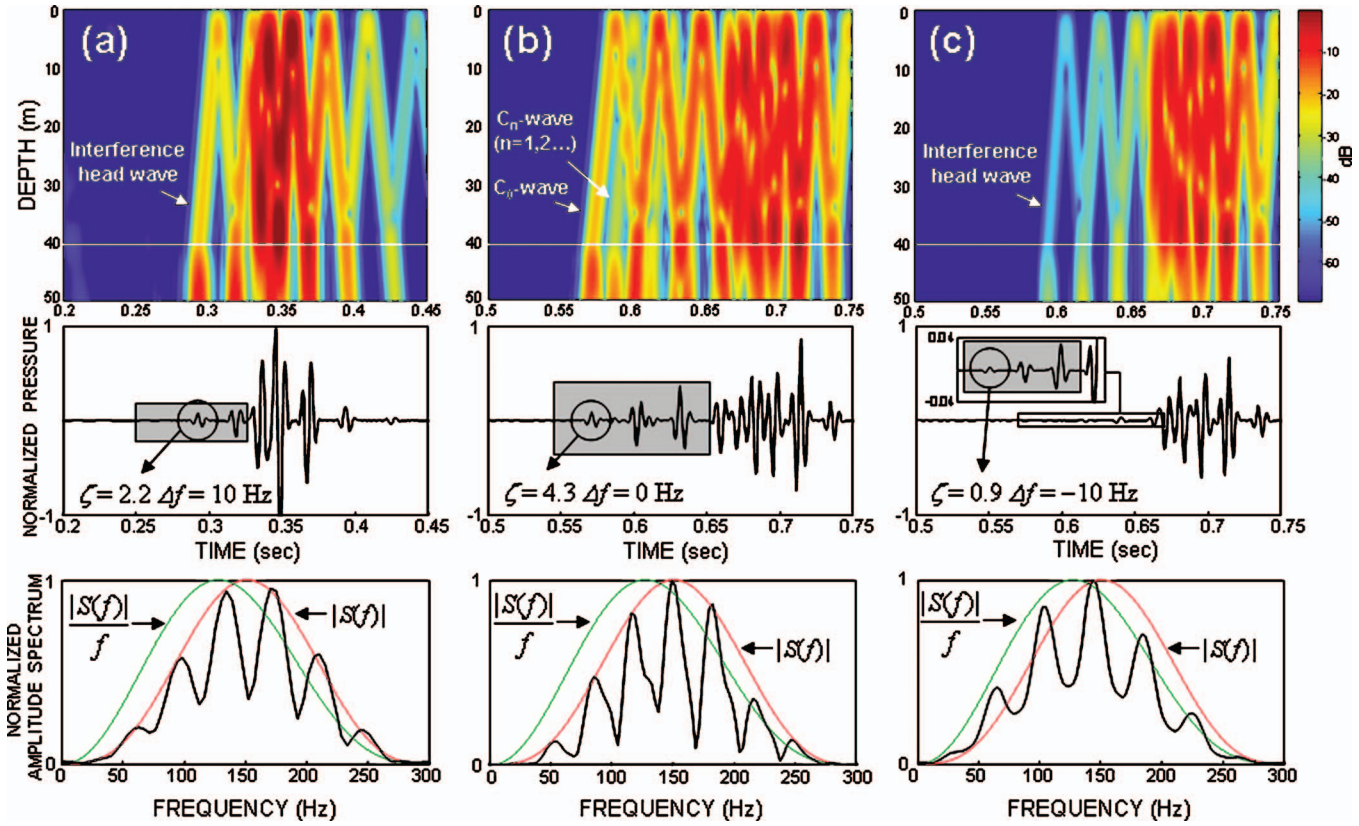


FIG. 13. PE simulated acoustic field (top), time signal for source and receiver depth of 40 m (middle), and the amplitude spectrum of the ground wave arrival (bottom) for the case of (a) $a=1.6$ s^{-1} and receiver range of 500 m, (b) $a=1.6$ s^{-1} and receiver range of 1000 m, and (c) $a=0.16$ s^{-1} and receiver range of 1000 m. (Note the expanded vertical scale.)

those used in our previous simulations involving a Pekeris waveguide (Fig. 5), with a gradient $a=1.6$ s^{-1} and c_{1s} equal to 1689 m/s. A source spectrum for which f_{ctr} and B are 150 and 90 Hz, respectively, is used in the PE simulation. For range 500 m [Fig. 13(a)] the ratio $\Delta T_{0\_1}/\delta t \approx 0.18$, indicating an interference head wave, and $\zeta \approx 2$. The signal received before 0.33 s (the onset of the water wave) is the developing ground wave composed of the interference head wave sequence. The grazing angle of the first arrival in the sequence is 29.7°, and that of the second is 28.3° (unlike the first-order counterpart, for which the grazing angle is constant and critical), and the arrival time difference T_p is 26 ms. By way of analogy to discussions of the Pekeris waveguide case in Sec. II (see Fig. 3), the frequency 19.4 Hz$(1/2T_p)$ can be identified as an effective mode 1 cutoff frequency for the waveguide at range $r=500$ m. [Here range/depth$=O(10)$ and a modal viewpoint is appropriate, e.g., see Weston and Rowlands.[14]] The frequency spectrum of the ground wave [Fig. 13(a), bottom] has (1) an envelope closely matched to the source spectrum, and (2) spectral peaks corresponding to odd multiples of 19.4 Hz. The second property is analogous to the first-order head wave sequence and the Pekeris waveguide case.

For the same waveguide, but at range 1000 m [Fig. 13(b)], the ground wave (the signal received before about 0.66 s) is more complex, owing to separated C_0 and C_n arrivals $(n=1,2,\ldots)$ that subsequently overlap after reflection from the sea surface. Continuing with the analogy to the Pekeris waveguide case, the mode 1 cutoff frequency here is determined by the geometry of the C_0 wave. For range 1000 m, the mode 1 cutoff frequency is 16.3 Hz, based on the arrival time difference (T_p) between the C_0 wave and the initial sea surface reflection C_0 wave, and the envelope of the ground wave spectrum [Fig. 13(b), bottom] is also well described by the source spectrum with peaks at odd multiples of $1/2T_p$ Hz. Defined in this manner, the mode 1 cutoff frequency (f_c) for a waveguide having a linear sound speed gradient in the sediment and constant sound speed in the water column decreases with increasing range, while falling within the bound, given by

$$\frac{c}{4h} < f_c < \frac{c}{4h\sqrt{1-(c/c_1)^2}}. \tag{4}$$

Note that inclusion of sediment attenuation increases the lower bound.

Finally, looking again at range 1000 m, but with the gradient reduced by a factor of 10 to 0.16 s^{-1} [Fig. 13(c)], an interference head wave with $\Delta T_{0\_1}/\delta t < 1$ and $\zeta < 1$ exists, and the arrival times are very similar to first order for a Pekeris waveguide with $c_1=1689$ m/s. Additionally, the spectrum [Fig. 13(c), lower] takes on first-order properties and a much reduced amplitude in the arrival structure [Fig. 13(c), middle]. Note that although the ground wave amplitude is clearly reduced compared with the case shown in Fig. 13(b), where $a=1.6$ s^{-1}, it is approximately 5 dB greater than for the equivalent Pekeris waveguide case. Further re-

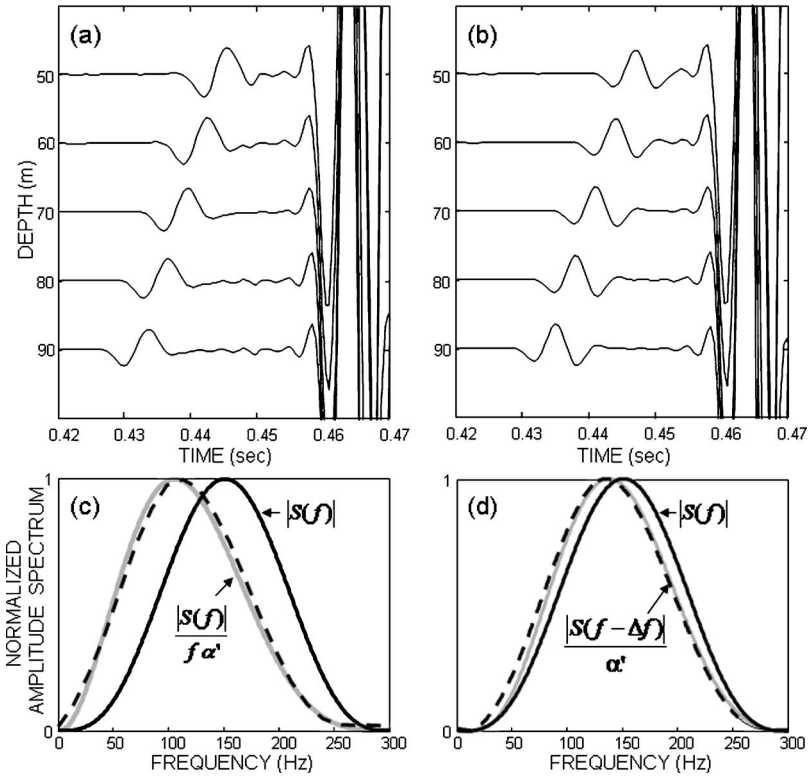


FIG. 14. PE simulated time series showing (a) first-order head wave arrivals as viewed from a vertical line array (VLA) for a Pekeris waveguide, and (b) zeroth-order interference head wave arrivals ($\zeta \approx 2$) for an equivalent gradient waveguide (see text for definition of equivalence). The water depth is 100 m, source depth is 50 m, and the VLA is at range 700 m. *The first-order head wave arrivals are amplified by a factor of 7* (17 dB) *in the region to the left of time* 0.45 s *in* (a); the amplitude of the water wave arrivals in each case are essentially equal and out of scale for this plot. The amplitude spectrum of (c) first-order head wave received at depth 90 m (dashed line), compared to a model using $\alpha_{sed}=0.1$ dB/m/kHz and $L_2=581$ m (light gray line) based on the source spectrum (black line) and (d) same for zeroth-order interference head wave, with $\Delta f = 7$ Hz used in the model function.

ducing the gradient to $0.016\ s^{-1}$ ($\zeta \approx 0.2$) puts the results nearly identical in amplitude to that in the Pekeris waveguide case.

V. SUMMARY AND DISCUSSION

Head waves, ground waves, and related precursor arrivals are important features of shallow water propagation exhibited in hydrophone data at relatively short ranges. They are directly linked with properties of the seabed, and thus provide a useful measure of seabed properties for geoacoustic inversion without the logistical effort of a longer range experiment. Although the term ground wave has been applied to different phenomena, here an interpretation of the ground wave as a head wave sequence is presented for a Pekeris waveguide, wherein the ground wave is viewed as a sequence of head waves that have undergone multiple reflections from the sea surface and seabed boundaries. For a Pekeris waveguide, the envelope of the spectrum of the ground wave arrival corresponds to the spectrum of a single (isolated) head wave arrival, and peaks in the spectrum correspond to odd multiples of the mode 1 cutoff frequency for the waveguide. In this case the head wave is first order in its ray series classification and its amplitude spectrum goes as $|S(f)|/f$, where $S(f)$ is the source amplitude spectrum. This relation is most clearly exhibited when both the source and receiver are located essentially on the bottom; in the case of either source or receiver located within the water column, the relative peak amplitudes in the ground wave spectrum corresponding to odd multiples of the mode 1 cutoff frequency are weighted by the corresponding depth-dependent modal eigenfunction. These results are illustrated using Fourier synthesis of a narrow band PE solution.

The effect of variations from a Pekeris waveguide were considered, albeit separately, for basic variations involving a downward refracting sound speed profile in the water column, sloping bottom, isospeed layering in the seabed, and a linear sound speed gradient in the seabed. (The effect of shear wave conversion was not addressed in this study, and our results apply to unconsolidated sediments where the shear speeds are quite low.[1,8]) The most notable variations involve layering and gradients. Layering in the sediment produces reflected waves that are zeroth order in their ray series classification in addition to first-order head waves propagating along interfaces. Because of this combination the amplitude spectrum follows neither the first-order characteristic of $|S(f)|/f$ nor the zeroth-order characteristic of $|S(f)|$. However, if layering is such that zeroth-order and first-order arrivals are not separated in time, the arrival structure and spectrum will be dominated by the zeroth-order contribution.

The sequence constituents of the ground wave, or isolated head waves, are strongly sensitive to the presence and strength of a sound speed gradient in the seabed. Two parameters, $\Delta T_{0\_1}/\delta t$ and ζ, are important. When both are less than unity, an interference head wave exists whose properties are akin to a first-order head wave with a spectrum going as $|S(f)|/f$, and a sequence also develops in a manner analogous to the first-order case. For $\zeta > 1$, the ground wave is composed primarily of an interference head wave ($\Delta T_{0\_1}/\delta t < 1$) sequence or a C_0 wave ($\Delta T_{0\_1}/\delta t > 1$) sequence, and the spectrum follows $|S(f)|$. For an interference head wave, the degree to which the spectrum follows $|S(f)|$ depends mildly on ζ. Also, in this case a mode 1 cutoff frequency f_c can be identified in a manner analogous to the Pekeris waveguide, with the difference that f_c decreases with increasing range while remaining within the upper (Pekeris waveguide) and lower (hard bottom, $c/4h$) bounds.

Sediment layering and a sound speed gradient in the seabed are thus two means by which arrivals that are zeroth order in ray series classification can be included in the precursor set. It is essential that these be distinguished from lower-amplitude, first-order head wave arrivals, e.g., in a physical model to use in an objective function and subsequent geoacoustic parameter estimation. An example is given in Fig. 14, which shows PE simulated arrivals for a Pekeris waveguide [Fig. 14(a)] and an equivalent gradient waveguide [Fig. 14(b)] with gradient $a=1.0\ \mathrm{s}^{-1}$. Equivalence in this case means that the horizontal arrival angle for the first arrival has been matched (26.8°) for a receiving vertical line array at range 700 m. Matching puts c_{1s} at 1689 m/s for the Pekeris waveguide case and at 1663.8 m/s for the gradient case. With a waveguide depth of 100 m, an isolated first arrival is evident and the sequence that gives rise to the ground wave has yet to develop. For the Pekeris waveguide case this arrival is a first-order head wave, and for the gradient case this arrival is a zeroth-order interference head wave ($\zeta \approx 2$), as evidenced by the much higher amplitude. In this example, a nominal sediment attenuation (α_{sed}) equal to 0.1 dB/m/kHz is used in each simulation and will be evident in the amplitude spectrum of the arrival. For inversion purposes a simple model for the Pekeris waveguide case is to divide $|S(f)|/f$ by $\alpha' = 10^{\hat{\alpha}_{\mathrm{sed}} f L_2/20000}$, where $\hat{\alpha}_{\mathrm{sed}}$ is a candidate estimate of α_{sed}, to be evaluated with an objective function, and $L_2=581$ m; for the gradient case $|S(f-\Delta f)|$, where $\Delta f=7$ Hz, must instead be used in this same operation. A satisfactory agreement is achieved [Figs. 14(c) and 14(d)] when α_{sed} is the model input.

Finally, we suggest that head wave nomenclature is improved by inclusion of the applicable ray series classification (i.e., either zeroth or first order) that, along with parameter ζ, characterizes both amplitude and spectral properties of the precursor arrival. We also find that a small shift in peak frequency for the first-arriving interference head wave, Δf, when expressed as the parameter $f_{\mathrm{ctr}}\Delta f/B^2$, where B is half power bandwidth, and f_{ctr} is center frequency of the source spectrum, is well described by a single curve that is a function of ζ. The maximum shift occurs in the vicinity of $\zeta \approx 2$.

ACKNOWLEDGMENTS

This work was funded by the Office of Naval Research Code 321 Ocean Acoustics. The authors wish to thank Dr. Robert I. Odom of APL-UW for helpful discussions.

[1] V. Červený and R. Ravindra, *Theory of Seismic Head Waves* (University of Toronto Press, Toronto, 1971), Chaps. 2, 3, 6, 7.
[2] G. V. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (PTR Prentice–Hall, New Jersey, 1994), Chaps. 4, 5.
[3] K. Aki and P. G. Richards, *Quantitative Seismology, Theory and Methods* (Freeman, San Francisco, 1980), Vol. **1**, Chap. 6.
[4] W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957), Chap. 4.
[5] One approach to the solution of the Helmholtz equation involves the ray series $p(r,f)=p_0(r)+(1/2i\pi f)p_1(r)+[1/(2i\pi f)^2]p_2(r)+\cdots$, for which the zeroth-order solution often suffices. However, a head wave associated with the interface between two isospeed layers of speed c and c_1 undergoes refraction at the critical angle, and the zeroth-order term vanishes (Ref. 1). The zeroth-order term, p_0, goes as $\sim 1/L$ with $L=[r\cot\theta_1(\mathrm{SH}/\cos^2\theta_1 + c_1\mathrm{RH}\cos\theta_1/c\cos^3\theta_2)]^{1/2}$, where θ_1 and θ_2 are the incident angle and the refracted angle, respectively, and SH and RH are source height and receiver height above the interface, respectively.
[6] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, New York, 1980), Chap. 39.
[7] C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley, New York, 1977), Chap. 9.
[8] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 2000), Chaps. 1, 2, 8.
[9] D. M. F. Chapman and D. D. Ellis, "The group velocity of normal modes," J. Acoust. Soc. Am. **74**, 973–979 (1983).
[10] O. A. Godin, N. R. Chapman, M. C. A. Laidlaw, and D. E. Hannay, "Head wave data inversion for geoacoustic parameters of the ocean bottom off Vancouver Island," J. Acoust. Soc. Am. **106**, 2540–2551 (1999).
[11] M. V. Hall, "Measurement of seabed sound speeds from head waves in shallow water," IEEE J. Ocean. Eng. **21**, 413–422 (1996).
[12] S. E. Dosso, G. H. Brooke, R. D. Huston, and J. P. Todoeschuck, "Ocean-bottom geoacoustic measurements in the ice-covered Lincoln Sea," in *Proc. Oceans '93*, Victoria, BC Canada (IEEE, New York, 1993), pp. 119–124.
[13] M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[14] D. E. Weston and P. B. Rowlands, "Guided acoustic waves in the ocean," Rep. Prog. Phys. **42**, 347–387 (1979).

The directionality of acoustic T-phase signals from small magnitude submarine earthquakes

N. R. Chapman[a)]
School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055, Victoria, B.C. V8W 3P6, Canada

R. Marrett
Defence Technology Agency, Auckland Naval Base, Auckland, New Zealand

(Received 4 August 2005; revised 24 February 2006; accepted 12 March 2006)

Acoustic transients radiated from small magnitude undersea earthquakes were studied in an experiment carried out using a towed horizontal line array operating in the South Fiji Basin. The transient signals consisted of P, S, and T phases, with the T-phase signal from each earthquake lasting up to 7 min. The directionality of the T-phase signal was determined by processing the array hydrophone data with a conventional beamformer. The weakest part of the signal arrived first on a direct bearing between the earthquake source and the array. However, subsequent stronger components arrived from different directions farther south of the source, in a region where a number of seamounts and ridges rose within the sound channel. A simple model based on ray path travel times for elastic wave travel in the earth and acoustic wave propagation in the water suggests that the later components of the T-phase signal were radiated into the water by downslope propagation from the seamounts and ridges. The initial weaker components may be scattered into the sound channel by leakage from the P and S phases relatively close to the array. [DOI: 10.1121/1.2195073]

PACS number(s): 43.30.Ma [RAS] Pages: 3669–3675

I. INTRODUCTION

Seismic transients radiated from small magnitude undersea earthquakes were studied in an experiment carried out in July 1982 using a towed horizontal line array to measure the level and directionality of ambient noise sources in the South Fiji Basin. The seismicity in this region of the SW Pacific is known to be very high from measurements recorded primarily at land based seismograph stations around the world. This is due to the presence of two subduction zones underlying the basin, one along the eastern boundary marked by the Kermadec Trench, and the other to the northwest along the Loyalty Islands chain where the plate motion is in the opposite sense.[1]

The low-frequency (<200 Hz) ambient ocean noise background in this region is significantly lower than that in the northern hemisphere, due to the bathymetric isolation of the basin from the rest of the South Pacific and the low density of local shipping.[2] As a result, the seismic transients from small magnitude earthquakes, which occur with relatively higher frequency than larger magnitude events, were observed over a broad frequency band from 10 to 90 Hz in this experiment. These events would not easily be detected in the northern oceans where the noise due to distant shipping dominates the ambient noise spectrum at low frequencies.

Seismic transients have been studied extensively over many years, initially using single hydrophones at stations near remote Pacific islands. The earthquake signal observed in the ocean usually consists of three components, the primary (P), secondary (S), and tertiary (T) phases.[3,4] The first two components are propagated as elastic waves in the earth and are converted to compressional waves in the water near the receiver, while the T phase is an acoustic signal that travels at relatively low speed in the deep ocean SOFAR (Sound Fixing And Ranging) sound channel (Fig. 1). The very low attenuation of signals propagating in the deep sound channel supports extremely efficient transmission of T-phase energy to very long ranges. Consequently, T phases provide excellent detection and localization capability for earthquake events. In recognition of these capabilities, hydroacoustic stations for T-phase detection play a prominent role in the International Monitoring System (IMS) mandated by the Comprehensive Test Band Treaty Organization.[5,6]

Due to the large impedance contrast between the oceanic crust and the water, seismic energy from an earthquake is injected into the water at very high grazing angles. The first explanation proposed for the generation of oceanic T phases was a downslope conversion mechanism at locations relatively close to the earthquake source.[3,7] In this model, the high angle bottom interacting energy that enters the water column is converted to shallow angle propagation paths that propagate in the deep sound channel to long ranges with relatively low loss.[8,9] This behavior has been observed at Hawaiian hydrophone stations for earthquakes in the North Pacific off Alaska and Japan,[7,10] and at Polynesian stations for Hawaiian events.[11] Another mechanism was proposed by Johnson *et al.* for T-phase signals that were generated in deep ocean abyssal plains.[12,13] They proposed that seismic energy from the earthquake was scattered into the sound channel at the earthquake source; the higher frequency content and

a)Electronic mail: chapman@uvic.ca

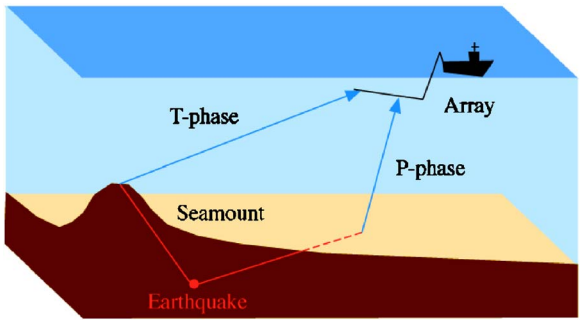


FIG. 1. (Color online) Propagation paths of P-phase and T-phase signals from a shallow marine earthquake. The T-phase signal propagates in the earth's crust and is injected into the water column at bathymetric features such as seamounts or ridges.

gradual onset of the abyssal T-phase signal was consistent with the scattering hypothesis. By comparison, the slope T phases are generally lower energy signals with sudden onsets. Park *et al.* have also proposed a mechanism based on modal scattering from a randomly rough sea bottom.[14] This mechanism can also apply in regions of nonplanar bathymetry. DeGroot-Hedlin and Orcutt have modeled the T-phase signal in terms of low order modes excited by uniformly distributed point sources on the sea floor. Their model of seismic to acoustic conversion is consistent with sea floor scattering along shallow sloping bathymetric features in the vicinity of the epicenter.[15]

In this paper, we present measurements of the signal level and directionality of acoustic T phases from shallow to intermediate depth, small magnitude undersea earthquakes in the South Fiji Basin. The measurements were made using a towed horizontal line array that provided high resolution estimates of the signal directions. A large number of small magnitude events were observed during the course of the experiment, but relatively few were reported as significant earthquakes in the seismological literature. The analysis is described here for three events for which the source locations and times were known from the literature. The directionality of the sound was determined from the horizontal line array data by comparing the measurements for the three events, which were observed at different array headings. Our results indicate that the T-phase signal was composed of several distinct components that were radiated from the directions of specific bathymetric features in the basin. This interpretation was tested using a simple model for the propagation of the earthquake signal. A similar approach using backazimuths and arrival times of T phases observed at IMS hydroacoustic triplet stations was reported by Graeber and Piserchia to map the locations of seismo-acoustic coupling for large earthquakes from the margin of the Indo-Australian plate.[16]

The paper is organized in the following way. In Sec. II we describe the experiment and the signal processing technique. We then describe the low-frequency noise background due to seismic transients, and present measurements of the signal for a typical earthquake from single hydrophone data. The beamformed array data are presented in Sec. III, and we describe the procedure that was used to determine the T-phase directionality. Finally the propagation model is described, and source mechanisms for the T-phase signal are suggested.

II. AMBIENT NOISE EXPERIMENT

The data reported in this paper were obtained in an experiment carried out in the South Fiji Basin during July 1982. Measurements of ambient ocean noise were made with a horizontal line array that was towed by the New Zealand research ship HMNZS TUI for a period of 12 days in the vicinity of a site at 24° S 176° 10′ E (Fig. 2). The array consisted of 32 hydrophones equispaced over an aperture of 300 m, so that the design frequency was 77.5 Hz. The towing depth was nominally 400 m, within the SOFAR sound channel, and the ship speed was 2–3 knots. The array had four depth sensors along the aperture, and two heading sensors, one at the head and the other at the tail. Although these data were recorded, they were used only to monitor array straightness. Instead of using the in line sensors, array headings were determined by monitoring the bearing of a signal from a 224 Hz beacon sound source that was deployed at a seamount about 400 km to the south. This signal, together with the array compass data, provided accurate information for determining the array heading to within a degree. In addition to the acoustic data, sound speed profiles were measured at the site throughout the experiment.

The directionality of the noise field was determined by beamforming the array hydrophone data. The hydrophone data were digitized at a rate of 1 kHz and processed with a cross spectral beamformer that used a discrete Fourier transform to carry out the beamforming on the cross spectral averaged fields. The spectra were estimated in 1 Hz bins over the band 0–255 Hz, and the data were averaged typically for 1 min. Up to 63 beam azimuths were used to establish the directionality from forward to rear endfire.

III. LOW-FREQUENCY NOISE BACKGROUND

An estimate of the variability of the ambient noise level at low frequency was obtained by sampling the single hydrophone data at 10 min intervals for the entire period of the experiment. The result is shown in Fig. 3 where 1 min averages of the noise level at 18 Hz are plotted versus time. The gap in the data around 16 July is due to an equipment failure, and the local increase in background level on 18–19 July is due to a passing storm. The spiky excursions of up to 10 dB from the base level of around 55 dB are generally due to seismic transients, although there are a few cases of noise bursts from mechanical pulling of the array during course changes. These latter events were easily recognized from the seismic signals by analysis of the spectral content of the noise estimate, but no attempt was made to remove them from the figure. This figure demonstrates the high level of seismic activity in this region of the SW Pacific; our analysis indicated that the frequency of seismic events was about 2–3/h during the experiment.

Only a small number of these events were also reported as significant earthquakes by the International Seismological Centre. The great majority of the seismic transients observed with the array were either aftershocks or T-phase signals from relatively small earthquakes (magnitude less than 4.0).

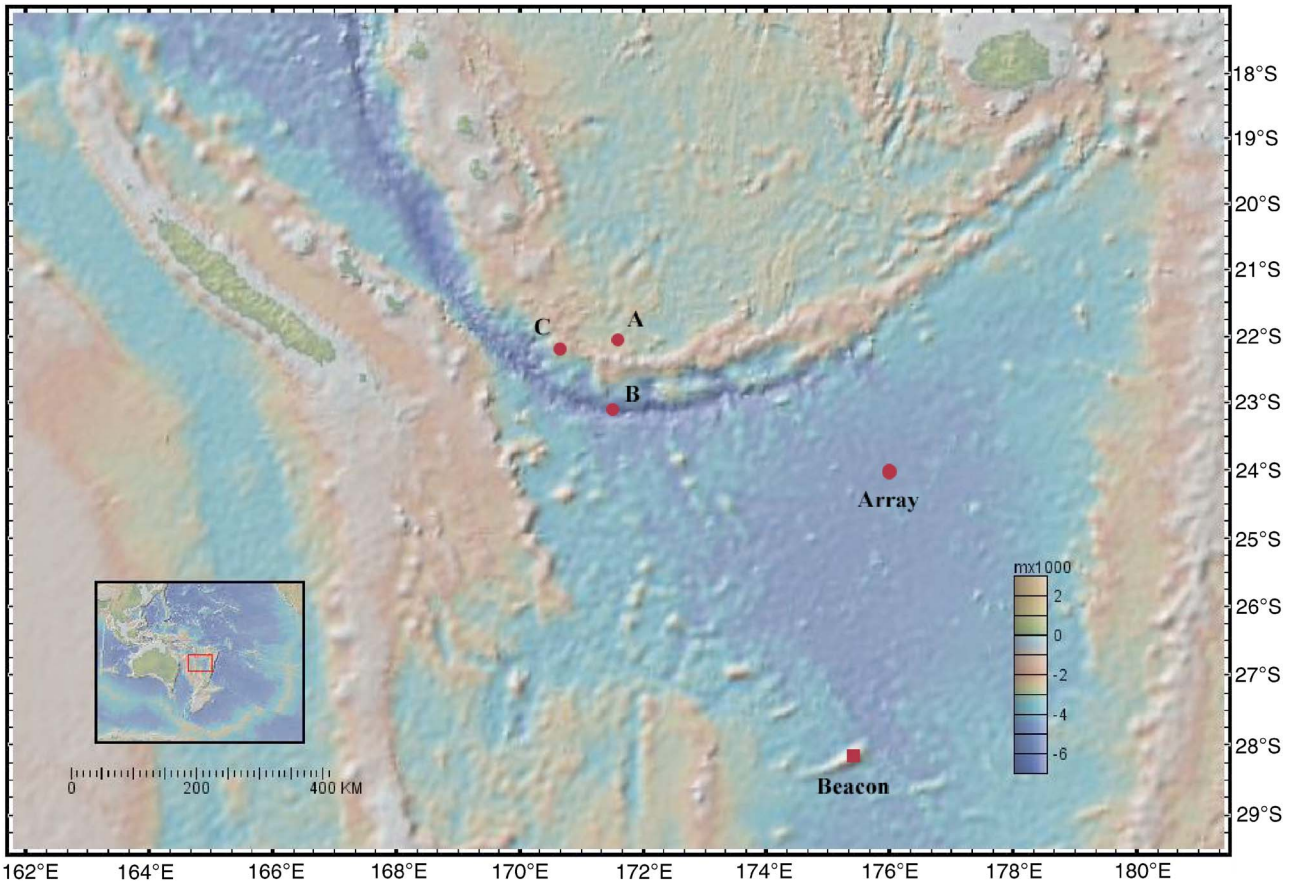


FIG. 2. (Color online) Regional chart of the experimental location (closed square) and the earthquake epicenters (circles) in the South Fiji Basin. The experimental site is at 24° S 176° 10′E.

The analysis reported in this paper for the T-phase directionality was carried out for three events that were all located in the vicinity of the New Hebrides Trench near the northwest border of the basin.[17] The locations and times of these earthquakes provided independent information that was used to establish the true directionality of the signals resolved by the array. This information, along with the magnitudes and depths of the events,[17] is listed in Table I, and the source locations are shown in the map in Fig. 2. The depths are deeper than the very shallow (<10 km) depths associated with small earthquakes at spreading centers and transform faults.[18] It should also be noted that the depth value of 33 km is assigned in cases for which the actual depth cannot be determined. The array headings and the range, bearing, and P-phase travel time of each event are listed in Table II.

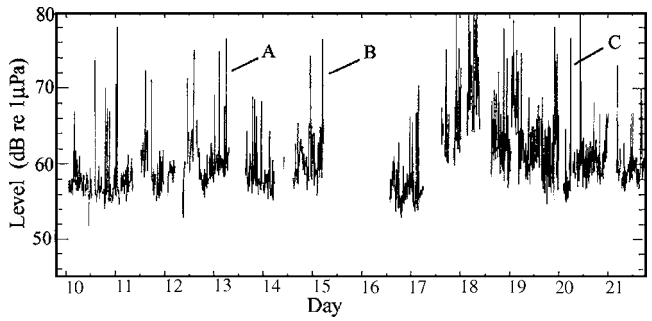


FIG. 3. Time history of the ambient noise level at 18 Hz (in dB re 1 μPa) measured during the experiment. The samples are 1 min averages of single hydrophone data.

IV. EXPERIMENTAL DATA

A. Single hydrophone data

The spectrum of the seismic transient recorded at a single array hydrophone for Event B (15 July) is plotted versus time in Fig. 4. The intensity is indicated by the color scale in dB, relative to the magnitude of the maximum intensity. The time history shown in the figure was typical of all three events listed in Table I. The signal is limited at low frequencies (<8 Hz) by flow noise at the towed array. However, the three phases of the seismic transient are evident over the band from 10 to at least 80 Hz. The duration of each of the P and S phases is 40–50 s, while the T phase begins about 105 s after the P-phase arrival, and persists for a period of about 400 s. During this time, the intensity of the

TABLE I. Earthquake source information for the three events.

Event	Epicenter location	Time (Z)	Focal depth	M_b
A (12 July Z)	22.0°S 171.5°E	12/18:42:05	33 km	<4.0
B (14 July Z)	23.2°S 171.3°E	14/17:17:27	33 km	4.7
C (19 July Z)	22.3°S 170.5°E	19/18:15:23	59 km	<4.0

TABLE II. Receiving geometry of earthquake events.

Event	Array heading (deg)	Bearing to epicenter (deg)	Range to epicenter (km)	Travel time (P phase) (s)
A (12 July Z)	112	292	529	79
B (14 July Z)	308	278	494	68
C (19 July Z)	136	284	600	73

signal varies over about 10 dB, with the maximum occurring 190–200 s after the onset of the T-phase signal, about 5 min after the arrival of the P phase. The travel time from the earthquake source to the array is about 68 s for the P phase.

B. Array beamformed data

The time history for the directionality of the seismic transient for Event B is shown in array beam space in Fig. 5. The data were filtered in the frequency band from 14 to 16 Hz and processed in beams. The vertical traces display the beam intensities in the color scale, from forward to rear endfire (left to right in the figure). Each horizontal slice is an average of 10 s of data, and the total time history in each beam is about 9 min.

The P and S phases are received about 1 min apart in the beams near broadside, which receive energy that propagates at high angles in the water. This is expected on the basis of the model shown in Fig. 1 for the propagation of these phases. The energy travels primarily in the earth at speeds characteristic of P and S waves, and is refracted into the water near the array at steep grazing angles due to the large contrast in sound speeds between the ocean bottom material and the water. Estimates of the elastic wave velocities were determined from the arrival times of the P and S waves that were recorded at the seismological station at Noumea, about 500 km to the southwest of the earthquakes.[17] The averaged velocities were 7.5 km/s for P, and 4.5 km/s for S.

The T-phase signal is initially observed in the beam at about 65° off broadside immediately following the S-phase arrivals. The signal is received over the next 6 min, and during this time there are four distinct components that are identified by changes in the signal level and bearing. The first component persists for about 60 s and is received in beams 3–4, roughly 65° off broadside. The second component is received in beam 10, about 43° off broadside, and the intensity of the signal increases by about 5 dB. The maximum of the T-phase signal occurs in the third component that is received in beams 6–7, roughly 55° off broadside, arriving about $4\frac{1}{2}$ min after the P phase. The intensity increases to about the same level as that of the P and S phases. Finally, the fourth strong component is observed in beams 9–10. A similar directional pattern for the T phases was observed for the signals from each of the three earthquakes.

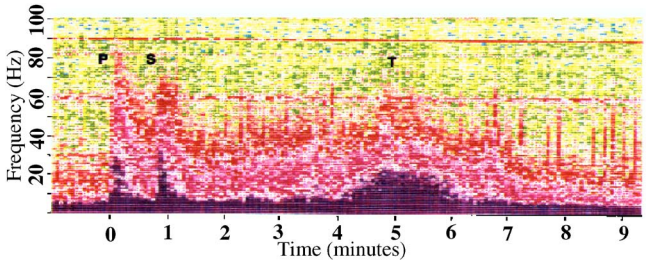


FIG. 4. (Color online) Time history of the spectral components of the seismic transient from event B measured at a single array hydrophone. The spectral data are limited by array flow noise below about 8 Hz. The T-phase signal persists for at least 7 min, with maximum intensity about 5 min after the arrival of the P phase.

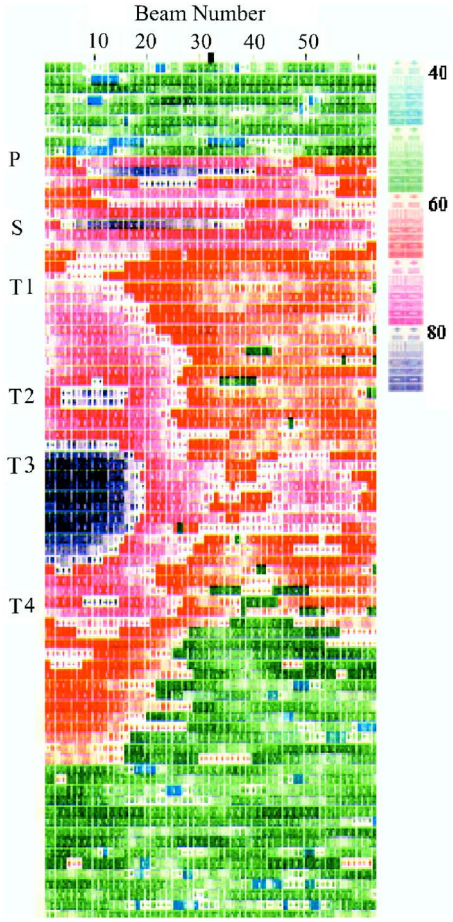


FIG. 5. (Color online) Beam intensity shown in color scale. The beams are numbered from forward endfire (beam 1) to aft endfire (beam 63). The broadside beam is beam 32. Each horizontal slice is an average over 10 s, so that the total time shown is about 9 min.

V. T-PHASE DIRECTIONALITY

The change in bearing of the T-phase signals observed in the beamformed data suggests that the signal was radiated to the array from specific locations in the basin. Since each earthquake was observed at a specific array heading for each event, the true directionality of the T-phase components cannot be obtained from the array beam data for any single event due to the intrinsic left-right ambiguity of line arrays. Instead, we have adopted a simple approach to use the data from all three events to establish the signal directionality. The method assumes that the T-phase signals are radiated from specific bathymetric features in the basin (such as ridges or seamounts), and that the propagation paths for the T phases from each earthquake are similar. This is not an unreasonable assumption, since the sources are all located in

TABLE III. Array directionality data for the T-phase components. The ambiguous bearings are also listed (in plain type) for each single event.

T-phase Component	Event A	Event B	Event C
I	**293**° ±6°/278° ±6°	**282**° ±3°/332° ±3°	**278**° ±2°/354° ±2°
II	**262**° ±3°/322° ±3°	**262**° ±3°/356° ±3°	**261**° ±2°/015° ±2°
III	**270**° ±3°/313° ±3°	**272**° ±3°/346° ±3°	**271**° ±2°/002° ±2°
IV	**264**° ±3°/320° ±3°	**261**° ±3°/358° ±3°	**262**° ±2°/010° ±2°

the same region near the New Hebrides Trench, and a similar pattern was observed in the beamformed T-phase arrival data for each event. On the basis of this approach, the left/right ambiguity can be resolved because each event was observed on a different array heading. The directionality data are listed in Table III, where the true (boldface) and ambiguous bearings are listed for each component of the T-phase signal for each event. It is clear that specific bearings are obtained consistently for each component, and the ambiguous bearing can easily be determined. We have assigned directions of 262° ±3°, 271° ±3°, and 262° ±3° for the second, third, and fourth T-phase components, respectively, and a direction along the bearing to the earthquake source for the first component.

The T-phase directionality for event B is plotted versus time in Fig. 6(a), where the reference time is the onset of the event. The intensity (in dB) is also shown [Fig. 6(b)] with respect to the same time origin. The first component, T1, arrives on a bearing of ~277°, approximately the direction to the earthquake source. The signal level increases at about 220 s, and the direction shifts southward to ~262° for the next 80 s. The maximum level of this second component, T2, occurs around 270 s. The third component, T3, arrives at about 320 s on a bearing of ~272°, and reaches a maximum signal level at ~340 s. This component persists for about 1 min, when the signal bearing shifts back to the bearing of the earthquake event ~279°. This time corresponds roughly to the travel time of a signal propagated to the array from the vicinity of the source. The signal level drops by about 5 dB from the maximum value in the third component. The fourth component, T4, appears around 450 s and persists for about 50 s. After this component, the signal level is weak, about 20 dB less than the maximum level, and the bearing is not well determined. The average value is about 275°.

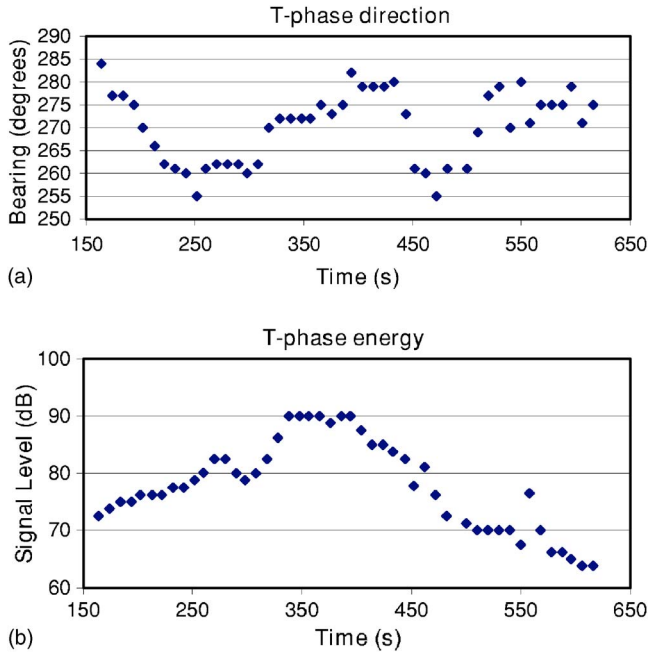


FIG. 6. (Color online) (a) Directionality of the T-phase signal from event B vs time. The origin of time is the earthquake event time. The second component, T2, arrives on a bearing of 262° at about 220 s; the third component arrives on a bearing of 272° at about 320 s, and the fourth component arrives on a bearing of 261° at about 450 s. (b) Energy of the T-phase signal from event B vs time.

These bearings indicate that the T-phase energy is radiated primarily from sources to the south of the earthquake epicenter, and not from the slopes along the subduction arc directly to the north. This result suggests different behavior from that reported for the source of T-phase energy in other subduction zone environments, for example, in the Indian Ocean where the T phases were observed to radiate along the slope between the trench and arc.[14,15] We propose in Sec. VI a simple model for the T-phase propagation to account for our observations of the directionality.

VI. T-PHASE PROPAGATION MODEL

We present in this section a simple propagation model to account for the travel time of the T-phase signals observed for the three earthquakes. The first component, which arrives from the direction of the earthquake source, propagates primarily as refracted elastic energy in the earth, is injected into the water column near the array and is scattered into the sound channel. This mechanism is similar to that proposed by Johnson *et al.* for the generation of abyssal T phases,[12] and T phases of this type have most recently been observed by Slack *et al.*[19] and Pan and Dziewonski.[20] However, in our case the signal enters the water column near the receiver; a simple calculation of the travel time indicates that this component does not propagate primarily as water borne energy that was scattered into the sound channel near the earthquake source.

We assume that the subsequent T-phase components are radiated from specific bathymetric features in the western part of the basin, south of the epicenters of the events. The source locations for these components were identified as ridges or seamounts that rise substantially into the water column, to depths as shallow as 2000–2500 m. These features lie along the bearings given in Sec. V (Table III) for the signal direction for each component, as indicated in Fig. 7.

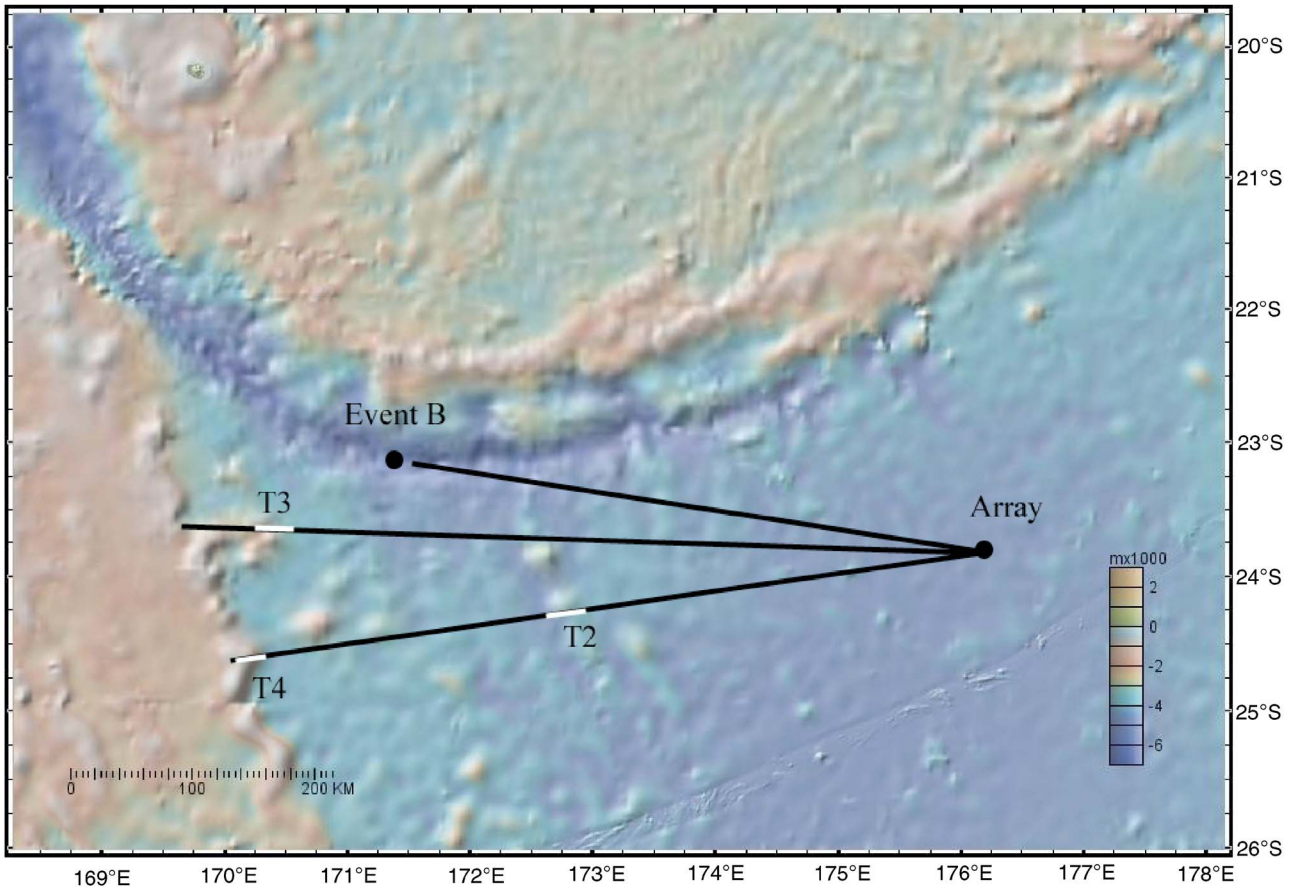


FIG. 7. (Color online) Directions of T-phase components (T2, T3, and T4) and earthquake location for event B. The signal directions for the T-phase components are distinct from the direction to the earthquake source, and extrapolate to seamounts and ridges to the south. The calculated travel times in Table IV are based on T-phase generation sites indicated by the white portions of the bearing lines.

The locations of the sea floor features that generate each of the T-phase components are indicated by the white portions of the lines along the bearings. We assume that the seismic energy is radiated into the water column at these sites. The converted acoustic waves are coupled into the sound channel, due to the sloping bottom, and propagate with low loss to long distances. The propagation model for these signal components assumes that the seismic energy travels as elastic waves in the earth from the earthquake source to the seamounts, and then travels as water-borne acoustic energy in the deep sound channel at a speed of 1.5 km/s to the array. Two options for the elastic wave propagation were tested: first as P waves propagating with speeds of ~7.5 km/s, and then as S waves propagating at ~4.5 km/s. The measured times and the predicted travel times for both models of elastic wave propagation are listed in Table IV. The predicted times are calculated based on the assumed T-phase source sites shown in Fig. 7. The simple propagation model fits the measured travel times reasonably well for all three events. Event B appears to favor T-phase coupling with S waves. However, it is difficult to determine generally whether P or S wave propagation is favored, because the travel paths in the earth are relatively short.

The surprising result of the analysis of these data is that the T-phase energy originates at bathymetric features that are located far from the earthquake source. Moreover, the locations of the features are not associated with the slope between the trench and arc of the subduction zone environment to the north of the array. However, the mechanism of seismo-acoustic energy conversion at the sites is not known. The conversion efficiency of each site is likely associated with the local bathymetric slope and the scale of the sea floor roughness.

VII. SUMMARY

The directionality of T phases from small magnitude earthquakes originating in the north west margin of the South Fiji Basin was determined to high resolution using a towed

TABLE IV. Travel times for T-phase signal components.

	T-phase component travel time (s)			
Event: Data/model	I	II	III	IV
Event A: data	170–240	295–335	340–470	460
Model: P-wave		260	405	430
Model: S-wave		280	425	445
Event B: data	155–205	220–290	310–420	440
Model: P-wave		260	315	430
Model: S-wave		270	325	450
Event C: data	175–235	235–305	335–450	470
Model: P-wave		265	400	430
Model: S-wave		295	415	450

horizontal line array. The T-phase signal from these events consisted of four distinct components that arrived on bearings that were associated with the directions to seamounts and ridges that extend into the SOFAR sound channel, south of the earthquake locations. A simple model for the travel times of the T-phase components, based on propagation from the earthquake source to the ridges and seamounts at elastic wave speeds in the earth followed by propagation from these features to the array in the sound channel, was consistent with the measured data. The model assumes that the seismic energy is coupled to the ocean efficiently at the seamount and ridge slopes. The conversion efficiency is likely related to the local sea floor slope at the conversion site.

ACKNOWLEDGMENTS

The authors want to acknowledge helpful discussions with D. Bohnenstiehl in revising this paper. His comments, and those of two anonymous referees, have greatly improved the paper.

[1]H.-W. Zhou, "Mapping of P-wave slab anomalies beneath the Tonga, Kermadec and New Hebrides arcs," Phys. Earth Planet. Inter. **61**, 199–229 (1990).
[2]R. Marrett and N. R. Chapman, "Low frequency ambient noise measurements in the South Fiji Basin," IEEE J. Ocean. Eng. **15**, 311–315 (1990).
[3]I. Tolstoy and M. Ewing, "The t-phases of shallow focus earthquakes," Bull. Seismol. Soc. Am. **40**, 25–51 (1950).
[4]D. H. Shurbet and M. W. Ewing, "T-phases at Bermuda and transformation of elastic waves," Bull. Seismol. Soc. Am. **47**, 251–262 (1957).
[5]M. Lawrence, M. Galindo, P. Grenard, and J. Newton, "Hydroacoustic monitoring system for the Comprehensive Nuclear Test Ban Treaty," Kerntechnik **66**(3), 90–95 (2001).
[6]E. A. Okal, "T-phase stations for the international monitoring system of the Comprehensive Nuclear Test Ban Treaty: A global perspective," Seismol. Res. Lett. **72**(2), 186–195 (2001).
[7]R. H. Johnson, J. Northrop, and R. Eppley, "Sources of Pacific T phases," J. Geophys. Res. **68**, 4251–4260 (1963).
[8]S. E. Dosso and N. R. Chapman, "Measurement and modeling of downslope acoustic propagation loss over a continental slope," J. Acoust. Soc. Am. **81**, 258–268 (1987).
[9]W. M. Carey, I. B. Gereben, and B. B. Brunson, "Measurement of sound propagation downslope to a bottom-limited sound channel," J. Acoust. Soc. Am. **81**, 244–257 (1987).
[10]D. A. Walker, C. S. McCreery, and Y. Hiyoshi, "T-phase spectra, seismic moments and tsunamigenesis," Bull. Seismol. Soc. Am. **82**, 1275–1305 (1992).
[11]J. Talandier and E. A. Okal, "On the mechanism of conversion of seismic waves to and from T-waves in the vicinity of island shores," Bull. Seismol. Soc. Am. **88**, 621–632 (1998).
[12]R. H. Johnson, R. A. Norris, and F. K. Dunnebier, "Abyssally generated T phases," in *The Crust and Upper Mantle of the Pacific Area*, edited by L. Knopoff, C. L. Drake, and P. J. Hart, [Am. Geophys. Union Geophysical Monograph **12**, pp. 70–78 (1968)].
[13]R. H. Johnson and R. A. Norris, "T-wave generation mechanisms," Hawaii Institute of Geophysics Report HIG-70-7, Hawaii Institute of Geophysics, University of Hawaii, 1970.
[14]M. Park, R. I. Odom, and D. J. Soukup, "Modal scattering: A key to understanding oceanic T-waves," Geophys. Res. Lett. **28**, 3401–3404 (2001).
[15]C. D. deGroot-Hedlin and J. A. Orcutt, "Synthesis of earthquake generated T-waves," Geophys. Res. Lett. **26**, 1227–1230 (1999).
[16]F. M. Graeber and P-F. Piserchia, "Zones of T-wave excitation in the NE Indian ocean mapped using avariations in backazimuth over time obtained from multi-channel correlation of IMS hydrophone triplet data," Geophys. J. Int. **158**, 239–256 (2004).
[17]Bulletin of the International Seismological Centre, Vol. **19**(7), (1982).
[18]C. G. Fox, H. Matsumoto, and T. K. A. Lau, "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," J. Geophys. Res. **106**, 4183–4206 (2001).
[19]P. D. Slack, C. G. Fox, and R. P. Dziak, "P wave detection thresholds, Pn velocity estimates, and T wave location uncertainty from oceanic hydrophones," J. Geophys. Res. **104**(B6), 13061–13072 (1999).
[20]J. Pan and A. M. Dziewonski, "Comparison of mid-oceanic earthquake epicentral differences of travel time, centroid locations and those determined by underwater hydrophone arrays," J. Geophys. Res. **110**, B07302, (2005).

A probabilistic model for the noise generated by breaking waves

Richard M. Heitmeyer[a)]
Naval Research Laboratory, 4555 Overlook Avenue, SW, Washington, DC 20375-5320

(Received 16 August 2005; revised 1 March 2006; accepted 1 March 2006)

We present a probability law on both the noise field and the beam-noise time waveforms that is based on the assumption that the breaking-wave occurrences are described by a space-time Poisson process and that the breaking-wave waveforms are independent Gaussian processes. The probability law specifies the probability density of all orders in terms of its characteristic function, which is determined as a space-time integral of a functional involving the characteristic function of the contribution of an individual breaking wave. As an illustration, we present examples of the first-order probability density and the correlation function for a noise field observed on a vertical array operating in shallow water. It is seen that for the smaller elevation angles and for deep phones, where the energy is dominated by distant breaking waves, the observed noise waveforms are essentially Gaussian processes. Conversely, for beams pointed toward the surface and for shallow phones, where the noise is dominated by a small number of breaking waves, the noise wave forms are not Gaussian. Finally, we present a measure of the Gaussianity derived from an approximation to the characteristic function and apply it to identify the elevation angles and phone depths where the Gaussian approximation is not valid.
[DOI: 10.1121/1.2188916]

PACS number(s): 43.30.Nb [EJS] Pages: 3676–3693

I. INTRODUCTION

Fundamental to the design and evaluation of statistical signal processing algorithms is a specification of the probability law on the observed noise as well as the signal processes. Such a probability law is required for such things as the calculation of Cramer Rao bounds on the location errors in matched field processing systems, see, for example, Thode *et al.*,[1] and the design of optimum likelihood ratio detectors and the calculation of their receiver operating characteristics, see, for example, Van Trees.[2]

In the application of statistical signal processing to SONAR system design, it is commonly assumed that wind-induced noise is a time-stationary Gaussian process, Makris.[3] To the extent that this assumption is valid, the full probability law is completely determined by the cross-correlation function. The spatial correlation function, itself, is conventionally determined from a "source sheet" model in which the sources are continuously distributed on a sheet located either on or slightly below the ocean surface, see, for example, Kuperman.[4] Parameters of these source models include the source strength spatial density and the spatial correlation of the source sheet elements or, alternatively, the vertical directionality of the source elements.

The modern theory of wind-induced noise holds that, at least at the higher frequencies, the noise is generated by the aggregate of the contributions from individual breaking waves. This theory is not inconsistent with the Gaussian assumption since central limit theorems assert that a sum of a large number of independent quantities is Gaussian distributed, provided that the sum is not dominated by only a few of those quantities. Under most circumstances this is a reasonable assumption. However, for receivers located near the surface or for hydrophone arrays with beams steered toward the surface, the sum can be dominated by a small number of breaking wave contributions and central limit theorems do not apply. An illustration of the burst-like nature of the noise due to breaking waves measured on the surface directed end-fire beam of a vertical array is presented in Hollett.[5]

In this paper we develop the probability law on the noise assuming a breaking-wave model where the breaking-wave occurrence times and locations are described by a space-time Poisson process. Under this assumption, it is shown that both the noise field and the beamformed noise field are filtered Poisson processes. For such processes, the joint probability density is determined in terms of its characteristic function (the multi-dimensional Fourier transform of the joint probability density) which is, itself, determined as a space-time integral of a functional involving the characteristic function of the contribution of an individual breaking wave, weighted by the space-time density of the breaking-wave occurrences. A corollary of this theory is that the moments of the breaking-wave noise process can be obtained from the classical source sheet model with a suitably chosen source strength density. A theoretical treatment of the one-dimensional filtered Poisson process in time, which includes the classical "shot noise" process, can be found in Parzen.[6] Another example of the application of filtered Poisson processes in underwater sound can be found in the reverberation model developed by Middleton.[7] Examples of the role of non-Gaussian process to represent underwater noise are found in Middleton[8] and Milne.[9]

The paper is organized as follows. In Sec. II, we specify the source model. The locations and occurrence times of the breaking waves are described by a (possibly nonhomogeneous) Poisson process. The stochastic waveforms of the

[a)]Electronic mail: heit@wave.nrl.navy.mil

sound radiated by the breaking waves, which are necessarily of finite duration, are described in terms of the two-point correlation function, or equivalently, the two-dimensional energy spectrum of the waveforms. In Sec. III we specify the resulting noise field and the time series of the output of a conventional (linear) beamformer that operates on that field. In Sec. IV, we specify the general form of the probability law on both the noise field and the beamformed noise to include the expressions for the cross-correlation function. In Sec. V, we present an example of the first-order probability density and the moments for a vertical array operating in a shallow water environment. These examples include cases where both the beamformed field and the noise field itself are not Gaussian. A summary and discussion is presented in Sec. VI.

II. THE BREAKING-WAVE SOURCE MODEL

The source model describes the locations and times of the individual breaking waves and the sound generated by each of those breaking waves. In this paper, we describe the breaking-wave occurrences as a Poisson process and the sound generated by each breaking wave as a Gaussian random waveform specified by its energy spectrum.

The Poisson process describing the breaking-wave occurrences is defined on the space-time volume $\Gamma = \Re \times T$ where the vector $\mathbf{r} \in \Re$ specifies the location of a breaking wave on the ocean surface and $\tau \in T$ specifies the occurrence time of the wave. Such a process is completely determined by a rate function $\lambda(\mathbf{r}, \tau)$ which represents the mean number of breaking-wave occurrences per unit area per unit time. In particular, the number of breaking waves that occur in the space-time region Γ is a Poisson random variable, N_Γ, with probability function,

$$\Pr[N_\Gamma = n] = \exp\{-M_\Gamma\}\frac{M_\Gamma^n}{n!}, \tag{1a}$$

where

$$M_\Gamma = \int_\Gamma \lambda(\mathbf{r}, \tau) dA(\mathbf{r}) d\tau \tag{1b}$$

is the mean number of breaking waves in Γ. Further, given that there are N_Γ breaking-wave occurrences in Γ, the locations and the times of those breaking waves, $\{(\mathbf{r}_i, \tau_i), i = 1, \ldots, N_\Gamma\}$, are independent and identically distributed with common probability density,

$$f(\mathbf{r}, \tau) = \frac{\lambda(\mathbf{r}, \tau)}{M_\Gamma} \quad \text{for } (\mathbf{r}, \tau) \in \Gamma. \tag{2}$$

Finally, given two disjoint space-time regions, the number of occurrences and the conditional densities of the locations of those occurrences are statistically independent.

The Poisson process of Eqs. (1) and (2) describes the breaking-wave occurrences as a point process that is nonhomogeneous in both space and time. This nonhomogeneity, which appears in the space-time dependence of the breaking-wave density, can be used to represent the spatial and temporal variations in the breaking-wave occurrences due to localized storms, near-shore effects, etc. In the special case where the breaking-wave occurrences are homogeneous, isotropic, and stationary, the rate function is independent of location and time, i.e., $\lambda(\mathbf{r}, \tau) = \lambda_0$, and the mean number of occurrences is $M_\Gamma = \lambda_0 A_\Re T$ where $A_\Re$ is the area of $\Re$ and T is the duration of T.

The sound generated by an individual breaking wave is represented by a monopole source which radiates a random function drawn from a stochastic process. The noise due to an ensemble of the breaking-waves with occurrences, $\{(\mathbf{r}_i, \tau_i), i = 1, \ldots, N_\Gamma\}$, is the ensemble of radiated noise waveforms, $\{s_i(t - \tau_i), i = 1, \ldots, N_\Gamma\}$, with the interpretation that $s_i(t - \tau_i)$ is the noise radiated by the ith breaking wave.

In this paper, it is convenient to assume that the radiated noise waveforms are described by a nonstationary Gaussian process. With this assumption, the joint probability density on the waveforms for an arbitrary set of observation times is completely determined by the waveform correlation function. Since the waveforms are of finite duration, this function depends on both observation times, rather than simply the difference between the observation times. Thus, the radiated noise correlation function is defined by

$$R_s(t_j, t_k) = E[s_i(t_j) s_i(t_k)], \tag{3}$$

where, $E[\cdot]$ is the expected value operator. Similarly, the spectrum of the radiated noise waveforms, which is defined as the Fourier transform of the two-dimensional (2D) correlation function, is a function of two frequencies. In this paper, we follow the convention that the 2D spectrum is defined by

$$\Lambda_s(f_1, f_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_s(t_1, t_2) e^{-i2\pi(f_1 t_1 - f_2 t_2)} dt_1 dt_2 \tag{4a}$$

so that the correlation function is given by

$$R_s(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Lambda_s(f_1, f_2) e^{i2\pi(f_1 t_1 - f_2 t_2)} df_1 df_2. \tag{4b}$$

Properties of the 2D correlation function and spectrum can be found in Papoulis.[10] Note that the definition of the spectrum in the third edition of Papoulis differs from the definition assumed here in the sign of the second argument. For later reference, we note that the 2D spectrum has the property that

$$\Lambda_s(f, f) = E[|S(f)|^2], \tag{5}$$

where $S(f)$ is the Fourier transform of $s(t)$. This follows immediately from Eq. (4a) by expressing $S(f)$ on the right-hand side of Eq. (5) in terms of the Fourier inversion formula.

A particularly simple form for the correlation function occurs for the case of broadband bursts. For this case, the radiated noise waveforms can be represented as

$$s(t) = \pi_\eta(t) s_0(t), \tag{6}$$

where $s_0(t)$ is a wide sense stationary process, $\pi_\eta(t)$ is a window function that is zero outside the interval $[0, \eta]$, and η is the burst duration. The correlation function for broadband bursts is

$$R_s(t_1, t_2) = \pi_\eta(t_1)\pi_\eta(t_2) R_0(t_1 - t_2), \tag{7a}$$

where

$$R_0(t_1 - t_2) = E[s_0(t_1)s_0(t_2)] \tag{7b}$$

and the 2D spectrum is

$$\Lambda_s(f_1,f_2) = \int_{-\infty}^{\infty} \Pi_\eta(f_1 - \alpha)\Pi_\eta(f_2 - \alpha)^* S_0(\alpha)d\alpha, \tag{8}$$

where $\Pi_\eta(f)$ and $S_0(f)$ are the Fourier transforms of $\pi_\eta(t)$ and $R_0(\tau)$, respectively. Equation (7) follows immediately from Eqs. (3) and (6). The derivation of Eq. (8) is similar to that presented on page 419 of Ref. 10.

Experimental support for the broadband burst representation can be found in the measured time-frequency resolutions of the breaking-wave noise wave forms presented in Ref. 5. These resolutions show no clear temporal variation in the frequency structure of breaking-wave noise over the duration of the event.

Experimental support for the Gaussian assumption is found in the results presented in Ding and Farmer.[11] This paper presents estimates of the probability density on the energy spectrum expressed on a decibel scale, i.e., $10\log|S(f)|^2$, at selected frequencies. These estimates are well approximated by an exponential probability density on $|S(f)|^2$ which is the probability density that results when $S(f)$ is a complex Gaussian random variable, or equivalently, when $s(t)$ is a Gaussian broadband burst.

To complete the specification of the radiated noise process, it is necessary to specify the spectrum, $S_0(f)$, and the window function, $\pi_\eta(t)$. To this end, we note from Eq. (8) that if $S_0(f)$ is approximately constant over the narrow frequency band where $\Pi_\eta(f)$ is concentrated, or equivalently, that $\pi_\eta(t)$ is slowly varying compared to $s_0(t)$, then

$$\Lambda_s(f_1,f_2) \approx S_0(f_2)\Delta(f_1 - f_2), \tag{9a}$$

where

$$\Delta(f_1 - f_2) = \int_{-\infty}^{\infty} \Pi_\eta(\alpha)\Pi_\eta(\alpha - (f_1 - f_2))^* d\alpha$$

$$= \int_{-\infty}^{\infty} \pi_\eta(t)^2 e^{-i2\pi(f_1 - f_2)t} dt. \tag{9b}$$

Substituting Eq. (9) into Eq. (5) yields

$$E[|S(f)|^2] \approx R_\pi(0)S_0(f), \tag{10a}$$

where

$$R_\pi(\xi) = \int_{-\infty}^{\infty} |\Pi_\eta(\alpha)|^2 e^{i2\pi\alpha\xi} d\alpha = \int_{-\infty}^{\infty} \pi(t)\pi(t + \xi)dt. \tag{10b}$$

In this paper, we assume that $\pi_\eta(t)$ is normalized so that $R_\eta(0) = 1$, e.g., $\pi_\eta(t) = \eta^{-1/2}$ for $t \in [0, \eta]$. With this assumption, $S_0(f)$ can be identified with measured estimates of the mean energy spectrum.

For numerical convenience, we assume that the monopole sources are located below a flat ocean surface at a common depth, z_s, which is close enough to the surface that its radiation pattern is a good approximation to that of an infinitesimal dipole located at the surface for all frequencies of interest. We can then use the experimental estimates of the probability densities on the dipole strengths in Ref. 11 to obtain a complete probabilistic model for the ensemble of radiated wave forms. As noted earlier, under the Gaussian assumption, the probability density on the spectrum level (on a linear scale) is exponential and, hence, is completely determined by the mean spectrum level.

For the numerical results presented in this paper we restrict attention to the frequency band 500–1000 Hz. For this band, several studies, Refs. 5 and 12–14 indicate that the source spectrum is approximately white so that its probability law is determined by a single probability density function. The required probability density is taken from the experimental results presented in Ref. 11. The relationship between the dipole source level and the monopole source level required to convert these results is also presented in Ref. 11. To simplify the theoretical model we take the wave-form durations to be constant with a value of 1.8 s. Measured probability densities of the wave-form durations can be found in Ding and Farmer.[15] These results, which are parametrized by wind speed, show mean duration values of 1.4 and 1.8 s with corresponding mean source levels of 94.5 and 102.0 $\mu\text{Pa}^2/\text{Hz}$ at 1 m for wind speeds between 6.1 and 12.4 m/s.

III. THE NOISE FIELD AND BEAMFORMER RESPONSE

The noise field due to the aggregate of the breaking waves is obtained as the sum of the contributions from the individual breaking waves. Thus, if $p(t,\mathbf{r}',z')$ is the noise observed at the point $(\mathbf{r}',z')$ due to a specific realization of breaking-wave occurrences, $\{(\mathbf{r}_i,\tau_i)i=1,\ldots,N_\Gamma\}$, then

$$p(t,\mathbf{r}',z') = \sum_{i=1}^{N_\Gamma} \hat{p}_i(t;\mathbf{r}',z';\mathbf{r}_i,\tau_i), \tag{11a}$$

where

$$\hat{p}_i(t;\mathbf{r}',z';\mathbf{r}_i,\tau_i) = \breve{p}_i(t - \tau_i;\mathbf{r}',z';\mathbf{r}_i) \tag{11b}$$

is the contribution from the ith breaking wave with

$$\breve{p}_i(t;\mathbf{r}',z';\mathbf{r}_i) = s_i(t) \otimes g(t,\mathbf{r}',z',\mathbf{r}_i,z_s), \tag{11c}$$

where $g(t,\mathbf{r}',z',\mathbf{r}_i,z_s)$ is the Green's function describing the acoustic propagation in the medium and "$\otimes$" denotes the convolution operation with respect to the time variable. Here, the over-symbol "˘" is used to denote the contribution from the breaking wave if it were to occur at time $t=0$; the over-symbol "^" is used to denote the contribution from the breaking wave at its actual occurrence time, $t=\tau$. This convention is followed throughout this paper. Also, we have dropped the explicit reference to the common source depth, z_s, on the left-hand side of Eqs. (11a)–(11c) to simplify the notation. This dependence will be left implicit throughout this paper.

The beamforming system uses a hydrophone array to spatially sample the pressure field and delay the phone out-

puts in time to produce an angular resolution of the pressure field. For such a system, the time waveform at the output of the beamformer is given by

$$b(t,\theta,\phi)=\sum_{n=1}^{N_p} w_n p(t-\Delta_n(\theta,\phi),\mathbf{r}'_n,z'_n), \tag{12}$$

where $\{(\mathbf{r}'_n,z'_n),n=1,\ldots,N_p\}$ are the positions of the N_p hydrophones, θ and ϕ are the azimuth and elevation resolution angles, respectively, and $\Delta_n(\theta,\phi)$ and w_n are the time-delay and the spatial shading coefficient for the nth phone. For a specific realization of the breaking-wave occurrences, the beam waveform can be written as a superposition of the contributions from the individual breaking waves by substituting Eq. (11) into Eq. (12) and interchanging the order of the summations. The result is

$$b(t,\theta,\phi)=\sum_{i=1}^{N_\Gamma} \hat{b}_i(t;\theta,\phi;\mathbf{r}_i,\tau_i), \tag{13a}$$

where

$$\hat{b}_i(t;\theta,\phi;\mathbf{r}_i,\tau_i)=\breve{b}_i(t-\tau_i,\theta,\phi;\mathbf{r}_i) \tag{13b}$$

is the beamformed contribution from the ith breaking wave with

$$\breve{b}_i(t;\theta,\phi;\mathbf{r}_i)=s_i(t)\otimes b_0(t;\theta,\phi;\mathbf{r}_i) \tag{13c}$$

and

$$\breve{b}_0(t;\theta,\phi;\mathbf{r}_i)=\sum_{n=1}^{N_p} w_n g(t-\Delta_n(\theta,\phi);\mathbf{r}'_n,z'_n;\mathbf{r}_i,z_s). \tag{13d}$$

In the following, it will be useful to have a common notation to express both the pressure field and the beamformer response. To this end, let δ denote the spatial observation points, $(\mathbf{r}',z')$, for the pressure field and the beam direction, (θ,φ), for the beamformer output. Further, let $x(t,\delta)$ denote either the observed pressure field or the observed beam response and let $\hat{x}(t,\delta;\mathbf{r},\tau)$ denote the contribution to $x(t,\delta)$ from a breaking wave at $(\mathbf{r},\tau)$. Then for a specific breaking-wave realization, both the pressure field and the beam response can be written in the form

$$x(t,\delta)=\sum_{i=1}^{N_\Gamma} \hat{x}_i(t;\delta;\mathbf{r}_i,\tau_i), \tag{14a}$$

where

$$\hat{x}_i(t;\delta;\mathbf{r}_i,\tau_i)=\breve{x}_i(t-\tau_i;\delta;\mathbf{r}_i) \tag{14b}$$

with

$$\breve{x}_i(t;\delta;\mathbf{r}_i)=s_i(t)\otimes v(t;\delta;\mathbf{r}_i), \tag{14c}$$

where

$$v(t;\delta;\mathbf{r}_i)=\begin{cases} g(t;\mathbf{r}',z';\mathbf{r}_i,z_s) & \text{pressure field} \\ b_0(t;\theta,\phi;\mathbf{r}_i)=\sum_{n=1}^{N_p} w_n g(t-\Delta_n(\theta,\phi);\mathbf{r}'_n,z'_n;\mathbf{r}_i,z_s) & \text{beam response.} \end{cases} \tag{14d}$$

For a *fixed realization* of the breaking-wave occurrences, the breaking-wave response, $x(t,\mathbf{r}',z')$, is a zero-mean Gaussian process since $s_i(t)$ is zero-mean Gaussian and hence, $\hat{x}_i(t;\mathbf{r}',z';\mathbf{r}_i,\tau_i)$ is zero-mean Gaussian. Furthermore, since the radiated noise waveforms are statistically independent as well as zero-mean, the space-time correlation function of the breaking-wave response is simply the sum of the space-time correlation functions of the individual breaking-wave contributions. Thus, for a fixed occurrence realization, the probability law on the pressure field for an arbitrary set of space-time points is completely determined by the correlation function of the breaking-wave contribution. This function is defined by

$$R_{\hat{x}}(t_j,t_k;\delta_j,\delta_k;\mathbf{r}_i,\tau_i)=E[\hat{x}_i(t_j;\delta_j;\mathbf{r}_i,\tau_i)\hat{x}_i(t_k;\delta_k;\mathbf{r}_i,\tau_i)], \tag{15}$$

where δ_j and δ_k are the two arbitrary observation points. Using the convolution property of the 2D correlation function, this function can be written as

$$R_{\hat{x}}(t_j,t_k;\delta_j,\delta_k;\mathbf{r}_i,\tau_i)=R_{\breve{x}}(t_j-\tau_i,t_k-\tau_i;\delta_j,\delta_k;\mathbf{r}_i), \tag{16a}$$

where

$$R_{\breve{x}}(t_j,t_k;\delta_j,\delta_k;\mathbf{r}_i)=R_s(t_j,t_k)\otimes v(t_j;\delta_j;\mathbf{r}_i)\otimes v(t_k,\delta_k;\mathbf{r}_i). \tag{16b}$$

Furthermore, it follows from the convolution property that the 2D spectrum of the breaking-wave contribution can be written as

$$\Lambda_{\hat{x}}(f_j,f_k;\delta_j,\delta_k;\mathbf{r}_i,\tau_i)=e^{-i2\pi(f_1-f_2)\tau_i}\Lambda_{\breve{x}}(f_j,f_k;\delta_j,\delta_k;\mathbf{r}_i), \tag{17a}$$

where

$$\Lambda_{\breve{x}}(f_j,f_k;\delta_j,\delta_k;\mathbf{r}_i)=\Lambda_s(f_j,f_k)V(f_j;\delta_j;\mathbf{r}_i)V(f_k;\delta_k;\mathbf{r}_i)^* \tag{17b}$$

and $V(f;\delta;\mathbf{r})$ is the Fourier transform of $v(t;\delta;\mathbf{r})$,

$$V(f;\delta;\mathbf{r}) = \begin{cases} G(f;\mathbf{r}',z';\mathbf{r}) & \text{for the pressure field} \\ B_0(f;\theta,\varphi;\mathbf{r}) & \text{for the beamformer output.} \end{cases} \tag{17c}$$

Thus, the correlation function of the contribution can be written in terms of the spectrum as

$$R_{\check{x}}(t_j,t_k;\delta_j,\delta_k;\mathbf{r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Lambda_s(f_j,f_k)V(f_j;\delta_j;\mathbf{r})V(f_k;\delta_k;\mathbf{r})^* \times e^{i2\pi(f_jt_j-f_kt_k)}df_jdf_k. \tag{18}$$

IV. THE PROBABILITY LAW ON THE NOISE FIELD AND THE BEAM RESPONSE

As seen in the preceding section, for a fixed, known, realization of the breaking-wave occurrences, the pressure field and the beam response are zero-mean Gaussian processes and the time-frequency resolution is a complex zero-mean Gaussian process. In this section, we describe the probability law on the pressure field and the beam response that results for all possible breaking-wave occurrence realizations. It is this probability law that governs the observed noise.

The full probability law is specified by the *joint* probability density on the breaking wave response $x(t,\delta)$ for an arbitrary set of observation times and spatial points or beam directions. To specify this probability density, we first introduce notation to represent the random vector whose components are the breaking wave responses for an arbitrary set of observation coordinates. To this end, let $\mathbf{t}=(t_1,\ldots,t_J)$ be a J dimensional vector of distinct observation times with $t_1<t_2,\ldots,<t_J$ and let $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_J)$ be the J dimensional vector with components $\delta_j=(r_j,z_j)$ for the pressure field and $\delta_j=(\theta_j,\varphi_j)$ for the beamformer response. Then for each choice of $\mathbf{t}$ and $\boldsymbol{\delta}$, the observed response is the J dimensional random vector $\mathbf{x}(\mathbf{t},\boldsymbol{\delta})=(x_1,\ldots x_J)$ where $x_j=x(t_j,\delta_j)$. To relate the observed response vector to the individual breaking wave contributions, we define the J dimensional random vectors by $\hat{\mathbf{x}}_i(\mathbf{t},\boldsymbol{\delta},\mathbf{r}_i,\tau_i)=(\hat{x}_i(t_1;\delta_1;\mathbf{r}_i,\tau_i),\ldots,\hat{x}_i(t_J;\delta_J;\mathbf{r}_i,\tau_i))$ and $\check{\mathbf{x}}_i(\mathbf{t};\boldsymbol{\delta};\mathbf{r}_i)=(\check{\mathbf{x}}_i(t_1;\delta_1;\mathbf{r}_i),\ldots,\check{x}_i(t_J;\delta_J;\mathbf{r}_i))$ where the components of these vectors are the breaking wave responses defined by Eq. (14). Then it follows from Eq. (14) that the observed response vector is given by

$$\mathbf{x}(\mathbf{t},\boldsymbol{\delta}) = \sum_{i=1}^{N_\Gamma} \hat{\mathbf{x}}_i(\mathbf{t},\boldsymbol{\delta};\mathbf{r}_i,\tau_i) \tag{19a}$$

and

$$\hat{\mathbf{x}}_i(\mathbf{t};\boldsymbol{\delta};\mathbf{r},\tau) = \check{\mathbf{x}}_i(\mathbf{t}-\tau;\boldsymbol{\delta};\mathbf{r}), \tag{19b}$$

where $\mathbf{t}-\tau=(t_1-\tau,\ldots,t_J-\tau)$.

In this paper, we specify the joint probability density on the observed response in terms its characteristic function. To this end, let $p_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{x})$ be the Jth order probability density on the J dimensional random vector $\mathbf{x}(\mathbf{t},\boldsymbol{\delta})$ and let

$$\varphi_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{u}) = E[\exp\{i\mathbf{u}\cdot\mathbf{x}(\mathbf{t},\boldsymbol{\delta})\}] \tag{20}$$

be the Jth order characteristic function, where $\mathbf{u}=(u_1,\ldots,u_J)$. Then $\mathbf{p}_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{x})$ is determined from $\varphi_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{u})$ by the J dimensional Fourier transform,

$$p_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathrm{x}) = (2\pi)^{-J}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\varphi_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{u})\exp\{-i\mathbf{u}\cdot\mathbf{x}\}d\mathbf{u} \tag{21}$$

The expression for the characteristic function is derived in Appendix A. There it is shown that

$$\varphi_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{u}) = \exp\left\{\int_{\Re}\left[\sum_{j=1}^{J}\int_{t_{j-1}}^{t_j}\lambda(\mathbf{r},\tau) \times(\varphi_{\check{\mathbf{x}}_{j,J}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau;\boldsymbol{\delta};\mathbf{r})-1)d\tau\right]dA(\mathbf{r})\right\} \tag{22a}$$

where

$$\varphi_{\check{\mathbf{x}}_{j,J}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau;\boldsymbol{\delta}_{j,J};\mathbf{r}) = E\left[\exp\left\{i\sum_{k=j}^{J}u_k\check{x}_k(t_k-\tau;\delta_k;\mathbf{r})\right\}\right], \quad \boldsymbol{\tau}\in[t_{j-1},t_j) \tag{22b}$$

$\mathbf{t}_{j,J}-\tau=(t_j-\tau,\ldots,t_J-\tau)$, $\mathbf{u}_{j,J}=(u_j,\ldots,u_J)$, and $\boldsymbol{\delta}_{j,J}=(\delta_j,\ldots,\delta_J)$ with $t_0=-\infty$. In addition, it is shown that the correlation function of the breaking-wave noise is given by

$$R_x(t_j,t_k;\delta_j,\delta_k) = \int_{\Re}\int_{-\infty}^{t}\lambda(\mathbf{r},\tau) \times R_{\check{x}}(t_j-\tau,t_k-\tau;\delta_j,\delta_k;\mathbf{r})d\tau dA(\mathbf{r}) \tag{23}$$

where $R_{\check{x}}(t_l,t_k;\delta_l,\delta_k;\mathbf{r})$ is the breaking-wave contribution correlation function given by Eq. (18)

In the special case of Gaussian waveforms,

$$\varphi_{\check{\mathbf{x}}_{j,J}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau;\boldsymbol{\delta}_{j,J};\mathbf{r}) = \exp\left\{-\frac{1}{2}\sum_{l=j}^{J}\sum_{k=j}^{J}u_lu_kR_{\check{x}} \times(t_l-\tau,t_k-\tau;\delta_l,\delta_k;\mathbf{r})\right\}. \tag{24}$$

Thus, for Gaussian waveforms the full probability law is determined by the contribution correlation function.

The function $\boldsymbol{\varphi}_{\check{\mathbf{x}}_{j,J}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau,\boldsymbol{\delta}_{j,J};\mathbf{r})$ in Eq. (22b) is the characteristic function of an individual breaking-wave contribution conditioned on the location and the time at which that breaking wave occurs. Thus, Eq. (22a) states that the characteristic function for the pressure field or the beamformer output is obtained by integrating the characteristic functions for an individual breaking-wave contribution over the relevant space-time region with respect to the breaking-wave occurrence density. Note that this result is a consequence of the Poisson assumption which describes the

breaking-wave occurrences as discrete space-time events. It does not result from a continuum approximation on those occurrences.

If the breaking-wave occurrence density is independent of time, $\lambda(\mathbf{r},\tau)=\lambda(\mathbf{r})$, then the probability law on the breaking-wave noise is *strictly* stationary in the sense that the probability density of all orders depends on the observation times only through their differences $\xi_{j,k}=t_k-t_j$. This follows by writing $t_k-\tau=t_j-\tau+\xi_{j,k}$ in the arguments of Eq. (22b) and then making the change of variable $\beta=t_j-\tau$ in each integral of Eq. (22a). The resulting integrals are independent of the absolute time provided that $\lambda(\mathbf{r},\tau)=\lambda(\mathbf{r})$ so that the full characteristic function, $\varphi_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{u})$ and hence, $p_{\mathbf{x}(\mathbf{t},\boldsymbol{\delta})}(\mathbf{x})$ depend only on the time differences. Note that in the special case, $J=1$, the first-order probability density is completely independent of time.

In practical terms, the breaking-wave density is strictly stationary provided that the "nominal" meteorological/oceanographic conditions over the region do not change over periods comparable to the total observation period, t_J-t_1. This follows from the fact that the breaking-wave density represents the *mean* number of breaking waves per unit time per unit area and, hence, is independent of time over periods during which the nominal conditions are constant. The stationarity result may seem counter-intuitive in that large broadband bursts can be clearly seen in a time-frequency resolution of the noise field, see Ref. 5. Nevertheless, the results are plausible in a probabilistic sense since the times at which those bursts occur are not known *a priori*.

To conclude this section, we relate the space-time correlation function of the breaking-wave noise, Eq. (23), to the spatial correlation function of the source sheet model of the classical theory, see, for example, Ref. 4. To this end, we assume that the breaking-wave sound pressure, $s_i(t)$, is a broadband burst given by Eq. (6) with the correlation function given by Eq. (7) and the spectrum given by the approximation of Eq. (9). Here, we are making the assumption that the breaking-wave window function is slowly varying compared to $\nu(t;\delta;\mathbf{r})$. The reasonableness of this assumption will be discussed in the context of the Greens function plots presented in the following section. Finally, we assume that the breaking-wave density is independent of time. With these assumptions, it is shown in Appendix B that the space-time correlation function is given by

$$R_x(\xi_{j,k};\delta_j,\delta_k)=R_\pi(\xi_{j,k})R_{xvs}(\xi_{j,k};\delta_j,\delta_k), \tag{25a}$$

where

$$R_{xvs}(\xi;\delta_j,\delta_k)=\int_{-\infty}^{\infty}C_{xvs}(f;\delta_j,\delta_k)\exp\{-i2\pi f\xi\}df, \tag{25b}$$

$$C_{xvs}(f;\delta_j,\delta_k)=\int_{\Re}\lambda(\mathbf{r})V(f;\delta_j;\mathbf{r})V(f;\delta_k;\mathbf{r})^*S_0(f)dA(\mathbf{r}) \tag{25c}$$

and

$$R_\pi(\xi)=\begin{cases}0, & |\xi|>\eta\\ \displaystyle\int_\xi^\eta \pi_\eta(\beta')^2d\beta', & |\xi|<\eta.\end{cases} \tag{25d}$$

According to Eq. (25), the space-time correlation function is the product of a time correlation function, $R_\pi(\xi)$, that depends only on the breaking-wave window function, $\pi_\eta(\xi)$, and the space-time correlation function, $R_{xvs}(\xi;\delta_j,\delta_k)$, which is the correlation function determined from the spatial cross-spectrum, $C_{xvs}(f;\delta_j,\delta_k)$, of Eq. (25c). The spatial cross spectrum itself can be identified with the cross-spectrum of the classic source sheet model by equating the source strength to the product $\lambda(\mathbf{r})S_0(f)$ and assuming that the source elements are statistically independent. This is intuitively reasonable because the product $\lambda(\mathbf{r})S_0(f)$ has the units of the number of breaking waves times the energy per unit area and we have assumed that the contribution of the individual breaking waves are statistically independent. Note that the spatial correlation function obtained by setting $\xi=0$ in Eq. (25a) is the spatial correlation function $R_{xvs}(0;\delta_j,\delta_k)$ since, by assumption, $R_\pi(0)=1$. Also, note that the temporal correlation function obtained from Eq. (25) for a common observation point, $\delta_j=\delta_k=\delta$, is determined from the spectrum

$$S_{xvs}(f;\delta)=\int_{\Re}\lambda(\mathbf{r})S_0(f)|V(f;\delta;\mathbf{r})|^2dA(\mathbf{r}). \tag{26}$$

In Sec. V, we present examples of the first-order probability law and the correlation function for both the pressure field and the beam response.

V. APPLICATION TO A VERTICAL ARRAY IN A HORIZONTALLY STRATIFIED MEDIUM

The examples presented in this section are obtained for a horizontally stratified, shallow-water, acoustic environment for a frequency band of 500–1000 Hz. The beam response examples were computed for two beams from a vertical array: the surface-endfire beam, which interrogates the energy generated by the waves that break directly above the array; and a "discrete-mode" beam, which interrogates the energy generated by distant breaking waves that propagates to the array through the discrete-mode component of the acoustic field. The pressure field examples were computed for two hydrophone depths: a near-surface depth which is dominated by the nearby breaking waves and a mid-water column phone which is not. To assist in the interpretation of the results, we also illustrate the properties of an angle-time-frequency realization of the breaking-wave noise which was presented in Finette and Heitmeyer.[16] This noise realization was obtained for the same acoustic environment, the same vertical array, and the same breaking-wave noise process as the ensemble property examples presented here. To provide a background for the results, we first describe the properties of the acoustic propagation functions. We then describe the angle-time-frequency noise realization. Finally we present the cumulative distribution functions and the correlation functions.

Propagation functions: The acoustic environment of interest here is the winter profile, hard bottom case from Ref.

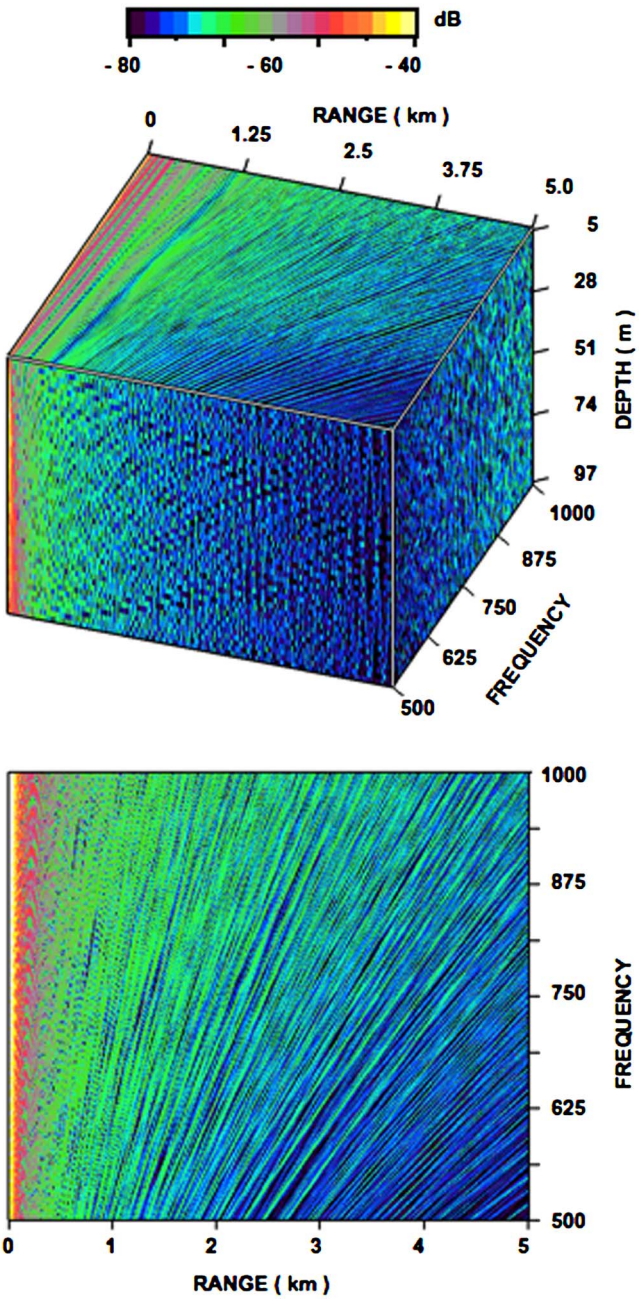


FIG. 1. (Color online) Acoustic field intensity: (top) range-frequency depth distribution; (bottom) range-frequency distribution at a depth of 50 m.

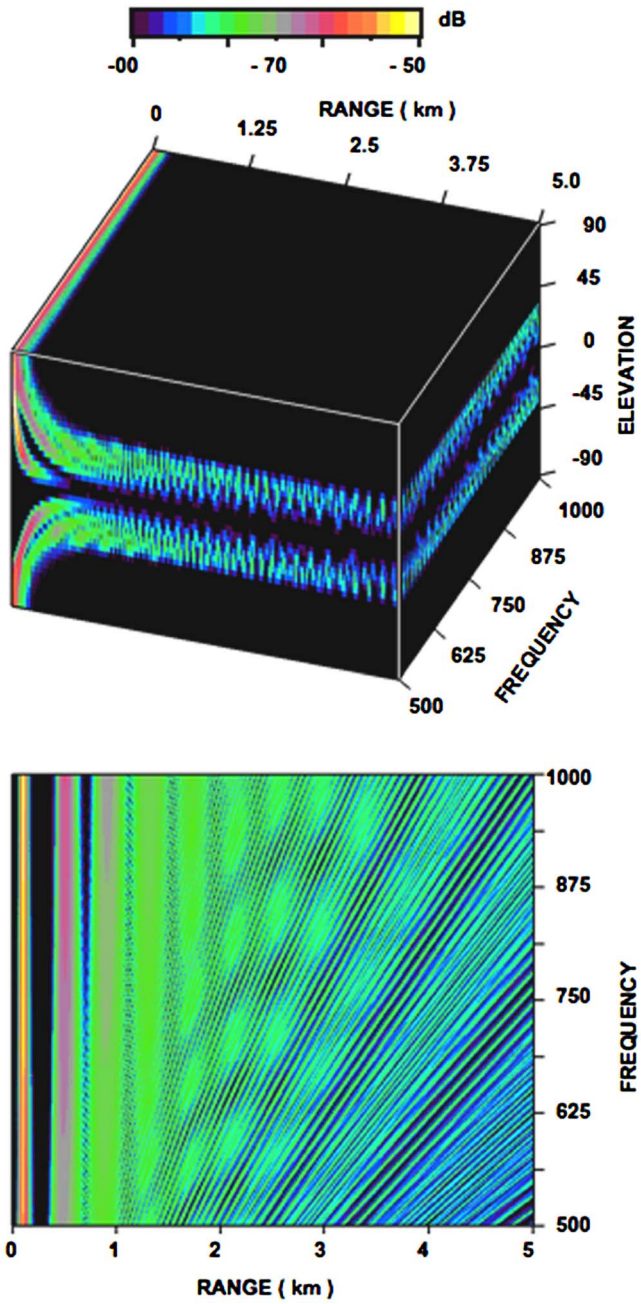


FIG. 2. (Color online) Beam intensity: (top) range-frequency-elevation distribution; (bottom) range-frequency distribution at an elevation of 26°.

16. This environment consists of a two layer medium where the sound speed in the 100 m deep water column is constant (1493 m/s) and the geoacoustic parameters are: density $\rho = 2.0$ g/cm; compressional speed, $c_p = 1800$ m/s and compressional attenuation coefficient, $\alpha_p = 2$ dB/λ. As in Ref. 16, the hydrophone array consists of 45 equally spaced phones distributed over a vertical aperture of 30 m centered in the water column. Hann shading was used in the beam forming. For later reference, we note that the endfire beam pointing toward the surface subtends a circular patch with a radius of about 20 m at the highest frequency in the 500–1000 Hz band and about 40 m at the lowest frequency.

The propagation functions for the acoustic field and the beam response are illustrated in Figs. 1–3. For the stratified-medium, vertical-array scenario, the propagation is the same in all directions and all of the phones are located along the depth axis at the same radial distance, which we take to be zero. Thus, the Greens function depends on position only through the radial distance to the breaking wave, $r_i = |\mathbf{r}_i|$, and the hydrophone depth z'; e.g., $G(f;\mathbf{r}',z';\mathbf{r}_i) = G(f;z';r_i)$. Furthermore, the beam propagation function depends only on the elevation angle, ϕ, and not on the azimuth, θ; e.g., $B_0(f;\theta,\phi;\mathbf{r}) = B_0(f;\phi;r)$. As in Ref. 16, a Fast Field Program[17] was used to compute the Green's function transform.

Figures 1 and 2 show the propagation transform functions for the acoustic field, $V(f;\delta;r) = G(f;z';r)$, and the beam response, $V(f;\delta;r) = B_0(f;\phi;r)$, respectively. As in

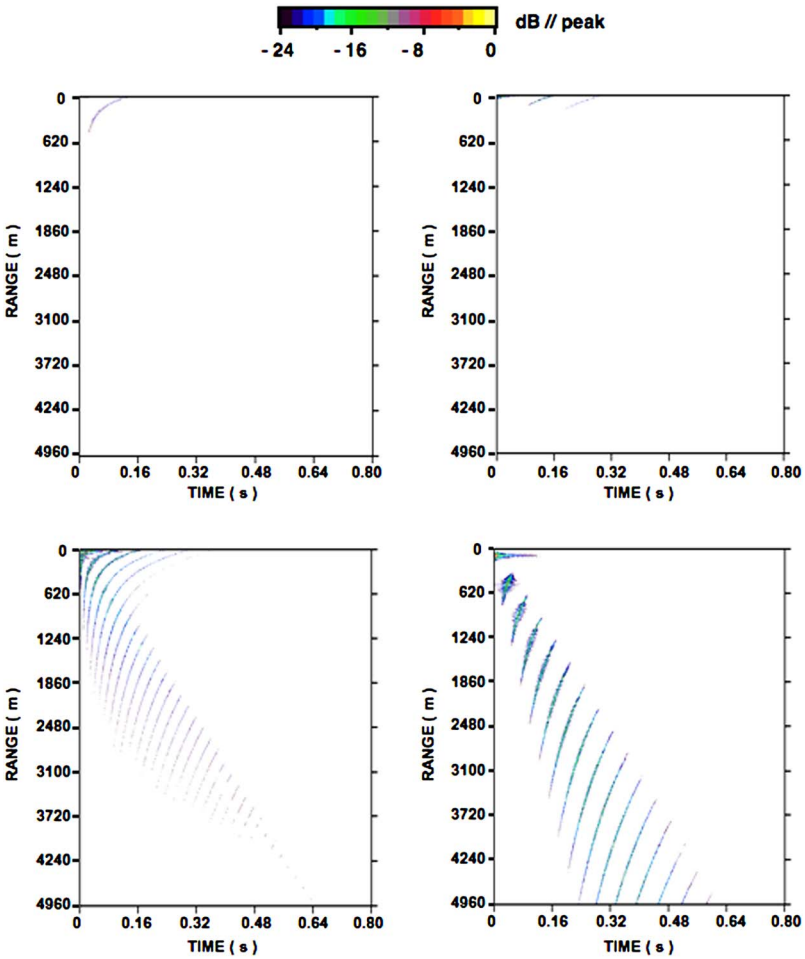


FIG. 3. (Color online) Acoustic field and beam response Greens functions vs source range: (top left) acoustic field 5 meter depth; (bottom left) acoustic field 50 m depth; (top right) beam response surface-endfire beam, (bottom right) beam response 45° above horizontal.

Ref. 16, these functions have been computed for a range interval of 5 km in 10 m increments, starting at 1 m, and a frequency band of 500–1000 Hz in 1 Hz increments. (In Ref. 16, the first range point was at 10 m and the frequency interval was 2 Hz.) The Greens function transform is shown in Fig. 1. The top plot shows the range-frequency-depth distribution over a depth interval from 5 to 97 m in 2.16 m increments. The upper face of this plot shows the range-frequency distribution at the near-surface depth of interest, 5 m. The bottom plot shows the range-frequency distribution at the mid-water depth of interest, 50 m. A detailed comparison of these two surfaces indicates that the energy at the near-surface depth at short ranges is greater than that at the mid-water depth at short ranges and less than that at the mid-water depth at the longer ranges. This is not surprising since at the short ranges, the energy falls off according to spherical spreading and, hence, is less for the deeper phone; whereas, at the long ranges, the energy is governed by cylindrical spreading and, hence, for the isovelocity profile of interest here, the deep receiver has added energy due to the mid-water sampling of the low order modes.

The beam propagation transform is shown in Fig. 2. The top plot shows the range-frequency-elevation distribution over the same range and frequency intervals as the Greens function transform and the full 180° elevation angle sector with +90° pointing toward the surface and −90° pointing toward the bottom. Accordingly, the upper face of this plot shows the range-frequency distribution for the surface-endfire beam of interest. As expected from geometrical considerations, the only significant contributions to this beam come from sources at small ranges which lie directly above the array. The bottom plot shows the range-frequency distribution for a beam at 26°. In contrast to the surface-endfire beam, this surface shows significant energy over the full range interval. This is expected since this beam points in a direction where the energy from the discrete-mode contributions arrive.

The propagation waveform functions for the acoustic field, $v(t;\delta;r)=g(t;z';r)$, and the beam response, $v(t;\delta;r)=b_0(t;\phi;r)$ are shown in Fig. 3. These waveforms have been obtained from the corresponding range-frequency surfaces shown in the preceding two plots as the inverse Fourier transform at each range with an upper frequency of 2048 Hz and, hence, a time resolution of 0.49 ms. In this figure, we have translated the waveforms by the nominal propagation time (r/c_0) to facilitate the display over the full range interval on the same time axis and we have plotted the absolute value of the waveform on a decibel scale to facilitate the display of the low level parts of the waveforms.

Figure 3 shows the range distribution of the Greens function waveforms for the acoustic field and for the beam response. The acoustic field Greens functions are shown in the left-hand plots for the two depths of interest; the near-

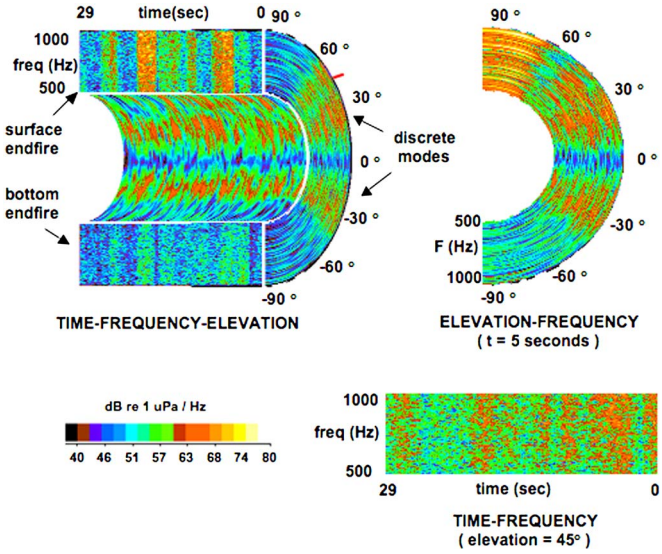


FIG. 4. (Color online) Breaking-wave beam-noise volume: (top left) time-frequency-elevation resolution; (top right) elevation-frequency resolution at $t=8$ s; (bottom right) time-frequency resolution at $\phi=45°$.

surface, 5 m depth (top left) and the mid-water, 50 m depth (bottom left). As seen in these plots, the multipath spread (the time span of significant arrivals) for all source ranges lie within 0.15 s for the near-surface receiver depth and within 0.4 s for the mid-water depth. That the multipath structure for the near-surface depth falls off faster with range is consistent with the energy fall-off seen in the corresponding frequency dependent Greens functions.

The beam response Greens functions are shown on the right-hand plots for the two elevation angles of interest; the surface-endfire, 90° beam (top right) and the discrete-mode, 26° beam (bottom right). As seen in these plots, the multipath spread for the beam waveforms is comparable to or less than that of the Greens function waveforms for all source ranges. Note that the multipath structure for the surface-endfire beam falls off even faster with source range than the multipath structure for the near-surface depth seen in the top left plot of Fig. 3. This is expected given the strong angular dependence of the endfire beam. On the other hand, the multipath structure for the discrete-mode beam falls off slower with source range than that of the mid-water waveform.

Angle-time-frequency noise realization: The elevation-time-frequency noise realization, taken from Ref. 16, was itself computed from a realization of the breaking-wave space-time occurrences obtained assuming an homogeneous Poisson process out to a range of 5 km with a breaking-wave density, $\lambda_0=2.46\times10^{-4}$ m^{-2} s^{-1}, which corresponds to a wind speed of 20 knots.[12] The total number of breaking waves in the realization used in these examples is approximately 5.7×10^5. The realization was computed using a short-term Fourier transform over a 28 s interval at a temporal resolution of 0.25 s obtained using a Hann window of the same duration.

The noise realization is shown in Fig. 4 in terms of the full angle-time-frequency volume along with an angle-frequency and a time-frequency surface obtained as cuts through the volume at $t=8$ s and $\phi=45°$, respectively. The upper and the lower faces of the noise volume are the time-frequency surfaces for the surface-endfire beam ($\phi=90°$) and the bottom-endfire beam ($\phi=-90°$), respectively. The right face of the volume is the angle-frequency surface for a time of about 3 s. Finally, the inner semi-cylinder of the noise volume is the angle-time surface for the lowest frequency in the band.

An inspection of the upper face of the noise volume in Fig. 4 shows that the noise contributions from the individual breaking waves are clearly resolved on the surface-endfire beam. This is not surprising due to the small number of breaking waves that occur in the circular patch directly above the array during the 28 s observation period. The time-frequency surface on the lower right of the plot indicates that even at ($\phi=45°$), individual breaking-wave contributions are still resolved. The reduced resolution in this surface is due to the fact that the 45° beam subtends a larger area on the sea surface and, hence, there are a larger number of breaking-wave events during the observation interval.

The angle-frequency structure of the noise realization is seen in the right hand face of the noise volume and in the angle-frequency cut on the upper right of the plot. The former corresponds to a time ($t=3$ s) when there are no breaking-wave contributions clearly evident on the surface-endfire beam; the latter to a time ($t=8$ s) when one of the two strong breaking-wave events is observed on the endfire beam. Both of these surfaces show two high noise regions in the vicinity of ±45°. The high noise region in the downward looking directions ($\phi<0°$) shows a sharp cutoff at −34°, which is approximately the critical angle for this environment. Thus, most of the noise in the high-noise region near −30° is due to the discrete-mode propagation of noise from waves that break at some distance from the array. Furthermore, since the vertical arrival structure for this environment is approximately symmetrical about the horizontal, a comparable amount of noise in the high noise region around +30° must be due to distant breaking-wave contributions. The noise in the remainder of the downward looking directions (−34° to −90°) arrives through continuous mode propagation and, hence, is largely due to the energy from nearby breaking waves that has interacted with the bottom.

The angle-time structure of the noise is seen on the inside semi-cylinder of the noise volume. This surface indicates that there can be considerable temporal variability in the noise both in the discrete-mode regions (±30°) and in the upward looking directions above the +30° discrete-mode region. The latter appears as a small number of large noise contributions; the former as a comparatively large number of smaller noise contributions. The temporal variability for the remaining directions (near horizontal and the downward looking directions below the −30° discrete-mode region) is considerably less.

Ensemble results: The first-order probability density and the correlation function examples presented here are obtained under the assumption that the breaking-wave sound pressure, $s_i(t)$, is a broadband burst given by Eq. (6) with $\pi_\eta(t)=\eta^{-1/2}$ from a Gaussian process with an energy spectrum that is flat over the 500–1000 Hz frequency band, i.e., $S(f)=S_0$. We further assume that the breaking-wave occur-

rence process is homogeneous over the 5 km radius area and the time interval of interest, i.e., $\lambda(r,\tau)=\lambda_0$. Finally, we make the additional assumption that the "bandwidth of the breaking-wave window function, $\Pi_\eta(f)$, is small compared to the frequency variation of the propagation transform $V(f;\delta;\mathbf{r})$. This last assumption is equivalent to the assumption that the duration of the window function, $\pi_\eta(t)$, is long compared to the multipath spread of the propagation waveform $\nu(t;\delta;\mathbf{r})$. This assumption is consistent with the multipath spreads observed in the plots of Fig. 3.

With these assumptions the noise space-time correlation function is determined from Eq. (25) by setting $R_\pi(\xi)=1-(|\xi|/\eta)$ in Eq. (25a) and setting $\lambda(\mathbf{r})=\lambda_0$ and $S_0(f)=S_0$ in Eqs. (25c) and (26). Furthermore, it is shown in Appendix B that the characteristic function of the noise amplitude, normalized to unit variance, can be written as

$$\varphi_x(u,\delta)=\exp\left\{(\eta\lambda)\int_{\Re}\times\left(\exp\left\{-\frac{u^2}{2}\frac{\sigma_v^2(\delta,\mathbf{r})}{\sigma_x^2(\delta)}\right\}-1\right)dA(\mathbf{r})\right\}, \tag{27a}$$

where

$$\sigma_v^2(\delta,\mathbf{r})=S_0\int_{f_1}^{f_2}|V(f;\delta;\mathbf{r})|^2df \tag{27b}$$

is the variance of the contribution of an individual breaking wave and

$$\sigma_x^2(\delta)=(\lambda\eta)\int_{\Re}\sigma_v^2(\delta;\mathbf{r})dA(\mathbf{r}) \tag{27c}$$

is the variance of the noise itself. The characteristic function for the normalized amplitude is determined by replacing u by u/σ_x in the expression for the un-normalized characteristic function. It is these equations that are used to compute the results described in the following.

We consider first the cumulative distribution function, the indefinite integral of the probability density. Figure 5 shows the cumulative distributions for both the beam response and the pressure field. For comparison purposes, we have normalized the noise amplitude to unit variance. The values of the variances relative to $S_0\eta^{-1}$, expressed on a dB scale, are shown in the legend on each plot. For reference purposes, we have also plotted the CDF of a unit variance Gaussian CDF as the black curve in each plot.

The cumulative distributions for the beam amplitudes are shown in the top plot. The CDF for the discrete-mode beam at 26°, shown as the blue curve, is indistinguishable from a zero-mean, unit variance Gaussian CDF. On the other hand, the CDF for the surface-endfire beam at 90° differs significantly from the Gaussian CDF. This is expected from the angle-time frequency plot shown in Fig. 4. Note that the variance for the discrete-mode beam is more than 8.7 dB larger than the variance of the surface-endfire beam.

The cumulative distributions for the phone amplitudes are shown in the bottom plot. As with the discrete-mode beam amplitude, the mid-water phone amplitude CDF is indistinguishable from the Gaussian CDF. The CDF for the near-surface phone shows a modest departure from the Gaussian CDF. In contrast to the beam amplitude, the departure from the Gaussian form for the near-surface phone is accompanied by an increase in the variance of 3.8 dB over the variance of the near Gaussian mid-water phone amplitude.

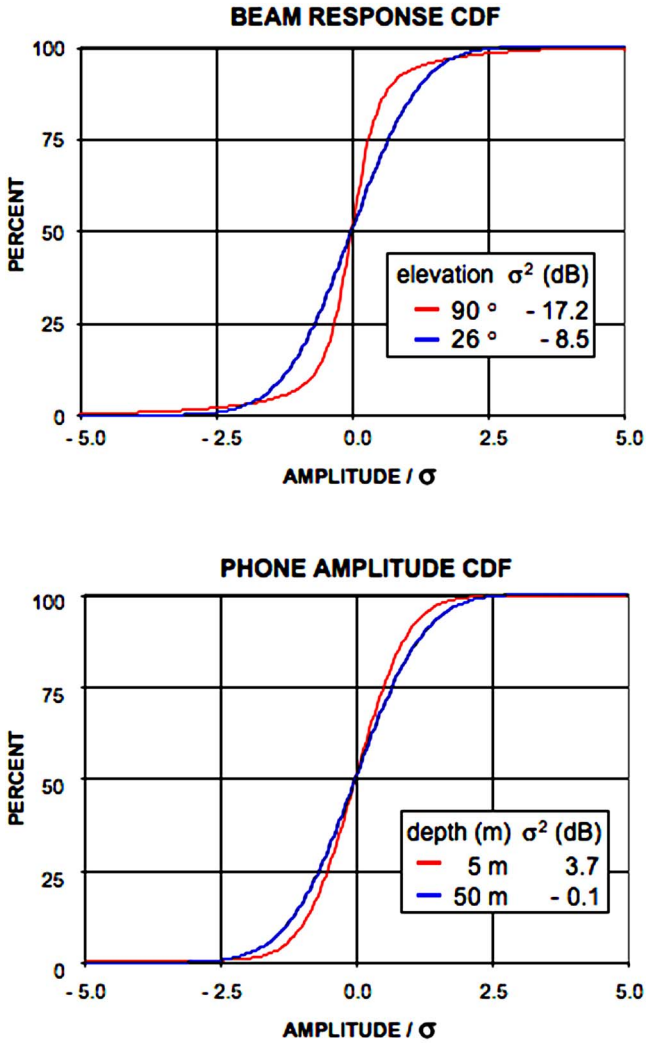


FIG. 5. (Color online) Cumulative distribution functions of the normalized amplitude: (top) beam amplitude for the surface-endfire beam (90°) and the discrete mode beam (26°); (bottom) phone amplitude for a 5 m phone depth and a 50 m phone depth. The variances relative to $S_0\eta^{-1}$ are shown in the ledger.

To provide a measure of the appropriateness of the Gaussian probability density as a function of the beam look-direction or the hydrophone depth, we develop an approximation to the characteristic function of Eq. (27). To this end, we first expand the exponential in the integrand of Eq. (27a) in an infinite series, substitute the result into Eq. (27a), and use the fact that

$$(\eta\lambda)\int_{\Re}\frac{\sigma_v^2(\delta;\mathbf{r})}{\sigma_x^2(\delta)}dA(\mathbf{r})=1$$

to obtain

$$\varphi_x\left(\frac{u}{\sigma_x};\delta\right)=\exp\left\{-\frac{u^2}{2}\right\}\Psi\left(\frac{u^2}{2};\delta\right), \tag{28a}$$

where

$$\Psi\left(\frac{u^2}{2};\delta\right)=\exp\left\{(\eta\lambda)\int_{\Re} r\left(\frac{u^2}{2};\delta;\mathbf{r}\right)dA(\mathbf{r})\right\} \tag{28b}$$

with

$$r\left(\frac{u^2}{2};\delta;\mathbf{r}\right)=\sum_{n=2}^{\infty}\frac{(-1)^n}{n!}\left(\frac{u^2}{2}\right)^n\left(\frac{\sigma_v^2(\delta;\mathbf{r})}{\sigma_x^2(\delta)}\right)^n. \tag{28c}$$

We then expand $\Psi(u^2/2;\delta)$ in an infinite series and distribute the integral across each term. The result can be written as

$$\varphi_x\left(\frac{u}{\sigma_x};\delta\right)=\exp\left\{-\frac{u^2}{2}\right\}\left[1+\psi\left(\frac{u^2}{2};\delta\right)\right], \tag{29a}$$

where

$$\psi\left(\frac{u^2}{2};\delta\right)=\sum_{p=1}^{\infty}\frac{1}{p!}\left(\sum_{n=2}^{\infty}\frac{(-1)^n}{n!}\left(\frac{u^2}{2}\right)^n\gamma_n\right)^p \tag{29b}$$

with

$$\gamma_n=\frac{(\eta\lambda)\int_{\Re}\sigma_v^2(\delta;\mathbf{r})^n dA(\mathbf{r})}{\sigma_x^2(\delta)^n}. \tag{29c}$$

The characteristic function will be well approximated by the Gaussian characteristic function [the first factor in Eq. (29a)] provided that $\psi(u^2/2;\delta)$ is small. A necessary condition for this to be the case is that the first term in Eq. (29c), γ_2, be small. Numerical examples suggest that if $\gamma_2<3$, then the density is well-approximated by a Gaussian. Plots of $(\gamma_2/3)^{1/4}$ as a function of the elevation angle for the beam response and depth for the acoustic field are presented in Fig. 6. It is seen that $\gamma_2<3$, $((\gamma_2/3)^{1/4}<1)$, occurs for beam angles greater than about 50° and for depths greater than about 7 m.

The normalized correlation functions (covariance functions) are shown in Fig. 7. For reference purposes, we have also plotted the exponential covariance function for a correlation time of unity as the black curve. As seen in the plots, there is essentially no difference between the surface-endfire and the discrete-mode covariance functions (top plot) and essentially no difference between the near-surface and the mid-water covariance functions (bottom plot). Further, as seen in the ledgers, the e^{-1} correlation times for the two beam covariance functions are the same, 0.8 ms, as are the correlation times for the two hydrophone amplitude covariance functions which are only a tenth of a millisecond smaller, 0.7 ms. The only noticeable difference between the two sets of covariance functions is the slightly larger tails in the phone amplitude covariance functions. This difference is not significant because of the very low levels involved.

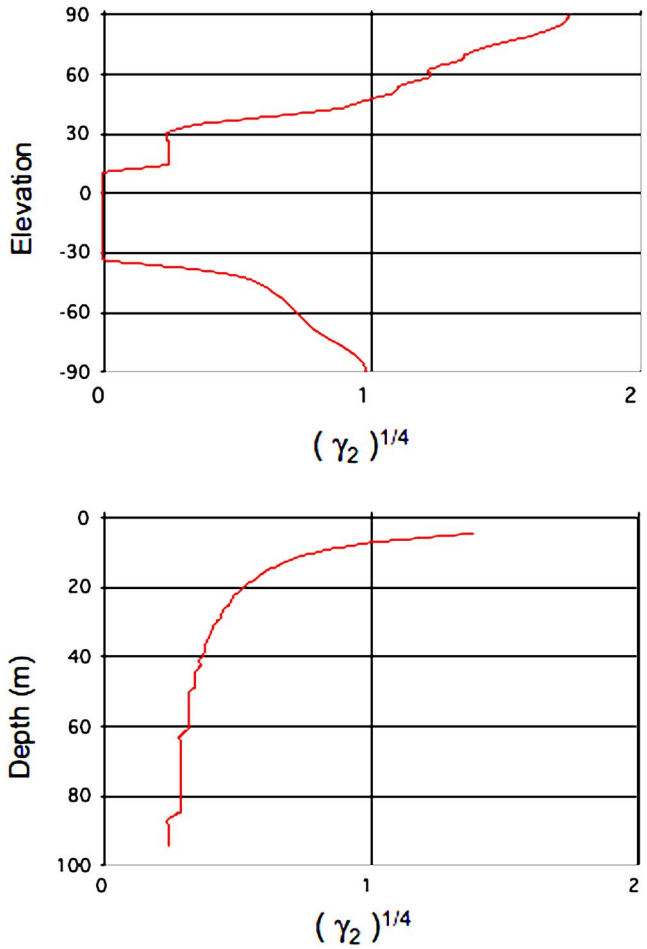


FIG. 6. (Color online) Gaussianity measure: (top) beam response, (bottom) acoustic field.

VI. SUMMARY AND DISCUSSION

In this paper, we have presented a probability law on both the noise field and the beam-noise time series that is based on the assumptions that the breaking-wave occurrence times and locations are described by a space-time Poisson process and that the breaking-wave waveforms are independent, nonstationary, Gaussian processes. The Poisson process is determined by space-time breaking-wave density which represents the mean number of breaking waves per unit area per unit time and the Gaussian waveform processes are determined by their two-point correlation function, or equivalently, their two-dimensional energy spectrum. The probability law determines the probability density of all orders in terms of its characteristic function, which itself is determined as a space-time integral of a functional involving the characteristic function of the contribution of an individual breaking wave, weighted by the space-time density of the breaking-wave occurrences. It was seen that when the breaking-wave density is independent of time over the time interval of interest, the noise process is time-stationary. Accordingly, the first-order probability density, and the moments derived from it, are independent of time. Furthermore, the correlation function depends only on the time difference and not on the absolute times. A corollary of the model is that the moments of the breaking-wave noise process can be obtained from a source-sheet model with a source-strength density computed from Eq. (25) in Sec. IV.

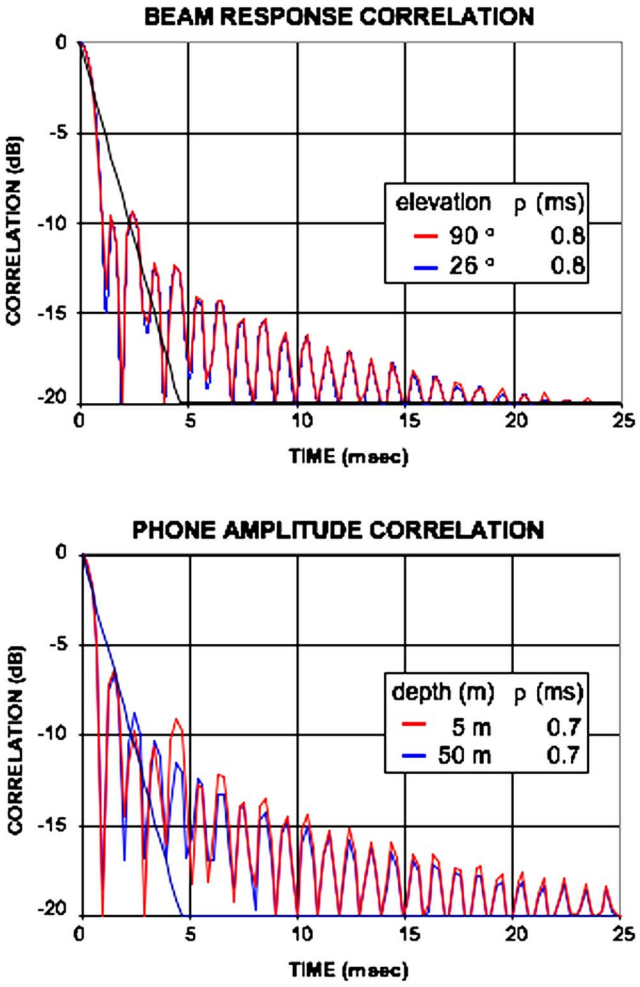


FIG. 7. (Color online) Covariance functions of the amplitude: (top) surface-endfire beam (90°) and the discrete mode beam (26°); (bottom) 5 m phone depth and a 50 m phone depth. The black curve is the exponential covariance function with unit correlation time. The e^{-1} correlation times are shown in the ledger.

To illustrate the properties of the model, we presented examples of the first-order probability density and the moments for a noise field observed on a vertical array operating in a shallow water environment. The beam response examples were computed for two beams, the surface-endfire beam which interrogates the energy generated by the waves that break directly above the array; and a "discrete-mode" beam which interrogates the energy generated by distant breaking waves that propagates to the array through the discrete-mode component of the acoustic field. The noise field examples were computed for two hydrophone depths: a near-surface depth which is dominated by the nearby breaking waves and a mid-water column phone which is not.

It was seen that for the discrete-mode beam and the mid-water phone, the cumulative distribution functions are indistinguishable from a Gaussian distribution and the e^{-1} correlation times are comparable to a few sample intervals. Thus, for these beam directions and water-depths, the noise is well-approximated by a Gaussian process as expected from central limit theorem arguments. On the other hand, for the surface-endfire beam and the near-surface water depth, the cumulative distribution function differs significantly from the Gaussian distribution, exhibiting a significantly larger high-amplitude tail. This is expected since the noise is dominated by a comparatively small number of breaking waves that occur infrequently. The correlation times, however, are comparable to those of the discrete-mode beam and the mid-water phone cases.

Finally, we presented a measure of the Gaussianity derived from an approximation to the characteristic function. An application of this measure to the specific examples indicates that the probability density is well-approximated by a Gaussian density for elevation angles greater than about 50° and for depths greater than about 7 m.

The examples were obtained for the near-constant (winter) sound speed profile and "hard-bottom" geoacoustic parameters used to compute on of the beam responses described in Ref. 16. In that paper, the dependence of the beam response on the sound speed profile and the geoacoustic parameters was examined through comparisons with results obtained for a downward refracting (summer) sound speed profile and for "soft-bottom" geoacoustic parameters. It was seen that the surface endfire noise is only weakly dependent on the sound speed profile and the geoacoustic parameters. This is not surprising since the breaking wave contributions that dominate this noise are from waves that break in the vicinity of the array and, hence, the received energy for these contributions is not strongly dependent on the acoustic environment. Accordingly, for this component of the noise, it is reasonable to expect that the probability density and the correlation function would also not be strongly dependent on the acoustic environment. On the other hand, it was also seen in Ref. 16 that the in the discrete-mode directions the *mean* noise power is strongly dependent on the environment with the sound speed profile controlling the width and depth of the noise notch and the geoacoustic parameters controlling the mean power levels. The results presented here indicate that for the winter-profile/hard-bottom case, the noise in the discrete mode directions is Gaussian distributed as would be expected from the central limit theorem since the noise is the sum of a large number of breaking wave generated contributions. Since the discrete mode noise in other acoustic environments is also the sum of a large number of contributions, it is reasonable to expect that the noise waveforms for these environments would also be Gaussian, with the only significant environmental effect being to control the variance of the probability density (the mean power) in the manner described in Ref. 16. The space-time structure of the acoustic field was not addressed in Ref. 16. Nevertheless, the same reasoning used to postulate the dependence of the beam response probability law on the acoustic environment, would suggest that the form of the probability law on the near-surface acoustic field waveforms would also be only weakly dependent on the acoustic environment and that the deep-receiver noise would be Gaussian for all environments with the mean power determined by the details of the environment. Verification of these hypotheses would require computing the probability density for a set of representative acoustic environments such as those found in Ref. 16. These computations would also serve to assess the environmental

dependence of the specific receiver-depths/elevation-angles where the probability density changes form.

ACKNOWLEDGMENT

The author is grateful for the financial support of The Office of Naval Research, Naval Research Laboratory Base Program.

APPENDIX A: THE PROBABILITY LAW

In this Appendix we derive the expressions for the characteristic functions and for the correlation function [Eqs. (22) and (23) in the text].

The expressions for the characteristic function follow from the following lemma.

Lemma 1: Let N_Γ be the number of breaking waves in the space-time region $\Gamma = \Re \times \mathrm{T}$ and let $(\mathbf{r}_k, \tau_k)$ be the positions and times of those breaking-wave occurrences. Assume that the breaking waves are distributed according to a Poisson process with rate function $\lambda(\mathbf{r}, \tau)$, and that the observed vector is given by Eq. (19) in the text. Then the characteristic function of the observed vector, Eq. (20), is given by

$$\varphi_{\mathbf{x}}(\mathbf{u}) = \exp\left\{ \int_\Gamma \lambda(\mathbf{r}, \tau)(\varphi_{\hat{\mathbf{x}}}(\mathbf{u}; \mathbf{r}, \tau) - 1) d\tau dA(\mathbf{r}) \right\}, \quad \text{(A1a)}$$

where

$$\varphi_{\hat{\mathbf{x}}}(\mathbf{u}; \mathbf{r}, \tau) = E[\exp\{i\mathbf{u} \cdot \hat{\mathbf{x}}(\mathbf{r}, \tau)\}]. \quad \text{(A1b)}$$

Proof: For notational convenience we denote the positions and times, $(\mathbf{r}_k, \tau_k)$, by $\mathbf{q}_k$, the breaking-wave density $\lambda(\mathbf{r}, \tau)$ by $\lambda(\mathbf{q})$, and the differential element $dA(\mathbf{r})d\tau$ by $d\mathbf{q}$. The proof is obtained by expanding the expected value operator in Eq. (20) in terms of expected value operators conditioned on the number of breaking waves in the region and using Eq. (19) to represent $\mathbf{x}(\mathbf{t}, \boldsymbol{\delta})$ for each of the conditional expected value operators. This results in

$$\begin{aligned}
\varphi_{\mathbf{x}}(\mathbf{u}) &= \Pr[N = 0] \\
&\quad + \sum_{n=1}^{n=\infty} E_{\mathbf{q}_1, \ldots, \mathbf{q}_n, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | N=n}[\exp\{i\mathbf{u} \cdot \mathbf{x}(\mathbf{t}, \delta)\} \cdot] \\
&\quad \times \Pr[N = n] = \Pr[N = 0] + \sum_{n=1}^{n=\infty} E_{\mathbf{q}_1, \ldots, \mathbf{q}_n, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | N=n} \\
&\quad \times \left[\exp\left\{ i\mathbf{u} \cdot \sum_{k=1}^{n} \hat{\mathbf{x}}_k(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k) \right\} \right] \Pr[N = n],
\end{aligned}$$

where $\Pr[N=n]$ is given by Eq. (1a) and $E_{\mathbf{q}_1, \ldots, \mathbf{q}_n, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | N=n}[\cdot]$ denotes the expectation over the occurrence random variables, $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ and the random waveform vectors, $\{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_J\}$. To facilitate the expectation, we use the identity

$$E_{\mathbf{q}_1, \ldots, \mathbf{q}_n, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | N=n}[\cdot] = E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n}[E_{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | \mathbf{q}_1, \ldots, \mathbf{q}_n}[\cdot]],$$

to rewrite the last equation as

$$\begin{aligned}
\varphi_{\mathbf{x}}(\mathbf{u}) &= \Pr[N = 0] + \sum_{n=1}^{n=\infty} E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n} \left[E_{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | \mathbf{q}_1, \ldots, \mathbf{q}_n} \right. \\
&\quad \left. \times \left[\exp\left\{ i\mathbf{u} \cdot \sum_{k=1}^{n} \hat{\mathbf{x}}_k(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k) \right\} \right] \right] \Pr[N = n].
\end{aligned} \quad \text{(A2)}$$

In this equation, the expected value operator $E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n}[\cdot]$ is determined from the product of the probability densities of Eq. (2) in the text and the expected value operator $E_{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | \mathbf{q}_1, \ldots, \mathbf{q}_n}[\cdot]$ is determined from the probability law on the individual source waveforms. Using the fact that the breaking-wave occurrences $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ are independent and identically distributed, as are the individual source contributions, the expected value in each term of the sum can be written as

$$\begin{aligned}
E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n} \left[E_{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | \mathbf{q}_1, \ldots, \mathbf{q}_n} \left[\exp\left\{ i\mathbf{u} \cdot \sum_{k=1}^{n} \hat{\mathbf{x}}_k(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k) \right\} \right] \right] &= E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n} \left[E_{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | \mathbf{q}_1, \ldots, \mathbf{q}_n} \left[\prod_{k=1}^{n} \exp\{i\mathbf{u} \cdot \hat{\mathbf{x}}_k(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k)\} \right] \right] \\
&= E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n} \left[\prod_{k=1}^{n} E_{\hat{\mathbf{x}}_k | \mathbf{q}_k} [\exp\{i\mathbf{u} \cdot \hat{\mathbf{x}}_k(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k)\}] \right]
\end{aligned}$$

or

$$E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n} \left[E_{\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_n | \mathbf{q}_1, \ldots, \mathbf{q}_n} \left[\exp\left\{ i\mathbf{u} \cdot \sum_{k=1}^{n} \hat{\mathbf{x}}_k(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k) \right\} \right] \right] = E_{\mathbf{q}_1, \ldots, \mathbf{q}_n | N=n} \left[\prod_{k=1}^{n} \varphi_{\hat{\mathbf{x}}}(\mathbf{u}; \mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k) \right], \quad \text{(A3a)}$$

where

$$\varphi_{\hat{\mathbf{x}}}(\mathbf{u}; \mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k) = E_{\hat{\mathbf{x}} | q}[\exp\{i\mathbf{u} \cdot \hat{\mathbf{x}}(\mathbf{t}, \boldsymbol{\delta}; \mathbf{q}_k)\}] \quad \text{(A3b)}$$

is the characteristic function for the breaking-wave contribution $\hat{\mathbf{x}}$. To evaluate the expected value on the right-hand-side of Eq. (A3a), we use the fact that for a Poisson process, given that there are a fixed number of positions, those positions are independent and identically distributed with a common probability density $\lambda(\mathbf{q})/M$. Thus, Eq. (A3a) can be written as

$$E_{\mathbf{q}_1,\ldots,\mathbf{q}_n|N=n}\left[E_{\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_n|\mathbf{q}_1,\ldots,\mathbf{q}_n}\left[\exp\left\{i\mathbf{u}\cdot\sum_{k=1}^{n}\hat{\mathbf{x}}_k(\mathbf{t},\boldsymbol{\delta};\mathbf{q}_k)\right\}\right]\right]=\prod_{l=1}^{n}E_{\mathbf{q}_l|N=n}\left[\prod_{k=1}^{n}\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q}_k)\right]$$
$$=\int_\Gamma\cdots\int_\Gamma\left[\prod_{k=1}^{n}\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q}_k)\right]\prod_{l=1}^{n}(\lambda(\mathbf{q}_l)/M)d\mathbf{q}_1\ldots d\mathbf{q}_n=\int_\Gamma\cdots\int_\Gamma\left[\prod_{l=1}^{n}\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q}_l)(\lambda(\mathbf{q}_l)/M)\right]d\mathbf{q}_1\ldots d\mathbf{q}_n$$
$$=\prod_{l=1}^{n}\int_\Gamma(\lambda(\mathbf{q}_l)/M)\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q})d\mathbf{q}_i=\left(\int_\Gamma(\lambda(\mathbf{q})/M)\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q})d\mathbf{q}\right)^n. \tag{A4}$$

Substituting Eq. (A4) into Eq. (A2) and using Eq. (1a) for $\Pr[N=0]$ and Eq. (1b) for M yields

$$\varphi_{\mathbf{x}}(\mathbf{u})=\sum_{n=0}^{n=\infty}\left(\int_\Gamma(\lambda(\mathbf{q})/M)\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q})d\mathbf{q}\right)^n \times\exp\{-M\}M^n/n!=\exp\{-M\}\exp\left\{\int_\Gamma\lambda(\mathbf{q})\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{q})d\mathbf{q}\right\}. \tag{A5}$$

Equation (A1a) follows since $M=\int_\Gamma\lambda(\mathbf{q})d\mathbf{q}$.

To complete the derivation of the characteristic function, it remains to show that Eq. (A1a) is equivalent to Eq. (22). To this end, we note that the contribution of the ith breaking wave at time t_j is zero if the breaking wave occurs after time t_j. Since the observations times are ordered, the only breaking waves that can contribute to the random vector $\mathbf{x}$ are those with occurrence times τ that lie in the interval $(-\infty,t_J]$. Furthermore, for $\tau<t_j<t_J$, the only components of the breaking-wave vector that are nonzero, are the components, $\hat{x}_l(t_l,\delta_l;\mathbf{r}_i,\tau_i)=\check{x}(t_l-\tau_i,\delta_l;\mathbf{r}_i)$, for $l\geqslant j$. Therefore, the breaking-wave contribution characteristic function can be written as

$$\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t},\boldsymbol{\delta};\mathbf{r},\tau)=\begin{cases}\varphi_{\check{\mathbf{x}}_{1,J}}(\mathbf{u}_{1,J};\mathbf{t}_{1,J}-\tau;\boldsymbol{\delta}_{1,J};\mathbf{r}) & \text{for } \tau<t_1\\ \varphi_{\check{\mathbf{x}}_{j,J}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau;\boldsymbol{\delta}_{j,J};\mathbf{r}) & \text{for } \tau\in[t_{j-1},t_j];\ j=2,\ldots,J\\ 1 & \text{for } \tau>t_J,\end{cases} \tag{A6a}$$

where

$$\varphi_{\check{\mathbf{x}}_{j,J}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau;\boldsymbol{\delta}_{j,J};\mathbf{r})=E\left[\exp\left\{i\sum_{k=j}^{J}u_k\check{x}_k(t_k-\tau;\delta_k;\mathbf{r})\right\}\right] \tag{A6b}$$

and $\mathbf{t}_{j,J}-\tau=(t_j-\tau,\ldots,t_J-\tau)$, $\mathbf{u}_{j,J}=(u_j,\ldots,u_J)$ and $\boldsymbol{\delta}_{j,J}=(\delta_j,\ldots,\delta_J)$ with $t_0=-\infty$. It follows that

$$\int_{-\infty}^{\infty}\lambda(\mathbf{r},\tau)(\varphi_{\hat{\mathbf{x}}}(\mathbf{u};\mathbf{t};\boldsymbol{\delta};\mathbf{r},\tau)-1)d\tau=\int_{-\infty}^{t_1}\lambda(\mathbf{r},\tau)\times(\varphi_{\check{\mathbf{x}}_{1,J}}(\mathbf{u}_{1,J};\mathbf{t}_{1,J}-\tau;\boldsymbol{\delta}_{1,J};\mathbf{r})-1)d\tau+\int_{t_1}^{t_2}\lambda(\mathbf{r},\tau)\times(\varphi_{\check{\mathbf{x}}_{2,J}}(\mathbf{u}_{2,J};\mathbf{t}_{2,J}-\tau;\boldsymbol{\delta}_{2,J};\mathbf{r})-1)d\tau+\ldots+\int_{t_{J-1}}^{t_J}\lambda(\mathbf{r},\tau)(\varphi_{\check{\mathbf{x}}_{J,J}}(\mathbf{u}_{J,J};\mathbf{t}_{J,J}-\tau;\boldsymbol{\delta}_{J,J};\mathbf{r})-1)d\tau$$
$$=\sum_{j=1}^{J}\int_{t_{j-1}}^{t_j}\lambda(\mathbf{r},\tau)(\varphi_{\check{\mathbf{x}}}(\mathbf{u}_{j,J};\mathbf{t}_{j,J}-\tau;\boldsymbol{\delta}_{j,J};\mathbf{r})-1)d\tau. \tag{A7}$$

Substitution of Eq. (A7) into Eq. (A1a) yields Eq. (22). This completes the derivation. The expression for the correlation function [Eq. (23)] is a consequence of the following lemma.

Lemma 2: For a filtered Poisson process, with zero-mean component processes, the variance and the correlation function are given by

$$\sigma_{\mathbf{x}}^2(t)=\int_\Gamma\lambda(\mathbf{r},\tau)\sigma_{\hat{\mathbf{x}}}(t;\mathbf{r},\tau)dA(\mathbf{r})d\tau \tag{A8a}$$

and

$$R_{\mathbf{x}}(t_k,t_j)=\int_\Gamma\lambda(\mathbf{r},\tau)R_{\hat{\mathbf{x}}}(t_k,t_j;\mathbf{r},\tau)dA(\mathbf{r})d\tau. \tag{A8b}$$

Derivation: It is shown in standard probability texts that the joint first-order moments of a random vector $\mathbf{x}$ are given by

$$E\left[\prod_{j=1}^{J} x_j\right] = i^J \left.\frac{d^J \varphi_{\mathbf{x}}(\mathbf{u})}{du_1 \ldots du_J}\right|_{\mathbf{u}=0}. \tag{A9}$$

To compute the derivatives, we write Eq. (A1a) in the form,

$$\varphi_{\mathbf{x}}(\mathbf{u}) = \exp\{M(\varphi_{\mathbf{x}_a}(\mathbf{u}) - 1)\}, \tag{A10a}$$

where

$$\varphi_{\mathbf{x}_a}(\mathbf{u}) = M^{-1} \int_{\Gamma} \lambda(\mathbf{r}, \tau) \varphi_{\hat{\mathbf{x}}}(\mathbf{u}; \mathbf{r}) dA(\mathbf{r}) d\tau. \tag{A10b}$$

The direct evaluation of the derivative on the right-hand side of Eq. (A9) for $J=2$ yields

$$\frac{\partial^2 \varphi_{\mathbf{x}}(\mathbf{u})}{\partial u_1 \partial u_2} = \left[M \frac{\partial \varphi_{\mathbf{x}_a}(\mathbf{u})}{\partial u_2} \frac{\partial \varphi_{\mathbf{x}_a}(\mathbf{u})}{\partial u_1} + \frac{\partial^2 \varphi_{\mathbf{x}_a}(\mathbf{u})}{\partial u_2 \partial u_1} \right] M \varphi_{\mathbf{x}}(\mathbf{u}). \tag{A11}$$

For zero-mean component processes, the first-order derivatives of the characteristic function are zero so that according to Eq. (A10b), the first-order derivative of $\varphi_{\mathbf{x}_a}(\mathbf{u})$ must be zero. Thus,

$$\begin{aligned} E[x_1, x_2] &= \left.\frac{\partial^2 \varphi_{\mathbf{x}}(\mathbf{u})}{\partial u_1 \partial u_2}\right|_{\mathbf{u}=0} = M \frac{\partial^2 \varphi_{\mathbf{x}_a}(\mathbf{u})}{\partial u_1 \partial u_2} \\ &= \int_{\Gamma} \lambda(\mathbf{r}, \tau) \left.\frac{\partial^2 \varphi_{\hat{\mathbf{x}}}(\mathbf{u}; \mathbf{r})}{\partial u_1 \partial u_2}\right|_{\mathbf{u}=0} dA(\mathbf{r}) d\tau \\ &= \int_{\Gamma} \lambda(\mathbf{r}, \tau) E[\hat{x}_1, \hat{x}_2] dA(\mathbf{r}) d\tau. \end{aligned} \tag{A12}$$

Equation (A8b) follows from Eqs. (A12) and (A9) and the definition of correlation function. Equation (A8a) follows from Eq. (A8b).

APPENDIX B: SPECIAL CASES

In this Appendix we derive the equations used to obtain the numerical results in Sec. V of the paper.

To begin, we specify the first-order characteristic function under the assumption of Gaussian breaking-wave wave forms. The equations are obtained by setting $J=1$ in Eq. (24) in the text and restricting the observations to the same time, $t_j = t_k = t$, and the same point, $\delta_j = \delta_k = \delta$. The result is

$$\varphi_{\mathbf{x}}(u; t, \delta) = \exp\left\{ \int_{\Re} \left[\int_0^{\infty} \lambda(\mathbf{r}, t-\beta) \times (\varphi_{\check{\mathbf{x}}}(u; \beta; \delta; \mathbf{r}) - 1) d\beta \right] dA(\mathbf{r}) \right\}, \tag{B1a}$$

where

$$\varphi_{\check{\mathbf{x}}}(u; \beta; \delta; \mathbf{r}) = \exp\left\{ -\frac{u^2}{2} \sigma_{\check{x}}^2(\beta; \delta; \mathbf{r}) \right\} \tag{B1b}$$

and

$$\sigma_{\check{x}}^2(\beta; \delta; \mathbf{r}) = R_{\check{x}}(\beta, \beta; \delta, \delta; \mathbf{r}) \tag{B1c}$$

is the variance of the breaking-wave contribution. To obtain Eq. (B1a) we have made the change of variable, $\beta = t - \tau$ in the integral. The first-order characteristic function for time-independent breaking-wave densities is obtained by setting $\lambda(\mathbf{r}, \tau) = \lambda(\mathbf{r})$ in the integral in Eq. (B1a).

Broadband bursts with long duration approximation: We next assume that the breaking-wave sound pressure, $s_i(t)$, is a broadband burst given by Eq. (6) in the text with the correlation function given by Eq. (7) and the spectrum given by the approximation of Eq. (9) in the text. Moreover, we assume that the duration of the breaking-wave window function is long compared to the duration of $v(t; \delta; \mathbf{r})$. In the frequency domain, this is equivalent to assuming that $v(f; \delta; \mathbf{r})$ is approximately constant over the frequencies where $\Pi_{\eta}(f - f')$ and hence, $\Delta(f - f')$ is significant. Support for this assumption is seen in the Greens function plots in Sec. V. We begin by developing the expression for the correlation function of the breaking-wave contribution. To this end, we first substitute Eq. (9a) in the text into Eq. (18) and interchange the order of integration. The result is

$$\begin{aligned} R_{\check{x}}(\beta, \beta + \xi_{j,k}; \delta_j, \delta_k; \mathbf{r}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_0(f') \Delta(f - f') V(f; \delta_j; \mathbf{r}) V(f'; \delta_k; \mathbf{r})^* e^{i2\pi[(f-f')\beta - f'\xi_{j,k}]} df df' \\ &\approx \int_{-\infty}^{\infty} S_0(f) V(f; \delta_j; \mathbf{r}) V(f; \delta_k; \mathbf{r})^* \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta(f - f') e^{i2\pi[(f-f')\beta - f'\xi_{j,k}]} df' \right] df, \end{aligned}$$

where we have used the fact that $V(f'; \delta_k; \mathbf{r})$ is approximately constant compared to $\Delta(f - f')$ for each f'. Using Eq. (9b) in the text, the inner integral can be written as

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta(f - f') e^{i2\pi[(f-f')\beta - f'\xi_{j,k}]} df' = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \pi_{\eta}(t)^2 e^{-i2\pi(f-f')t} dt \right] e^{i2\pi[(f-f')\beta - f'\xi]} df' \\ &= \int_{-\infty}^{\infty} e^{-i2\pi(t-\beta)f} \left[\int_{-\infty}^{\infty} e^{i2\pi(t-\beta)f'} e^{-i2\pi f'\xi} df' \right] \pi_{\eta}(t)^2 dt = \int_{-\infty}^{\infty} e^{-i2\pi(t-\beta)f} \left[\int_{-\infty}^{\infty} e^{i2\pi(t-\beta-\xi)f'} df' \right] \pi_{\eta}(t)^2 dt = \int_{-\infty}^{\infty} e^{-i2\pi(t-\beta)f} \delta(t - \beta \\ &- \xi) \pi_{\eta}(t)^2 dt = e^{-i2\pi\xi f} \pi_{\eta}(\beta + \xi)^2 \end{aligned}$$

so that

$$R_{\breve{x}}(\beta,\beta+\xi_{j,k};\delta_j,\delta_k;\mathbf{r}) \approx \pi_\eta(\beta+\xi_{j,k})^2 R_{vs}(\xi_{j,k};\delta_j,\delta_k;\mathbf{r}), \tag{B2a}$$

where

$$R_{vs}(\xi_{j,k};\delta_j,\delta_k;\mathbf{r}) = \int_{-\infty}^{\infty} V(f;\delta_j,\mathbf{r})V(f;\delta_k;\mathbf{r})^* \times e^{-i2\pi f\xi_{j,k}} S_0(f)df. \tag{B2b}$$

The breaking-wave contribution variance is obtained by setting $\xi=0$ and $\delta_j=\delta_k=\delta$. The result is

$$\sigma_{\breve{x}}^2(\beta;\delta;\mathbf{r}) = \pi_\eta(\beta)^2 R_{vs}(0;\delta,\delta;\mathbf{r}) = \pi_\eta(\beta)^2 \int_{-\infty}^{\infty} |V(f;\delta;\mathbf{r})|^2 S_0(f)df. \tag{B3}$$

The noise cross-correlation is obtained by substituting Eq. (B2) into Eq. (18) in the text,

$$R_x(t_j,t_k;\delta_j,\delta_k) = \int_{\Re}\int_0^{\infty} \lambda(\mathbf{r},t_j-\beta)\pi_\eta(\beta+\xi_{j,k})^2 R_{vs}(\xi_{j,k};\delta_j,\delta_k;\mathbf{r})d\beta dA(\mathbf{r}) = \int_{\Re}\left[\int_0^{\infty} \lambda(\mathbf{r},t_j-\beta)\pi_\eta(\beta + \xi_{j,k})^2 d\beta\right] R_{vs}(\xi_{j,k};\delta_j,\delta_k;\mathbf{r})dA(\mathbf{r}).$$

If the breaking-wave density is independent of time, then

$$R_x(t_j,t_k;\delta_j,\delta_k) = R_\pi(\xi_{j,k})\int_{\Re} \lambda(\mathbf{r}) R_{vs}(\xi_{j,k};\delta_j;\delta_k;\mathbf{r})dA(\mathbf{r}), \tag{B4a}$$

where

$$R_\pi(\xi) = \int_0^{\infty} \pi_\eta(\beta+\xi)^2 d\beta = \int_\xi^{\infty} \pi_\eta(\beta')^2 d\beta' = \begin{cases} 0 & |\xi| > \eta \\ \displaystyle\int_\xi^{\eta} \pi_\eta(\beta')^2 d\beta' & |\xi| < \eta \end{cases}. \tag{B4b}$$

Equations (B4) and (B2b) are equivalent to Eq. (25) in the text. The noise variance is

$$\sigma_x^2(\delta) = \int_{\Re} \lambda(\mathbf{r}) R_{vs}(0;\delta,\delta;\mathbf{r})dA(\mathbf{r}), \tag{B5}$$

where we have used the fact that $\pi_\eta(t)$ is normalized so that $R_\pi(0)=1$ (see Sec. II of the text).

To obtain the characteristic function, we note from Eq. (B3) and the fact that $\pi_\eta(\beta)=0$ for $\beta \notin [0,\eta]$ that

$$\varphi_{\hat{x}}(u;\beta;\delta,\mathbf{r}) - 1 = \begin{cases} \exp\left\{-\dfrac{u^2}{2}\pi_\eta(\beta)^2 R_{vs}(0;\delta;\mathbf{r})\right\} - 1 & \beta \in [0,\eta] \\ 0 & \text{otherwise.} \end{cases}$$

Substituting into Eq. (B1a), yields

$$\varphi_x(\mathrm{u};\delta) = \exp\left\{\int_A \lambda(\mathbf{r})\left[\int_0^{\eta}\left(\exp\left\{-\frac{u^2}{2}\pi_\eta(\beta)^2 R_{vs}(0;\delta;\mathbf{r})\right\} - 1\right)d\beta\right]dA(\mathbf{r})\right\}. \tag{B6}$$

The characteristic function of the noise normalized to unit variance is obtained by using Eq. (B5) for the noise variance to write

$$\varphi_x\left(\frac{u}{\sigma_x};\delta\right) = \exp\left\{\int_A \lambda(\mathbf{r})\left[\int_0^{\eta}\left(\exp\left\{-\frac{u^2}{2}\frac{\pi_\eta(\beta)^2 R_{vs}(0;\delta;\mathbf{r})}{\displaystyle\int_{\Re}\lambda(\mathbf{r})R_{vs}(0;\delta;\mathbf{r})dA(\mathbf{r})}\right\} - 1\right)d\beta\right]dA(\mathbf{r})\right\}. \tag{B7}$$

Special case: Constant window function:

$$\pi(t) = \begin{cases} \eta^{-1/2} & t \in [0,\eta] \\ 0 & \text{otherwise,} \end{cases} \tag{B8}$$

$$R_\pi(\xi) = \int_\xi^{\eta} \pi_\eta(\beta')^2 d\beta' = \eta^{-1}(\eta-\xi) = \left(1-\frac{\xi}{\eta}\right). \tag{B9}$$

It follows that

$$R_x(\xi;\delta_j,\delta_k)=\left(1-\frac{|\xi|}{\eta}\right)\int_{\Re}\lambda(\mathbf{r})R_{vs}(\xi;\delta_j,\delta_k;\mathbf{r})dA(\mathbf{r}) \tag{B10}$$

and

$$\sigma_x^2(\delta)=\int_{\Re}\lambda(\mathbf{r})R_{vs}(0;\delta;\mathbf{r})dA(\mathbf{r}) \tag{B11}$$

and

$$\varphi_x\left(\frac{u}{\sigma_{\mathbf{x}}};\delta\right)=\exp\left\{\int_A(\eta\lambda(\mathbf{r}))\left(\exp\left\{-\frac{u^2}{2}\frac{R_{vs}(0,\delta;\mathbf{r})}{\oint_{\Re}(\eta\lambda(\mathbf{r}))R_{vs}(0,\delta;\mathbf{r})dA(\mathbf{r})}\right\}-1\right)dA(\mathbf{r})\right\}. \tag{B12}$$

Special case: Constant breaking-wave density $\lambda(\mathbf{r})=\lambda$. It follows that

$$R_x(\xi;\delta_j,\delta_k)=\lambda\left(1-\frac{|\xi|}{\eta}\right)\int_{\Re}R_{vs}(\xi;\delta_j,\delta_k;\mathbf{r})dA(\mathbf{r}) \tag{B13}$$

and

$$\sigma_x^2(\delta)=\lambda\int_{\Re}R_{vs}(0,\delta;\mathbf{r})dA(\mathbf{r}) \tag{B14}$$

and

$$\varphi_x\left(\frac{u}{\sigma_{\mathbf{x}}};\delta\right)=\exp\left\{(\eta\lambda)\oint_{\Re}\left(\exp\left\{-\frac{u^2}{2}\frac{R_{vs}(0;\delta;\mathbf{r})}{(\eta\lambda)\oint_{\Re}R_{vs}(0;\delta;\mathbf{r})dA(\mathbf{r})}\right\}-1\right)dA(\mathbf{r})\right\}. \tag{B15}$$

Special case: Band-limited white spectrum $S_0(f)=S_0$; $f\in[f_1,f_2]$. It follows that

$$\begin{aligned}R_{vs}(\xi;\delta_j,\delta_k;\mathbf{r})&=S_0\int_{f_1}^{f_2}V(f;\delta_j;\mathbf{r})V(f;\delta_k;\mathbf{r})^*e^{i2\pi f\xi}df\\&=S_0R_v(\xi;\delta_j,\delta_k;\mathbf{r}),\end{aligned} \tag{B16a}$$

where

$$R_v(\xi;\delta_j,\delta_k;\mathbf{r})=\int_{f_1}^{f_2}V(f;\delta_j;\mathbf{r})V(f;\delta_k;\mathbf{r})^*e^{i2\pi f\xi}df. \tag{B16b}$$

Thus

$$R_x(\xi,\delta_j,\delta_k)=\lambda S_0\left(1-\frac{|\xi|}{\eta}\right)\int_{\Re}R_v(\xi;\delta_j,\delta_k;\mathbf{r})dA(\mathbf{r}) \tag{B17}$$

and

$$\sigma_x^2(\delta)=\lambda S_0\int_{\Re}R_v(0;\delta;\mathbf{r})dA(\mathbf{r}) \tag{B18}$$

and

$$\varphi_x\left(\frac{u}{\sigma_x};\delta\right)=\exp\left\{\int_{\Re}(\eta\lambda(\mathbf{r}))\left(\exp\left\{-\frac{u^2}{2}\frac{R_v(0;\delta;\mathbf{r})}{\oint_{\Re}(\eta\lambda(\mathbf{r}))R_v(0;\delta;\mathbf{r})dA(\mathbf{r})}\right\}-1\right)dA(\mathbf{r})\right\}. \tag{B19}$$

[1]A. Thode, M. Zanolin, E. Naftali, I. Ingram, P. Ratilal, and N. Makris, "Necessary conditions for a maximum likelihood estimate to become asymptotically unbiased and attain the Cramer-Rao lower bound. II. Range and depth localization of a sound source on an ocean waveguide," J. Acoust. Soc. Am. **121**, 1890–1910 (2002).
[2]H. L. Van Trees, *Detection, Estimation, and Linear Modulation Theory—Part III* (Wiley, New York, 1971).
[3]N. Makris, "The statistics of ocean-acoustic ambient noise," in *Sea Surface Sound*, edited by T. Leighton (Kluwer Academic, Dordrecht, 1997).
[4]W. A. Kuperman and F. Ingenito, "Spatial-correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[5]R. D. Hollett, "Observations of underwater sound at frequencies below 1500 Hz from breaking waves at sea," J. Acoust. Soc. Am. **95**, 165–170 (1994).
[6]E. Parsen, *Stochastic Processes*, (Holden-Day, 1962).
[7]D. Middleton, "A statistical theory of reverberation and similar first order scattered fields. I. Waveform and general processes," IEEE Trans. Inf. Theory **IT-13**(3), (1967).
[8]D. Middleton, "On the theory of random noise, phenomenological models I, II," J. Appl. Phys. **22**, 1143 (1951); **22**, 1153 (1951); **22**, 1326 (1951).
[9]A. Milne, "Statistical description of noise under shore fast sea ice," J. Acoust. Soc. Am. **39**, 1174–1182 (1966).
[10]A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd ed. (McGraw-Hill, New York, 1991).
[11]L. Ding and D. M. Farmer, "On the dipole acoustic source level of breaking waves," J. Acoust. Soc. Am. **96**, 3036–3044 (1994).
[12]R. M. Kennedy, "Sea surface dipole sound source dependence on wave-breaking variables," J. Acoust. Soc. Am. **91**, 1974–1982 (1992).
[13]R. M. Kennedy and T. V. Goodnow, "Measuring the vertical directional spectra caused by sea surface sound," IEEE J. Ocean. Eng. **15**, 299–310 (1990).
[14]D. J. Kewley, D. G. Browning, and W. M. Carey, "Low-frequency wind-generated ambient noise source levels," J. Acoust. Soc. Am. **88**, 1894–1902 (1990).
[15]L. Ding and D. M. Farmer, "Observations of breaking surface wave statistics," J. Phys. Oceanogr. **24**, 1368–1387 (1994).
[16]S. Finette and R. M. Heitmeyer, "Angle-time-frequency resolution of the noise field generated by wind-induced breaking waves," J. Acoust. Soc. Am. **99**, 209–222 (1996).
[17]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP Press, New York, 1994).

Low-frequency acoustic propagation loss in the Arctic Ocean: Results of the Arctic climate observations using underwater sound experiment

Alexander N. Gavrilov
Centre for Marine Science and Technology, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia

Peter N. Mikhalevsky[a)]
Science Applications International Corporation, 4001 N. Fairfax Dr., Suite 850, Arlington, Virginia 22203

(Received 4 March 2004; revised 16 March 2006; accepted 17 March 2006)

Acoustic data from the Arctic climate observations using underwater sound (ACOUS) experiment are analyzed to determine the correlation between acoustic propagation loss and the seasonal variability of sea ice thickness. The objective of this research is to provide long-term synoptic monitoring of sea ice thickness, an important global climate variable, using acoustic remote sensing. As part of the ACOUS program an autonomous acoustic source deployed northwest of Franz Josef Land transmitted tomographic signals at 20.5 Hz once every four days from October 1998 until December 1999. These signals were received on a vertical array in the Lincoln Sea 1250 km away. Two of the signals transmitted in April 1999 were received on a vertical array at ice camp APLIS in the Chukchi Sea north of Point Barrow, Alaska, at a distance of approximately 2720 km from the source. Temporal variations of the modal propagation loss are examined. The influence of ice parameters, variations of the sound speed profile, and mode-coupling effects on the propagation losses of individual modes is studied. The experimental results are compared to the results of the earlier experiments and the theoretical prediction using numerical modeling. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195255]

PACS number(s): 43.30.Pc, 43.30.Hw, 43.30.Qd [WLS] Pages: 3694–3706

I. INTRODUCTION

Acoustic ice reflection and scattering have been subjects of many experimental and theoretical studies since the earlier 1960s, because the interaction of sound waves with the rough sea ice cover is the major factor that governs attenuation of acoustic signals propagated in the Arctic Ocean. The signal transmission loss increases rapidly with frequency due to ice scattering. Therefore long-range acoustic transmissions in the Arctic Ocean are possible only at low frequencies of tens of Hz, and consequently low-frequency acoustic propagation is of particular research interest. The total energy loss of transmitted signals has been studied in several acoustic propagation experiments in the Arctic.[1–3] Attenuation of individual normal modes at low frequencies is much less investigated experimentally. Knowing the modal propagation loss is important, because this makes it possible to predict the range and depth dependence of the sound field. Moreover, theoretical models of acoustic ice reflection and scattering at low frequencies can be examined in more detail by comparing the results of numerical prediction for modal attenuation due to ice scattering with the modal propagation loss measured experimentally.

Propagation losses of individual modes in the signals transmitted over the entire Arctic Basin were measured for the first time in the Transarctic Acoustic Propagation (TAP) experiment[4–6] in 1994. The TAP experiment results provided us with important data that were used for verifying the most recent models of ice scattering, propagation loss, and reverberation.[7] However, the TAP measurements were conducted for a relatively short time period of a week in April, which did not allow us to study temporal variations of the modal propagation loss and modal attenuation with changes in the ice cover characteristics and, in particular, seasonal variations of the mean ice thickness and roughness along the path. If this correlation is experimentally examined and determined, it would prove the capability of acoustics to remotely monitor the sea ice thickness year round over large regions of the Arctic Ocean.

Long-term acoustic transmissions on a stationary trans-Arctic path were conducted for the first time in the Arctic climate observations using underwater sound (ACOUS) experiment in 1998–1999. Acoustic observations on the path of 1250 km long from the Franz Victoria Strait to the Lincoln Sea lasted for 14 months. Two of the ACOUS signals were also received on a vertical array at ice camp APLIS drifting in the northern part of the Chukchi Sea 2750 km away from the source. Signal reception on vertical arrays allowed us to separate individual modes of the signals and analyze different parameters of the modes, including the modal propagation losses. The main goal of the ACOUS experiment was to study long-term changes in the ocean temperature and heat content along the trans-Arctic path by means of acoustic thermometry. In addition, the data obtained gave us an opportunity to investigate temporal variations of modal attenu-

a)Electronic mail: mikhalevskyp@saic.com

 0001-4966/2006/119(6)/3694/13/$22.50

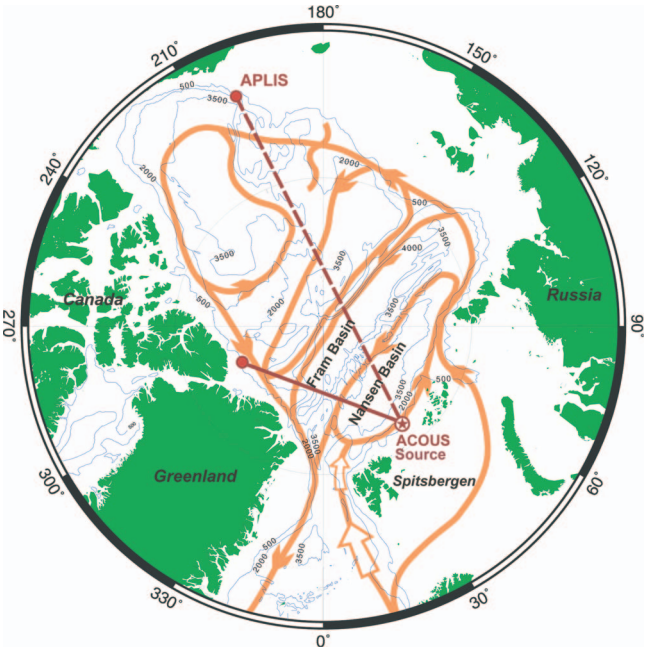


FIG. 1. ACOUS experiment paths in the Arctic Ocean. The orange-colored lines and arrows show the recently proposed scheme of Atlantic water circulation (Ref. 8).

ation due to various factors including seasonal change in the sea ice characteristics, variations of the sound speed, and the mode coupling effects.

It is commonly believed that the acoustic propagation loss due to ice scattering in the deep Arctic Basin depends primarily on morphological and physical characteristics of the rough sea ice cover. A model sound speed profile typical for the Arctic Basin is often used to numerically model acoustic transmission losses due to ice scattering. Although the Arctic sound speed profiles are generally upward refracting and look nearly alike all over the deep Arctic Basin, actually certain important characteristics of profiles, such as the thickness and sharpness of the upper under-ice acoustic duct above the thermocline, vary considerably from one region to another. Relatively small spatial and temporal variations of the sound speed profile may lead to considerable changes in the modal group velocities and attenuation. The other factor that may strongly affect the modal propagation loss is mode coupling over range-dependent sections of acoustic paths. The influence of these factors on the modal propagation loss on the ACOUS path is considered in this paper along with the analysis of the temporal variations of modal attenuation due to seasonal change in the ice characteristics. The results of numerical modeling are used to interpret the experimental observations.

II. ACOUS EXPERIMENT

A. Experimental scheme

The ACOUS paths to the Lincoln Sea and the Chukchi Sea are shown in Fig. 1. The ACOUS source was moored at the edge of the Eurasian continental shelf at a depth of 60 m below the sea surface in water depth of 440 m. Every four days at 00:00 GMT, the source transmitted an M-sequence signal of 255 digits at the 20.5-Hz carrier phase-modulated with a deflection of ±60°. Ten periods of the carrier formed each digit of the sequence. The whole transmitted signal consisted of 10 repetitions of the 255-digit M-sequence and lasted for about 20.7 minutes. The signal level was nominally 195 dB re. 1 μPa at 1 m. Two accelerometers installed on the membranes of the source measured their oscillations and controlled the signal level through an onboard signal processor. If the amplitude of oscillations corresponded to the nominal source level within ±0.5 dB, the octal code of the M-sequence was 703. Otherwise the signal would be modulated with a different M-sequence, with the octal code set according to a look-up table corresponding to the measured source level.

The autonomous vertical receive array in the Lincoln Sea consisted of eight hydrophones spaced equidistantly at a 70-m interval. The array was moored near the edge of the Canadian continental shelf in water depth of 545 m, so that it spanned almost all the water column. Five self-recording micro-CTD's distributed from the top to the bottom of the array logged hydrostatic pressure, water temperature and salinity every 10 minutes. Examination of acoustic recordings after recovery of the array in March 2001 revealed that the source transmitted a total of 107 signals and stopped operating in December 1999. The M-sequence code in all transmissions was 703, which meant that the level of the transmitted signal remained at the nominal value of 195 dB.

On April 9 and 13, 1999, two ACOUS signals were received on a vertical array at ice camp APLIS established at the northern edge of the Chukchi Sea to support the SCICEX-1999 expedition. This flexible array was suspended from sea ice and consisted of eight equidistant hydrophones from 150 m to 675 m below the sea surface. The array tilt was tracked via an ice-mounted acoustic navigation system.

B. Environmental conditions

The vertical section of the sound speed field along the path to the Lincoln Sea is shown as a contour plot along with bathymetry in Fig. 2. The path crosses both Eurasian and Canadian continental slopes, which are steep enough to induce strong coupling of propagated acoustic modes. The Gakkel Ridge in the middle of the path and the western edge of the Lomonosov Ridge north of the Lincoln Sea are not so high as to substantially affect propagation of the low-order modes at 20 Hz in the upward refracting Arctic acoustic channel.

The ACOUS path to the Lincoln Sea crosses the major gyre and two minor branches of Atlantic Intermediate Water (AIW) circulation in the Arctic Ocean[8] shown in Fig. 1 with the orange lines and arrows. The AIW flow coming into the Nansen Basin with the boundary current along the Eurasian continental margins brings warm waters of which the temperature exceeds 2 °C and constitutes a contrasting warmer core in the intermediate layer. This core can be distinguished in the sound speed field in Fig. 2 within the initial 300-km section of the ACOUS path at depths from 100 m to 500 m. The outgoing flow along the Canadian continental slope is much less warm and cannot be spatially discriminated in the

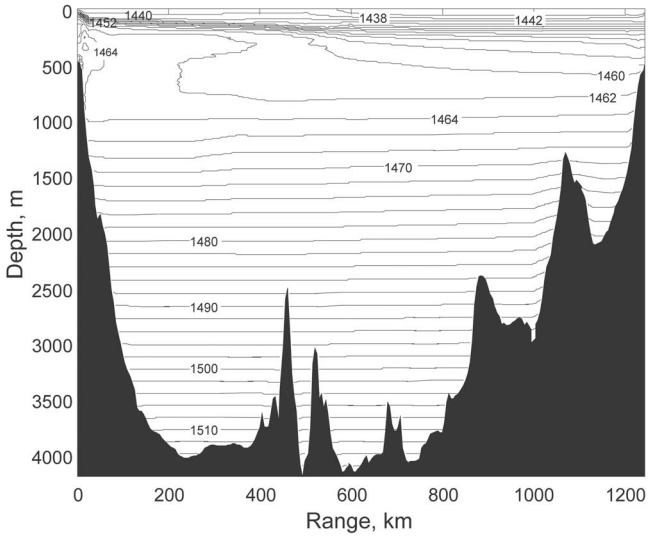


FIG. 2. Bathymetry and the typical sound speed contours along the acoustic path to the Lincoln Sea.

sound speed field, as well as the minor branches of AIW circulation along the Lomonosov and Gakkel Ridges. A statistical analysis of the historical oceanographic data collected in the regions along the acoustic path has shown that the AIW temperature is most variable along the initial section of the acoustic path in the Nansen Basin and, in particular, over the Eurasia continental slope. In the Fram Basin and the Lincoln Sea the temperature profiles are more stable and vary a small amount relative to the climatology profiles.[9]

The data on ice concentration[10] show that during the period of experimental observation the sea surface was covered year-round with ice everywhere along the acoustic path. The mean ice draft was expected to be approximately 2 m in the Nansen Basins in the so-called Arctic summer season (July–November).[11,12] The other regions along the path and the Arctic winter season (February–May) are not well represented with the recent submarine upward looking sonar data.

The acoustic path to APLIS crosses the Eurasian continental slope, the Lomonosov and Mendeleyev Ridges, and the Chukchi Plateau (Fig. 3). The Lomonosov and Mendeleyev Ridges do not rise above a depth of 1000 m and therefore do not have significant influence on the low-order modes at 20 Hz. The sea floor topography in the region of the Chukchi Plateau is very uneven. The acoustic path crossed some parts of the plateau over which the depth was about 600 m and hence the sea floor influenced every mode of the ACOUS signal except for mode 1. The Scientific Ice Expeditions (SCICEX) oceanographic transect conducted by USS Hawkbill in April 1999 was almost coincident with the ACOUS path to APLIS and made within a month after the ACOUS signal receptions. Therefore the sound speed profiles were well defined everywhere along the acoustic path except for the initial 400-km section that was not spanned by the SCICEX transect. The oceanographic data collected in the regions along this section in the 1990s and the SCICEX-1999 data were used to synthesize the sound speed field along the acoustic path shown in Fig. 3. In April 1999, all regions along the ACOUS path to APLIS were almost entirely covered with sea ice. In the absence of the SCICEX-1999 upward looking sonar data, we used the ice draft data collected in the earlier submarine cruises[12] across the Arctic Ocean to model statistical characteristics of the rough ice cover along the acoustic path.

III. EXPERIMENTAL RESULTS

A. Processing of the ACOUS signals

After signal demodulation and pulse compression, the modal propagation loss, relative to the reference level of modal amplitudes at a distance of 1 m from the source, can be estimated as[13]

$$PL_m = 20\log(2^{-3/2} B\, A_{m0} \sin^2\theta) - 20\log(\max\{|p_m|\}) - 10\log(r), \quad (1)$$

where $p_m(t)$ is the pulse-compressed wave form of individual modes filtered on the array, $\max\{|p_m|\}$ is the peak magnitude of the modal pulse, $B=\sqrt{2}\exp(195/20)$ is the peak

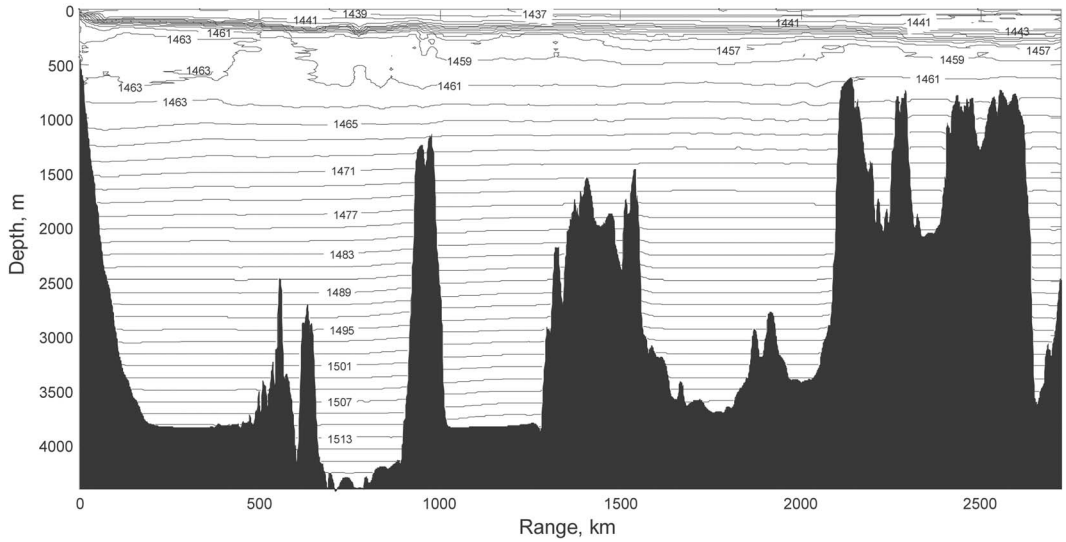


FIG. 3. Bathymetry and the sound speed contours along the acoustic path to APLIS.

signal level of the ACOUS source at 1 m, $\theta=\pi/3$ is the phase deflection, and

$$A_{m0}=\left(\frac{2\pi}{k_m(0,\omega)}\right)^{1/2}Z_m(0,z_0,\omega)$$

are the modal excitation coefficients that are determined from the modal eigenfunctions $Z_m(0,z_0,\omega)$ and wavenumbers $k_m(0,\omega)$ at the source location.

Expression (1) is exact for an ideal triangular pulse resulting from pulse compression in an infinite frequency band. Because of the finite frequency band of the source-to-receiver transfer function, the actual shape of pulse-compressed signals is not exactly triangular and the peak magnitude of the arrival pulse is considerably smaller. Moreover, the frequency dependence of modal wavenumbers is not absolutely linear, which may cause certain time dispersion of the modal pulse and a reduction of its maximum. The energy of the modal arrival pulse is invariant under changes of the pulse shape and hence can be used as a more accurate measure of the modal propagation loss. To allow for pulse distortion in Eq. (1), we use the effective modal amplitude $\hat{p}_m$ instead of $\max\{|p_m|\}$. This effective amplitude is defined from the pulse energy

$$E_m=\int p_m(t)p_m^*(t)dt$$

as

$$\hat{p}_m=\left(\frac{3E_m}{2T}\right)^{1/2}, \tag{2}$$

where the modal pulse p_m is integrated over its whole duration and T is the length of one digit of the 255 digit M-sequence.[6] Strong mode coupling also leads to a distortion of the modal pulse. In that case, the estimate of the effective modal amplitude $\hat{p}_m$ of mode m also includes the energy contribution from the other modes.

The parameters of the Lincoln Sea array were nearly optimum for accurate spatial filtering of the eight low-order modes that contributed most to the sound field on the array. The method of mode filtering on the Lincoln Sea array is discussed in Ref. 14, where the filtering algorithm is examined with the use of the signals numerically modeled on the hydrophones of the array. In contrast to the Lincoln Sea array, the vertical eight-element array deployed at APLIS spanned only a relatively small section of the deep-water Arctic acoustic channel. The vertical locations of the array hydrophones were such that the matrix of modal eigenfunctions was nearly singular with respect to four of the eight eigenvalues.[15] Therefore the eigenvector method[16] was used for filtering the modes on the APLIS array. Numerical modeling showed that the results of mode filtering should be sufficiently accurate if the contribution of modes higher than four to the sound field is relatively small. Processing of the ACOUS signals has confirmed this assumption, which will be shown in Sec. III C.

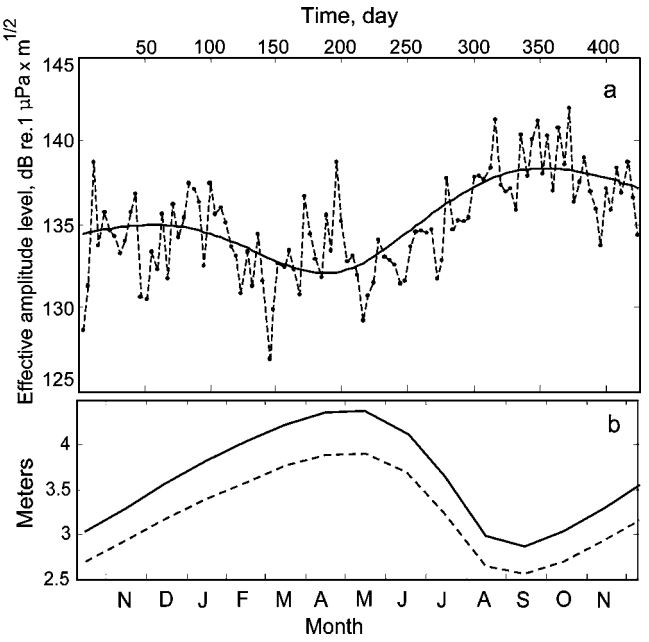


FIG. 4. (a) Variations of the group effective amplitude of modes 1–8 filtered on the Lincoln Sea array. The solid line shows the result of wavelet filtering of the long-term component of variations. (b) Modeled seasonal cycle of ice draft (dashed) and thickness (solid) (Ref. 11).

B. Signals on the Lincoln Sea array

The effective modal amplitudes $\hat{p}_m$, travel times, and phase were measured after filtering individual modes on the Lincoln Sea array. The temporal coherence of the received signals was high so that the total gain of signal processing approached its maximum of 34 dB. The resulting signal-to-noise ratio in the post-processed signals on individual hydrophones exceeded 40 dB for almost all of the ACOUS signal recordings. The tilt of the array under variable currents was examined using the data from the pressure sensor of the micro-CTD deployed at the top of the array and a hydrodynamic model of the array mooring system. From the results of numerical modeling, it was assumed that the array tilt would be negligible for mode filtering if the vertical displacement of the top sensor relative to its uppermost position did not exceed a threshold of 0.5 m, which corresponded to horizontal displacement of approximately 20 m. Sixty-three of 107 ACOUS receptions met this condition.

We first examined the variation of the total energy of the eight modes filtered on the array, which was almost equivalent to the depth-integrated signal energy. All of the 107 signals were used for this analysis, because cross-talk of the modes on the tilted array did not considerably affect the total mode energy. Figure 4(a) shows the variation of the level of the group effective amplitude $\hat{p}=[\Sigma_{m=1}^{8}\hat{p}_m^2]^{1/2}$ of modes 1–8 for 14 months of observations. In the temporal variation of the modal amplitude, one can notice the long-term change in the presence of superimposed short-term fluctuations. To separate the long-term change from short-term fluctuations, we applied the wavelet de-noising procedure, decomposing the signal by the "coiflets" family wavelets.[17] The solid line in Fig. 4(a) shows the results of such low-pass filtering. The residual short-term variations behave as a random, delta-

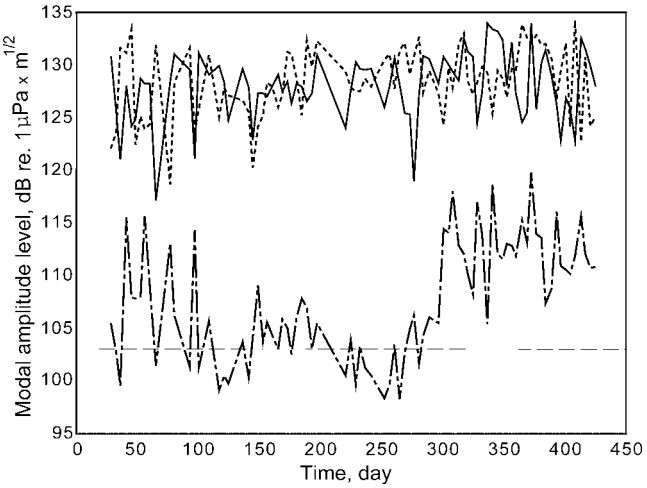


FIG. 5. Variations of the effective amplitude of modes 1 (dashed-dotted), 2 (solid), and 3 (dotted) filtered on the Lincoln Sea array during the ACOUS experiment. The dashed line shows the average noise level in the mode 1 filter channel.

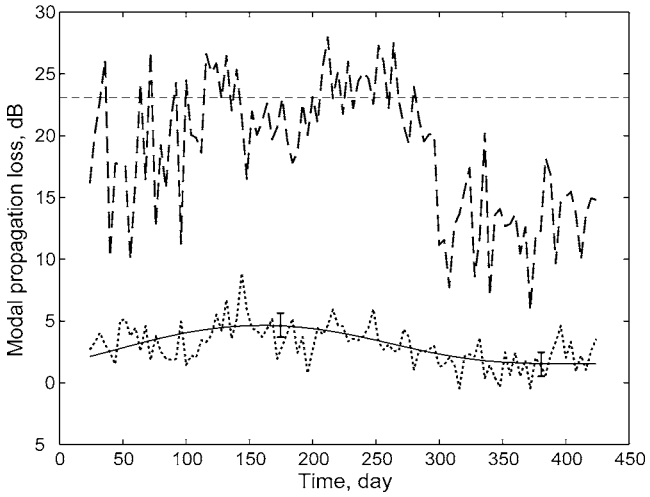


FIG. 6. Propagation loss [see Eq. (1)] of mode 1 (dashed) and the group of modes 2–4 (dotted) received on the Lincoln Sea array during the ACOUS experiment. The estimates of mode 1 propagation loss above the thin dashed line are not valid, because the mode 1 signal level was below the average noise level. The seasonal component of modes 2–4 propagation loss variation is shown with the solid line. Error bars demonstrate a 95% confidence interval for the seasonal component derived in the presence of random short-term fluctuations.

correlated process with a zero mean and a standard deviation of approximately 1 dB. The filtered long-term change correlates very well with the seasonal changes in the mean sea ice thickness [Fig. 4(b)] derived from the historical data on ice draft and the Arctic ice-ocean model.[11] The most likely cause of this correlation is the influence of the seasonal change in the ice cover on the acoustic propagation loss, which will be discussed in detail in Sec. IV.

Temporal variations of the effective amplitude $\hat{p}_m$ of modes 1–3 are shown in Fig. 5. Here the samplings are not uniformly spaced in time, because they do not include the events of excessively large tilt of the array at the time of signal reception. Consequently, the procedure of wavelet denoising cannot be correctly applied to these data sets. Mode 1 of the ACOUS signals arrived at the Lincoln Sea array 6.5–8 seconds later than mode 3.[14] The main arrival pulse of mode 1 exhibited considerable fluctuations and spreading, and could not be detected in some of the signal receptions. In such cases, the estimates of mode 1 effective amplitude, made by integration of the mode-filtered signal within a time window from 5.5 s to 9 s after mode 3 arrival, represent variations of the noise level in the mode 1 filter channel. The dashed line in Fig. 5 shows the average noise level in this channel. Despite very large short-term fluctuations, it can be seen in Fig. 5 that the propagation loss of mode 1 underwent substantial seasonal variations: the pulse of mode 1 can be distinguished with high confidence in approximately 50% of signal recordings made in October–December 1998 (days 1–90 in Fig. 5); in February–June 1999 (days 110–260 in Fig. 5), mode 1 almost disappeared and could be detected in few ACOUS signals; in August 1999 (days 290–420 in Fig. 5), the amplitude of this mode increased rapidly so that the main arrival pulse could be clearly seen in almost all of the signal recordings until the end of observations in December. It is impossible to discriminate any long-term constituent in the temporal variations of the amplitude of modes 2 and 3 in Fig. 5. According to Fig. 4(a) and the results of numerical modeling discussed in the next section, the seasonal peak-to-peak variations of individual modes 2, 3 and higher are expected to be less than 5 dB, and hence they might be hidden in the presence of the short-term fluctuations that have a standard deviation of about 3 dB.

Modes 2, 3, and 4 attenuate along the major, deep-water section of the ACOUS path with approximately the same rate. Therefore we calculated the group effective amplitude and the mode-average propagation losses for these modes, which allowed us to considerably reduce the short-term fluctuations due to possible effects of variable mode coupling. Figure 6 shows the variations of the propagation loss of mode 1 and the group of modes 2–4 on this path for 14 months of observations, disregarding the loss due to cylindrical spreading. The seasonal cycle in the variation of modes 2–4 propagation losses is not prominent, but visible. In the Arctic peak winter season (approximately days 120–260), the propagation loss of this modal group was typically 3–5 dB. In the peak summer season of 1999 (days 290–420), the propagation loss of the modal group decreased to 1.5–2 dB on average. The propagation loss of mode 1 was varied from 18 dB to 22 dB and higher values above the noise-limited threshold in winter and within 8–12 dB in the summer period of 1999.

The modal propagation loss observed in winter corresponds to the path-average attenuation coefficients of 0.015–0.018 dB/km for mode 1 and 0.003–0.004 dB/km for modes 2–4. These numbers are similar to the modal attenuation coefficients measured in the TAP experiment[6] along the Turpan-to-SIMI path, but much smaller than those measured on the Turpan-to-Narwhal path which was almost coincident with the 1000-km section of the ACOUS path from the Nansen Basin to the Lincoln Sea. The integral propagation loss along the path to Narwhal was about 30 dB for mode 1 and 10–13 dB for modes 2–4. The higher signal attenuation on this path was attributed to the fact that the ice

cover in the Lincoln Sea and surrounding deep-water regions was generally thicker and rougher than that in the Central Arctic Basin.

It is necessary to note that the propagation loss of modes 2–4 measured on this path are small and may have considerable relative errors due to imperfect calibration of the array hydrophones and some uncertainty in the actual level of the transmitted signal. Indeed, the calibration accuracy of the hydrophones of the Lincoln Sea array is about ±1.5 dB. Although the amplitude of oscillations of the source membrane was automatically adjusted to the reference sound pressure with a resolution of 0.5 dB, the actual uncertainty of the absolute source level is expected to be at least ±1 dB. Therefore the total instrumental error of the absolute propagation loss measurements may amount to 2–5 dB. The dominant component of instrumental errors is expected to be regular.

C. Signals on the APLIS array

The arrival patterns of modes 1–4 of the ACOUS signal filtered on the vertical array at ice camp APLIS on April 9 and 13, 1999 are shown in Figs. 7(a) and 7(b), respectively. The amplitude of mode 1 shown in this figure is multiplied by a factor of 2. Mode 2 dominates the other mode in the ACOUS signals received at APLIS. The arrival pulse of this mode in the first signal is split, which testifies to strong mode coupling. Mode 1 is weak, but distinguishable in both signals. Mode 4 is practically vanishing in these signals.

The propagation loss of mode 1 measured on the APLIS array is 41 dB for the first signal and 45 dB for the second signal, which corresponds to the path-average attenuation coefficients of 0.015 dB/km and 0.0165 dB/km, respectively. These numbers are similar to the attenuation coefficient of mode 1 measured on the Turpan-to-SIMI path in the TAP experiment (43.5±2.5 dB and 0.0165±0.001 dB/km, respectively). Moreover, the path-average attenuation coefficients of mode 1 in April 1999 on both ACOUS paths were approximately the same. In contrast to mode 1, the propagation loss of mode 2 in the two ACOUS signals received at APLIS was much higher than that measured in the TAP experiment. The propagation loss of mode 2 was approximately 21 dB and 24 dB in these two ACOUS signals, versus 13±1.5 dB in the TAP experiment. The path-average attenuation coefficient of this mode estimated on the path to APLIS is 0.008–0.009 dB/km, which is approximately two times higher than that measured on the ACOUS path to the Lincoln Sea. The excessive attenuation of mode 2 on this path was, most likely, a result of strong interaction of this mode with the sea floor in the region of the Chukchi Plateau, which will be shown in the next section.

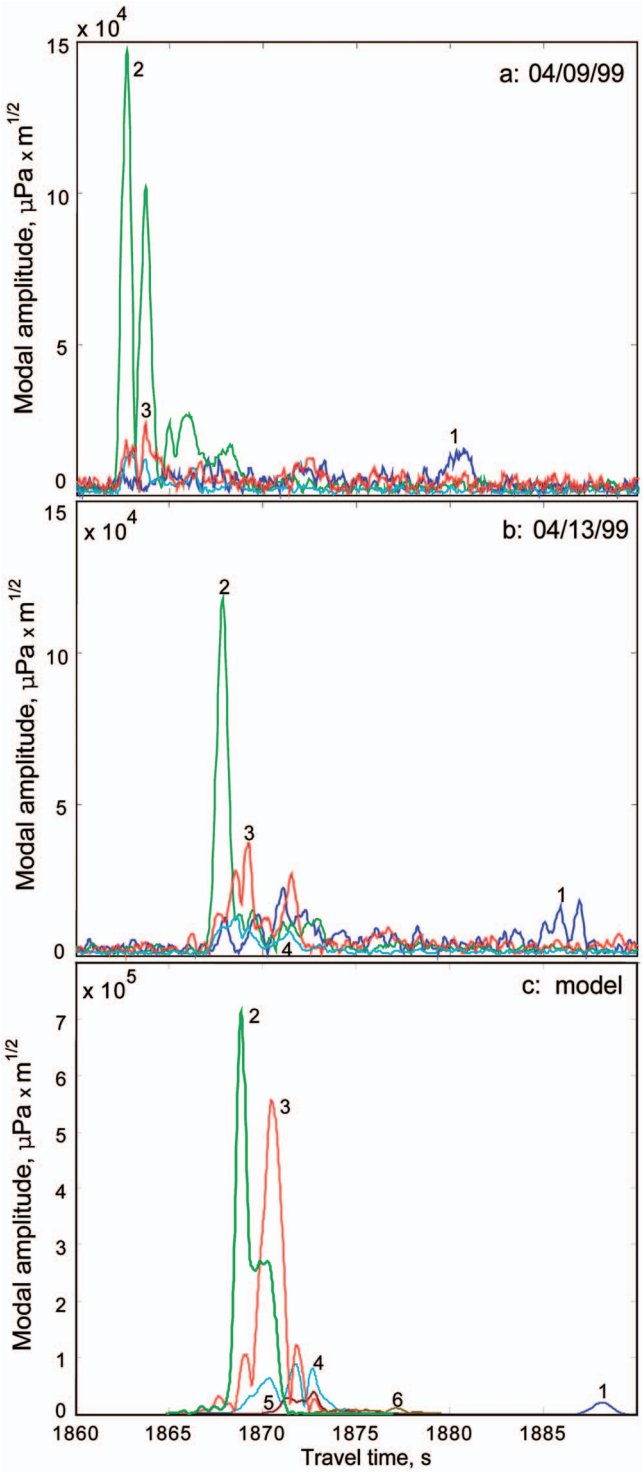


FIG. 7. The arrival patterns of modes 1–4 of the ACOUS signal filtered on the vertical array at ice camp APLIS on 9 April at range R=2714.84 km from the source (a) and 13 April at range R=2722.44 km (b). The bottom panel (c) shows the modeled arrival patterns of modes 1–6 of the ACOUS signal on the APLIS array at R=2722.44 km. The amplitude of mode 1 is shown doubled in all three plots.

IV. MODEL OF THE ACOUS SIGNAL PROPAGATION

Bathymetry and acoustic properties of the seafloor do not change in time, and hence the temporal variations of the acoustic propagation loss on the ACOUS path could only result from changes in the ice cover and variations of the sound speed field along the path. The morphological characteristics of the ice cover, such as the height and correlation length of ice roughness, influence directly the acoustic propagation through ice scattering. Variations of the sound speed profiles alter the interaction of acoustic waves with the seafloor and the sea surface, which may also influence the propagation loss. In this section, the most likely causes of the observed propagation losses and their temporal variations on the ACOUS paths will be examined using the results of numerical modeling.

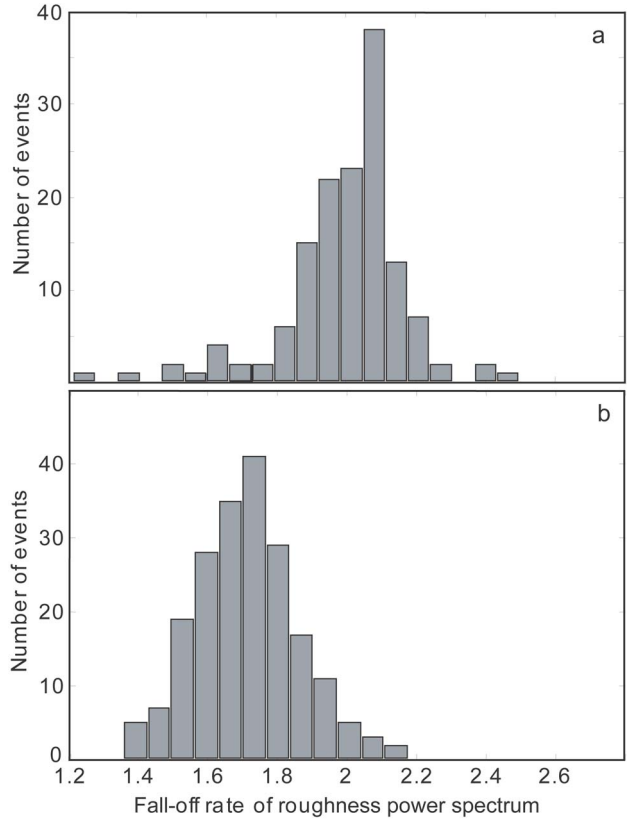


FIG. 8. Histograms of the fall-off rate of the ice roughness power spectra calculated for the ice draft profiles that were observed in the spring of 1991 (a) and the autumn of 1996 (b). The ice roughness spectrum was calculated for each 500-m segment of the ice draft profile and then averaged within 50-km sections.

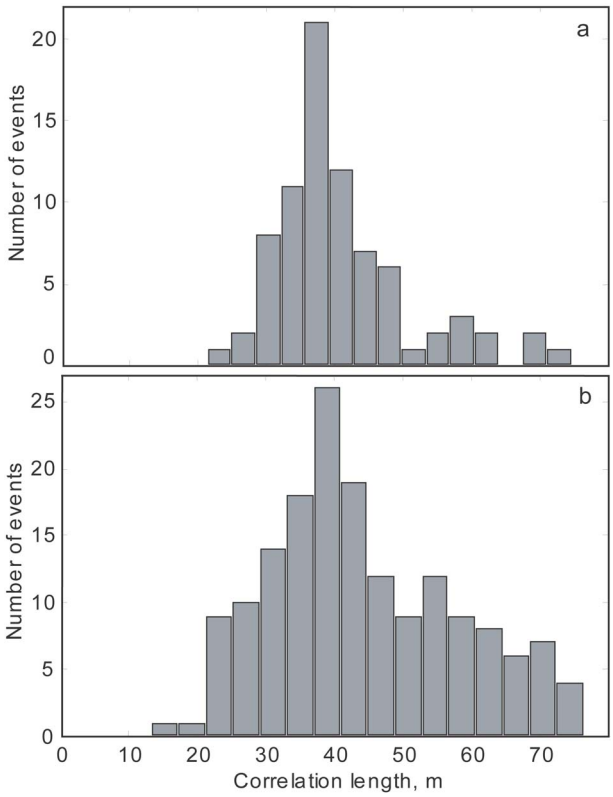


FIG. 9. Histograms of the ice roughness correlation length calculated for the ice draft profiles that were observed in the spring of 1991 (a) and the autumn of 1996 (b). The correlation function was derived from the power spectra averaged within 50-km sections.

A. Ice scattering loss

The coherent reflection coefficient from Arctic sea ice was computed with the use of Kudryshov's ice scattering model.[18] Proper determination of the ice statistic characteristics including the roughness spectrum is critical for ice reflection and scattering modeling.[19,20] The morphological characteristics of the Arctic ice canopy have substantially changed over the last 2 decades.[11] For that reason, we analyzed the data on Arctic ice draft available from the recent submarine upward-looking sonar observations, in order to determine the most likely spectrum of the ice roughness and the typical numbers of mean ice thickness, standard deviation, and roughness correlation length. The ice roughness spectra were calculated for 500-m segments of the ice draft profile and then averaged within each 50-km section. The mean ice thickness and standard deviation of ice roughness were also calculated for each 50-km section. The analysis showed that the spectral fall-off rate of 2 would be most probable for a one-dimensional (1D) ice draft profile in the Arctic winter season [Fig. 8(a)], while it was somewhat smaller and less definite in the summer season [Fig. 8(b)]. For the Goff-Jordan model of the roughness spectrum,[21] the frequency dependence f^2 of a 1D roughness spectrum corresponds to a fractal dimension of 2.5 and the normalized two-dimensional (2D) power spectrum

$$G(\mathbf{u}) = (2\pi)^{-2} l_0^2 [1 + (\mathbf{u} l_0)^2]^{-3/2}, \tag{3}$$

where l_0 is the roughness correlation length. The roughness power spectrum (3) corresponds to azimuthal isotropic roughness with the correlation function $R(l) = \exp(-|r|/l_0)$.

The distribution of the ice roughness correlation length derived from the ice draft data is shown in Fig. 9. In the winter season, the correlation length of 35–40 m is apparently most probable. In Arctic summer, the correlation length of about 40 m is also frequent, but the variance is much greater than that in winter. A large variance of the ice roughness correlation length results from ice-free zones spreading in summer.

The ice statistic parameters, such as the mean ice thickness and roughness standard deviation, vary considerably over the Arctic Ocean regions and seasons. According to the most recent observations, in winter the mean ice thickness is about 2–3 m in the Canada Basin and the Beaufort Sea and 3–4 m in the Fram and Nansen Basins. In summer, the mean ice thickness decreases to 1.5–2 m in the Canada Basin and 2–3 m in the Fram and Nansen Basins. In contrast to the ice thickness, the relationship between the roughness standard deviation and the mean thickness of ice is much less variable. In winter, the standard deviation is most likely 0.6–0.65 of the mean thickness, as can be seen in Fig. 10(a). In summer, the values from 0.5 to 0.7 are almost equiprobable [Fig. 10(b)].

According to the Archimedean principle, the roughness of both surfaces of a floating ice sheet must be partly corre-

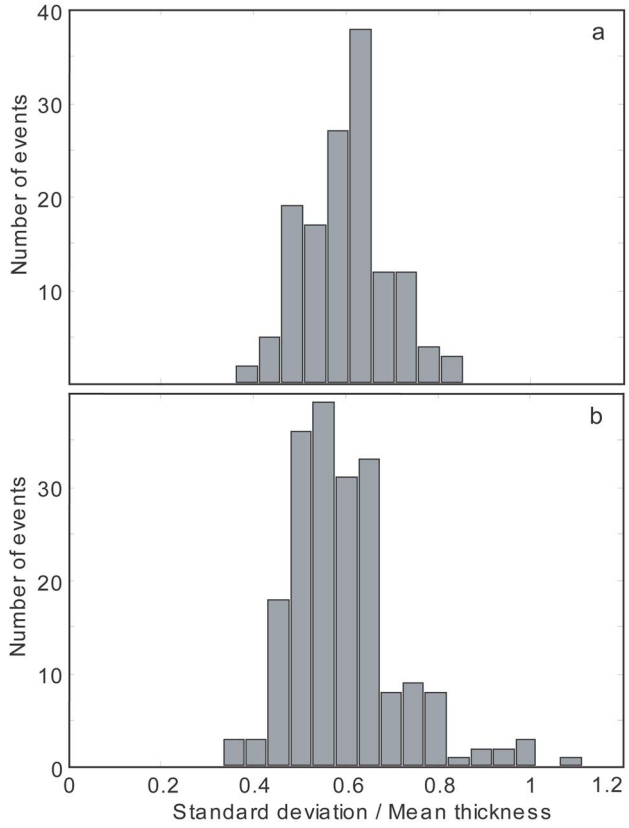


FIG. 10. The ratio of roughness standard deviation and mean ice thickness averaged within 50-km sections of the ice draft profiles made in spring of 1991 (a) and the autumn 1996 (b).

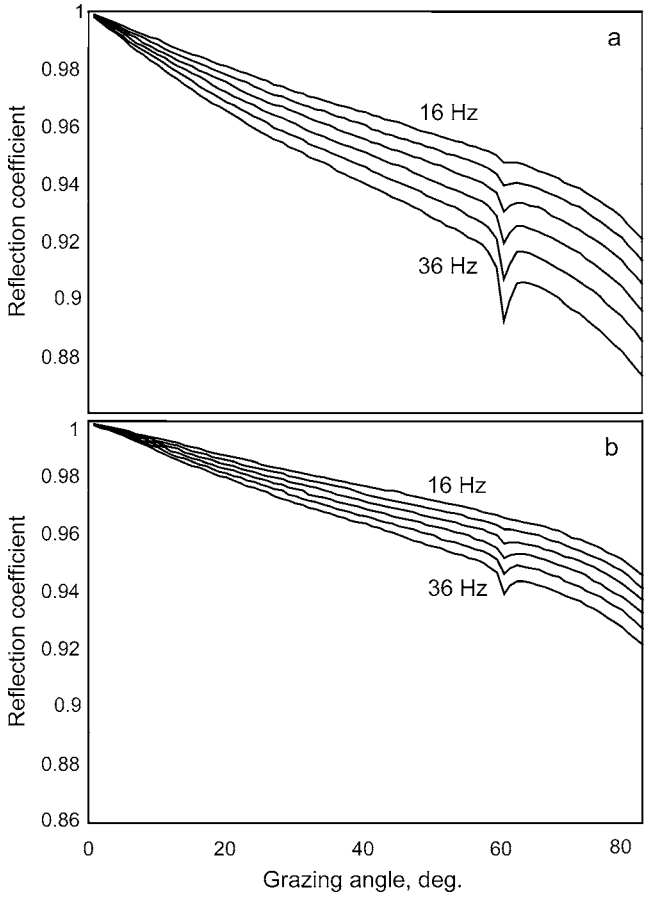


FIG. 11. Coherent reflection coefficient versus grazing angle at frequencies 16, 20, 24, 28, 32, and 36 Hz for Arctic sea ice with the typical winter parameters [(a) mean thickness=3.0 m; standard deviation of bottom ice boundary=1.8 m; standard deviation of top ice boundary=0.45 m; correlation length=40 m] and the typical summer parameters [(b) mean thickness =2.0 m; standard deviation of bottom ice boundary=1.2 m; standard deviation of top ice boundary=0.3 m; correlation length=40 m]. The following acoustic properties of ice were used in the model: compressional wave velocity=3000 m/s; shear wave velocity=1800 m/s; compressional wave attenuation=0.45 dB/λ; shear wave attenuation=0.9 dB/λ; relative density=0.91.

lated and have similar spectra especially at smaller wavenumbers, for which the spatial periods of variation are much larger than the ice thickness. In Kudryashov's ice scattering model, the correlation between the top and surface ice roughness is taken into account and characterized by the correlation coefficient and the ratio of standard deviation of the top and bottom ice roughness. The estimates of these two parameters available from *in-situ* measurements at drifting ice stations "North Pole" are as follows: the standard deviation of the top surface is about four times less than that of the bottom surface, and the correlation between the top and bottom roughness is about 0.7. These estimates were used in the ice scattering model. Note that the influence of the top-side roughness on the ice scattering loss modeled at low frequencies, such as 20 Hz, is stronger than the effect of correlation between two rough surfaces. Consequently, knowing the value of standard deviation of the top-side ice roughness is important for accurate numerical prediction of the acoustic propagation loss due to ice scattering. On the other hand, it is an established fact[22] that the heights of top and bottom ice roughness vary almost synchronously with seasonal growth and melting of sea ice. Therefore the results of numerical prediction for the relative change of acoustic propagation loss due to the seasonal variation of ice statistic characteristics should not be particularly sensitive to the roughness standard deviation ratio accepted in the model.

Figure 11(a) demonstrates the magnitude of the ice reflection coefficient versus grazing angle calculated at six frequencies spaced at 4 Hz from 16 Hz to 36 Hz, for the ice parameters expected in winter in the Central Arctic Basin. Small spikes at a grazing angle of 60° coincide with the critical angle for the longitudinal waves in ice. The angular dependence of the ice reflection coefficient shown in Fig. 8(b) was calculated for the typical ice characteristics observed in the Nansen Basin in August–October in the 1990s. The modal attenuation coefficients were calculated for the sound speed field along the ACOUS paths shown in Figs. 2 and 3 using the KRAKENC code[23] with the modeled ice reflection coefficients as the top boundary conditions. Figures 12(a) and 12(b) show the variation of the attenuation coefficients of modes 1–4 at 20.5 Hz along the path to the Lincoln Sea calculated for the ice parameters given in the captions of Figs. 11(a) and 11(b), respectively. The total propagation losses of modes 1–5 on the ACOUS path derived from the attenuation coefficients are shown in Fig. 13. It is evident from the comparison of modeling results in Fig. 13, that seasonal change in the ice cover may lead to noticeable change of the modal propagation loss at 20 Hz on a 1 -Mm transarctic path, especially for mode 1. Moreover, the

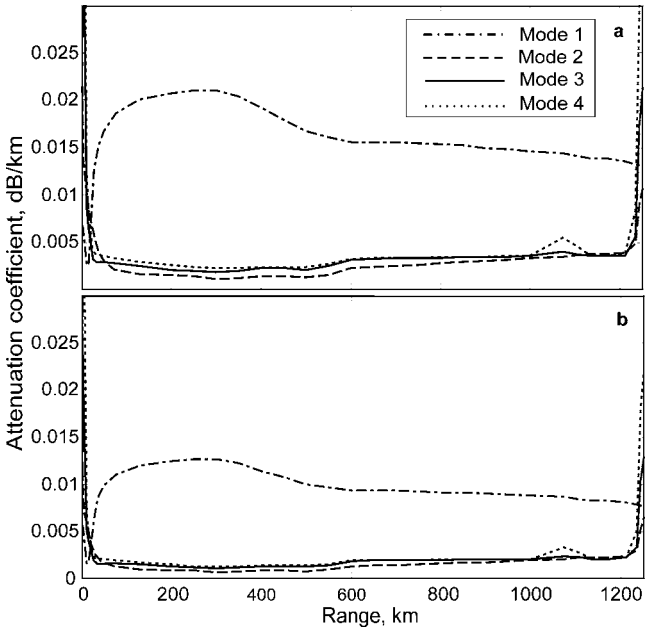


FIG. 12. Attenuation coefficients of modes 1–4 at 20.5 Hz along the ACOUS path modeled for the winter (a) and summer (b) ice conditions corresponding to the ice reflection coefficient shown in Figs. 11(a) and 11(b), respectively. This model prediction was obtained using the sound speed profiles along the ACOUS path shown in Fig. 2.

total propagation losses of mode 1 and the group of higher modes modeled for the winter and summer conditions are comparable to the numbers obtained from the ACOUS observations in the corresponding seasons (shown in Fig. 6).

For lack of recent upward looking sonar data on ice draft for the regions north of Greenland and the Canadian Archipelago, we assumed in the numerical model that the ice cover characteristics were uniform everywhere along the ACOUS path. However, it is known from the earlier observations of ice draft in the 1970s–1980s (Ref. 24) that the mean thickness and standard deviation of sea ice in the Lincoln Sea and the adjacent regions were commonly larger than those in the Nansen and Fram Basins. Hence the modal attenuation derived from the numerical model for the final 200–300 km section of the ACOUS path could be somewhat underestimated relative to the real conditions.

Note that the attenuation coefficient of individual modes also varies with the horizontal variation of the sound speed profile along the path, which influences the modal cycle lengh. Change of the modal attenuation due to sound speed variations is discussed in more detail in the next section.

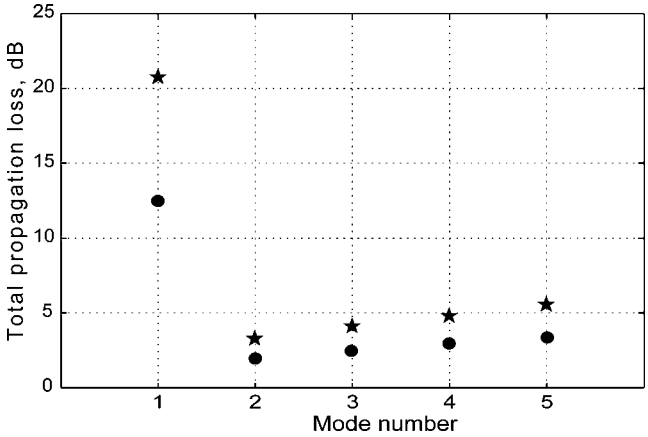


FIG. 13. Total propagation loss of modes 1–5 on the ACOUS path to the Lincoln Sea, disregarding cylindrical spreading, modeled for the winter (stars) and summer (dots) ice conditions and the sound speed field shown in Fig. 2.

B. Effects of mode coupling and variation of the sound speed profile

The propagation of a 20-Hz acoustic signal from the Franz Victoria Strait to the Lincoln Sea was modeled earlier, before the ACOUS experiment, using the climatology sound speed profiles and the ETOPO5 bathymetry database. The coupled-mode algorithm used for numerical modeling and the results of modeling are thoroughly considered in Ref. 25, where it is shown that the mode coupling greatly influences the modal propagation loss along the path. It is also shown that temporal variations of the sound speed over any rapidly or slowly varying section of the path may lead to large variations of the modal propagation loss due to changes in the mode coupling effects over range-dependent sections. The newer modeling results presented below were obtained for the actual location of the ACOUS source and the Lincoln Sea array with the use of the recent International Bathymetry Chart of the Arctic Ocean (IBCAO) database[26] which is much more accurate than ETOPO5. Moreover the real sound speed profiles used in the new model are more realistic than the climatology data.

Acoustic interaction with the seabed increases attenuation of the signals propagated over shallow-water sections of the acoustic path. For the ACOUS paths, such sections are the northern edge of the Eurasian continental shelf in the Franz Victoria Strait, approximately 15 km long; the northern edge of the Canadian continental shelf in the Lincoln Sea, about 20 km long; and the Chukchi Plateau that spans a 500-km section of the ACOUS path to APLIS near its end. Over these regions, almost every mode of the ACOUS signal underwent significant interaction with the seabed. In the deep basins, interaction of the ACOUS signals with the seabed was negligible. Acoustic properties of the seabed in the Lincoln Sea and Franz Victoria Strait are more or less known,[27] while no geoacoustic investigation of the seabed on the Chukchi Plateau has yet been done. For that reason, the geoacoustic characteristics typical for the Arctic shelf regions, such as the Lincoln and Barents Seas, were used for acoustic modeling the seabed in the region of the Chukchi Plateau.

Variations of the modal amplitudes at the ACOUS carrier frequency along the path to the Lincoln Sea were modeled earlier (see Fig. 3 in Ref. 14). In the primary (unperturbed) environmental model, the sound speed field along the path (Fig. 2) was compiled from (1) the oceanographic data of a 100-km long oceanographic section across the Eurasia continental slope north of the Franz Victoria Strait in 1993, (2) several temperature/salinity (T/S) profiles of the 1990s at distances from 150 to 550 km, (3) the climatology T/S profiles from 600 to 1150 km, and (4) the ICESHELF oceano-

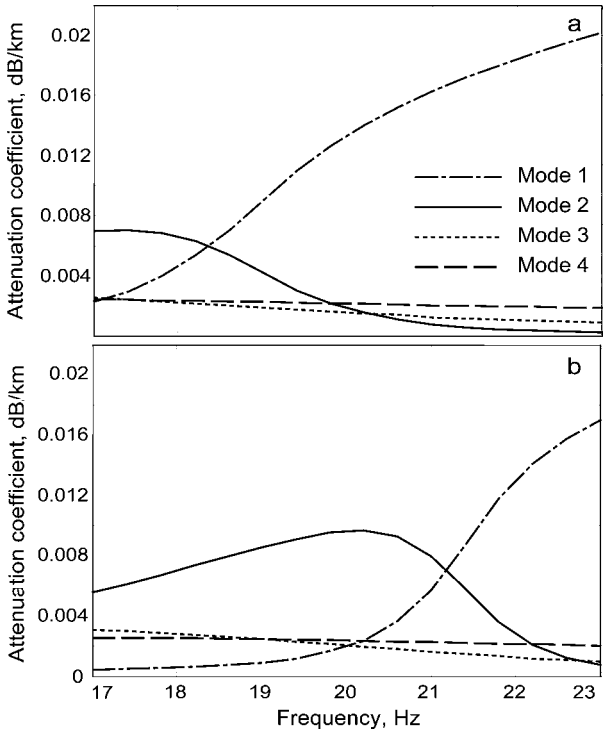


FIG. 14. Attenuation coefficients of modes 1–4 versus frequency for the ice reflection coefficient shown in Fig. 11(b) and the typical (a) and "warm" (b) sound speed profiles shown in Fig. 6 in Ref. 14.

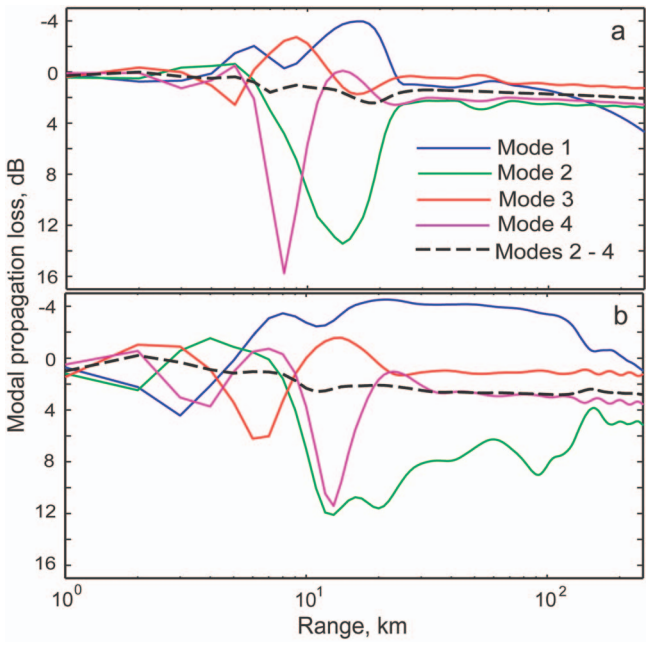


FIG. 15. Variations of the propagation loss of modes 1–4 (disregarding cylindrical spreading) in the frequency band of 20.5±1 Hz modeled along the initial 250-km section of the ACOUS path to the Lincoln Sea for two different environmental conditions: (a) the typical conditions compiled from the sound speed profiles observed in the 1990s and (b) the abnormally "warm" conditions reconstructed through inversion of the modal travel times and shapes observed on the Lincoln Sea array in November–December 1999. The black dashed lines show the range dependence of the propagation loss for the group of modes 2–4.

graphic data of 1990s in the Lincoln Sea.[28] It was shown that all modes suffered strong coupling along the initial and final section sections over the continental slopes. Moreover, a significant increase of the AIW temperature within the initial 300-km section in the Nansen Basin that was detected in August–December 1999 via inversion of the modal travel times and shapes, should also influence the mode-coupling effects and the propagation loss of individual modes. In the oceanographic conditions typical for the deep-water Arctic Basin, mode 1 at 20 Hz is trapped in the upper Arctic acoustic duct formed by the sea surface and the thermocline. As a result, this mode propagates considerably slower and attenuates faster than the higher, mid-depth refracted modes. However, if the maximum AIW temperature exceeds 2.5 °C and the AIW layer expands above a 100-m depth, which took place north of the ACOUS source in the Nansen Basin in autumn 1999, the upper duct cannot trap mode 1 at 20 Hz, which makes this mode propagating faster and attenuating slower than mode 2. Figure 14 shows the attenuation coefficients of modes 1–4 versus frequency calculated for the typical (a) and "warm" (b) temperature and sound speed profiles observed in the Nansen Basin north of the ACOUS source (see Fig. 6 in Ref. 14). The effective frequency band of the ACOUS signal is approximately 20.5±1 Hz. Hence relatively small and likely variations of the AIW characteristics may lead to change of the critical frequency of modes 1 in the upper acoustic duct, which greatly affects the attenuation coefficient of this mode. In contrast to mode 1, the attenuation of mode 2 increases in this transient frequency band.

Mode coupling behaves in a different way at different frequencies. Therefore, when modeling spatial variations of the modal amplitudes of the ACOUS signal, we calculated the rms amplitude and propagation loss of individual modes within the frequency band of 20.5±1 Hz. The influence of the sound speed field on the propagation loss of modes 1–4 (color lines) along the initial 250-km section of the ACOUS path is demonstrated in Fig. 15. The propagation losses shown in the top panel were calculated for the unperturbed sound speed field that approximately corresponded to the environmental conditions in winter and spring 1999, and the winter ice conditions. The bottom panel shows the result of modeling for the sound speed field reconstructed through inversion of the modal travel times and shapes observed on the Lincoln Sea array in November–December 1999 (see Fig. 11 in Ref. 14). It is seen from a comparison of modeling results that modes 1 and 2 are most sensitive to the change in the environmental conditions reconstructed from the ACOUS observations. Attenuation of mode 1 considerably decreases in the "warmer" conditions, while the propagation loss of mode 2 increases. The dashed lines in Fig. 15 show the group propagation loss for modes 2–4. This modal group is subject to much weaker amplitude variations due to mode coupling than the individual modes. This fact is not surprising, because the intermodal energy exchange caused by mode coupling is most intense between the modes of adjacent numbers. Also note that such significant change in the environmental conditions did not affect much the group propagation loss of modes 2–4 at the end of this 250-km section. Moreover, the level of group energy of modes 2–4 decreased a little, by roughly 0.5 dB, in the "warmer" conditions, while the level of mode 1 increased by 4 dB. Numerical modeling of the modal propagation loss along the

initial section of the ACOUS path has been performed for various hydrological conditions observed in the northern Franz Victoria Strait and the Nansen Basin. The modeled spatial variations of the amplitude of individual modes are different for different environmental conditions, but the magnitude of variation and the group propagation loss of modes 2–4 always remain similar to those shown in Fig. 15.

Based on the results of numerical modeling, one can conclude that the short-term fluctuations of the modal amplitudes on the Lincoln Sea array were most likely a result of mesoscale variations of the sound speed profiles, which occurred primarily within the initial 200–300-km section of the path and altered substantially the mode coupling effects. According to historical oceanographic data, the sound speed profiles in the Fram Basin and the Lincoln Sea are much less variable in time than those in the Franz Victoria Strait and the Nansen Basin. Consequently, the influence of sound speed variations in the Fram Basin and the Lincoln Sea on the propagation loss of individual modes and modal groups in particular is expected to be much smaller than that demonstrated in Fig. 15.

Comparing the results of numerical prediction for the modal propagation loss due to ice scattering in different seasons (Fig. 13) with the numerically examined variations of the modal amplitudes caused by change in the sound speed field, one can conclude that the seasonal change of the modal propagation loss observed on the ACOUS path (Figs. 5 and 6) was most likely a result of the seasonal change in the sea ice cover. According to Fig. 15, the change in the hydrological conditions in the Nansen Basin in autumn–winter 1999 could also contribute to the large decrease of the total propagation loss of mode 1 observed in the experiment. However, as follows from Fig. 5, the contribution of mode 1 to the total energy of all modes filtered on the Lincoln Sea array is very small, so that the long-term variation of the modal group propagation loss can be attributed primarily to the seasonal change in the sea ice cover.

To verify this conclusion, it is also necessary to investigate if variations of the sound speed profile along the shallow-water sections and continental slopes could affect the modal propagation loss in a different way if the seabed acoustic properties are different. In order to examine different bottom scenarios, the group propagation loss of modes 2–4 with range along the initial 250-km section was also calculated for a different acoustic model of the seabed. Based on the available geoacoustic data, the primary acoustic model of the seabed in the northern Franz Victoria Strait was accepted to be a "sand-clay" model, of which the principal parameters of the upper 50-m fluid sediment layer were as follows: sound speed at the water-sediment interface of 1800 m/s; sound speed gradient of 2 s^{-1}; density of 1800 kg/m^3; and attenuation of 0.1 dB/km Hz. In the alternative model, these parameters were substantially modified: the sound speed at the interface decreased to 1600 m/s; the gradient decreased to 1 s^{-1}; the density changed to 1700 kg/m^3; and attenuation increased to 0.3 dB/km Hz. Figure 16 demonstrates the effect of different acoustic models of the seabed on the group propagation loss of modes 2–4 in two different environmental conditions discussed earlier in this section. The attenuation of modal group energy noticeably increased over an acoustically softer and more absorbing sediment (red lines) relative to the primary model of the seafloor (blue lines). However, the changes of attenuation are very similar in different hydrological conditions designated by the solid and dashed lines. This means that the numerical prediction of modal propagation loss might have some absolute, but time-invariant error, if the acoustic model of the seabed in the shallow-water regions was chosen incorrectly. This also means that the influence of sound speed variations on the modal propagation loss is invariant to some extent under changes in the seabed properties. Thus the conclusions made from the results of numerical modeling with respect to the seasonal change of the modal propagation loss should be valid even if the acoustic model of the seabed was not accurate.

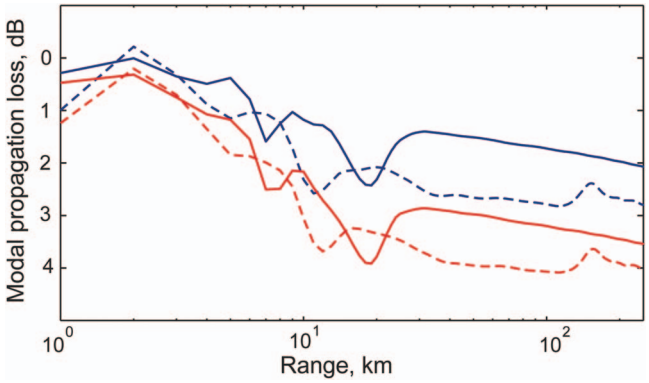


FIG. 16. Variations of the group propagation loss of modes 2–4 (disregarding cylindrical spreading) in the frequency band of 20.5±1 Hz modeled along the initial 250-km section of the ACOUS path for the typical (solid) and "warm" (dashed) hydrological conditions and for the data-based (blue) and alternative (red) acoustical models of the seabed.

Propagation losses of modes 1–4 modeled along the acoustic path to ice camp APLIS are shown in Fig. 17. Mode coupling is significant for every mode along the initial section over the continental slope. The Lomonosov Ridge (at 950–1000 km) does not influence mode 1, affects slightly modes 2 and 3, and causes strong coupling of modes 4 and higher. For the accepted model of bathymetry along the path, the Chukchi Plateau should not influence mode 1. The other modes are subject to significant coupling over the Chukchi Plateau. This numerical prediction is partly confirmed by the ACOUS signals received on the APLIS array. The width of the split pulse of mode 2 [Fig. 7(a)] indicates that strong coupling of mode 2 with the higher modes occurred along the final section of the propagation path over the Chukchi Plateau. In contrast to the numerical prediction of the modal pulse forms on the APLIS array shown in Fig. 7(c), the arrival pulse of mode 3 in the received signals was very weak and deformed. Also the propagation loss of mode 2 predicted from the numerical model is much lower than that measured at APLIS. An obvious discrepancy between the experimental and modeled arrival patterns of modes on the APLIS array is most likely a result of incorrect modeling of the bathymetry and acoustic properties of the sea floor along the section of the acoustic path over the Chukchi Plateau.

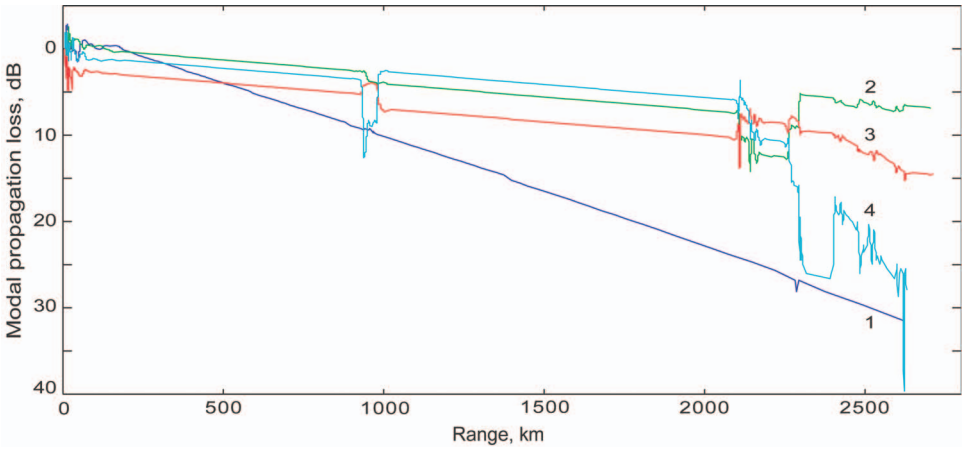


FIG. 17. Propagation loss of modes 1–4 modeled at the central frequency of the ACOUS signal along the path to APLIS, disregarding cylindrical spreading. The sound speed vertical section along the path was combined from (1) the historical oceanographic data of 1990s for the initial 350-km section of the path and (2) the SCICEX-1999 T/S profiles observed along the major section from 350 km to 2720 km.

V. CONCLUSIONS

The ACOUS experiment has proven the capability of low-frequency cross-Arctic acoustic transmissions to observe long-term basin-scale changes in the Arctic ice cover. The seasonal cycle in the variations of the total energy loss of the transmitted signal and the propagation loss of individual modes was observed. According to the theory of ice scattering, coherent losses of the low-frequency signals propagated under ice in the Arctic Ocean depend primarily on the height and correlation length of the ice roughness. As proven by upward-looking sonar observations of sea ice draft, the height of ice roughness closely mirrors the change in the mean ice thickness through a nearly linear relationship. For ice draft in excess of 1 m, the standard deviation of the bottom ice interface is about 0.5–0.6 of the mean ice thickness. Seasonal melting and growing of sea ice result basically in variations of the mean ice thickness and standard deviation of ice roughness, which influences considerably the acoustic propagation losses and hence can be observed by means of long-range acoustic transmissions, as was done in the ACOUS experiment.

The results of propagation loss modeling along the ACOUS path to the Lincoln Sea, using the most recent ice draft data and the model of seasonal change of ice thickness, are basically in agreement with the results of experimental observations on this path. This allows us to derive the correlation between the modal propagation loss and ice thickness and hence gives a way to build an acoustic inversion method for remote observations of ice thickness. However, the accuracy of such acoustic monitoring of Arctic sea ice needs to be examined in the future experiments by means of ice draft profiling conducted along the acoustic path at the time of acoustic observations.

Attenuation of mode 1 at 20 Hz is much more sensitive to changes of the ice cover than the higher modes. Remote observations of sea ice by means of low-frequency cross-Arctic acoustic transmissions would be much more accurate, if mode 1 is not subject to significant interaction with the sea floor and strong coupling with the higher modes due to variable bathymetry and horizontal changes in the sound speed profile. Such propagation conditions for mode 1 can be achieved through choosing either better locations of the source and the receiving array (e.g., moving the source and receivers further out the shelves into deeper waters) or a more suitable frequency band of the signal. For example, at 23 Hz mode 1 would experience somewhat weaker interaction with the sea floor along the initial section of the ACOUS path and much less significant coupling with the higher modes over the frontal zone in the Nansen Basin than that observed at 20.5 Hz in the ACOUS experiment.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research and the Civilian Research and Development Foundation (Grant No. RG2-2407). The authors would like to express their deep appreciation to John M. Lents of SAIC who managed field operations in the ACOUS experiment. The authors also thank Vladimir Kudryashov for valuable suggestions and comments in relation to the algorithm for numerical modeling of the coherent acoustic reflection from Arctic sea ice.

[1]B. M. Buck and C. R. Green, "Arctic deep-water propagation measurements," J. Acoust. Soc. Am. **36**, 1526–1533 (1964).
[2]F. DiNapoli and R. Mellen, "Low frequency attenuation in the Arctic Ocean," in *Ocean Seismo Acoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York, 1986), pp. 387–395.
[3]E. Livingston and O. Diachok, "Estimation of average under-ice reflection amplitudes and phases using matched-field processing," J. Acoust. Soc. Am. **86**, 1909–1919 (1989).
[4]P. N. Mikhalevsky, A. B. Baggeroer, A. N. Gavrilov, and M. M. Slavinsky, "Experiment tests use of acoustics to monitor temperature and ice in the Arctic Ocean," EOS (Wash. D.C.) **76**, 265–269 (1995).
[5]R. Pawlowicz, D. Farmer, B. Sotirin, and S. Ozard, "Shallow-water receptions from the transarctic acoustic propagation experiment," J. Acoust. Soc. Am. **100**, 1482–1492 (1996).
[6]P. N. Mikhalevsky, A. N. Gavrilov, and A. B. Baggeroer, "The transarctic acoustic propagation experiment and climate monitoring in the Arctic," IEEE J. Ocean. Eng. **24**, 182–202 (1999).
[7]G. Duckworth, K. LePage, and T. Farrel, "Low-frequency long-range propagation and reverberation in the Central Arctic: Analysis of experimental results," J. Acoust. Soc. Am. **110**, 747–760 (2001).
[8]B. Rudels, E. P. Jones, L. G. Anderson, and G. Kattner, "On the intermediate depth waters of the Arctic Ocean," in *The Polar Oceans and their Role in Shaping the Global Environment*, edited by Johannessen *et al.* (American Geophysical Union, Washington, DC, 1994), pp. 33–46.
[9]Environmental Working Group (EWG), Joint U.S. Russian Atlas of the Arctic Ocean (CD-ROM), National Snow and Ice Data Center, Boulder, CO, 1998.
[10]DMSP SSM/I daily and monthly polar gridded sea ice concentrations (National Snow and Ice Data Center, University of Colorado; digital data from http://nsidc.org/data/nsidc-0002.html, last viewed in January 2004).

[11]D. A. Rothrock, Y. Yu, and G. A. Maykut, "Thinning of the Arctic sea-ice cover," Geophys. Res. Lett. **26**, 3469–3472 (1999).
[12]Submarine upward looking sonar ice draft profile data and statistics (National Snow and Ice Data Center University of Colorado; digital data from ftp://sidads.colorado.edu/pub/DATASETS/NOAA/G01360/, last viewed in January 2004).
[13]A. N. Gavrilov and P. N. Mikhalevsky, "Phase acoustic tomography with broadband signals in the Arctic," in *Proceedings of the 3-th European Conf. on Underwater Acoustic*, edited by J. S. Papadakis (FORTH—IACM, Heraklion, Crete, 1996), Vol. **2**, pp. 851–856.
[14]A. N. Gavrilov and P. N. Mikhalevsky, "Recent results of the ACOUS (Arctic Climate Observations using Underwater Sound) Program," Acta. Acust. Acust. **88**, 783–791 (2002).
[15]P. N. Mikhalevsky and A. N. Gavrilov, "Acoustic thermometry in the Arctic Ocean," in *Proceedings of the 5-th European Conf. on Underwater Acoustic*, edited by M. Zakharia (LASSSO-ESCPE, Lyon, France, 2000), Vol. **2**, pp. 1291–1296.
[16]T. C. Yang, "A method of range and depth estimation by modal decomposition," J. Acoust. Soc. Am. **82**, 1736–1745 (1987).
[17]R. Coifman, Y. Meyer, S. Quake, and V. Wickerhauser, "Entropy based algorithms for best basis selections," IEEE Trans. Inf. Theory **38**, 713–718 (1992).
[18]V. M. Kudryashov, "Sound reflection from ice cover," Phys. Acoust. **42**, 247–253 (1996).
[19]K. LePage and H. Schmidt, "Modeling of low-frequency transmission loss in the central Arctic," J. Acoust. Soc. Am. **96**, 1783–1795 (1994).
[20]W. A. Kuperman and H. Schmidt, "Self-consistent perturbation approach to rough surface scattering in stratified elastic media," J. Acoust. Soc. Am. **86**, 1511–1522 (1989).
[21]J. Goff and T. Jordan, "Stochastic modeling of seafloor morphology: version of Sea Beam data for second order statistics," J. Geophys. Res. **93**, 13589–13608 (1988).
[22]H. Eicken, W. B. Tucker, and D. K. Perovich, "Indirect measurements of the mass balance of summer Arctic sea ice with an electromagnetic induction technique," Ann. Glaciol. **33**, 194–200 (2001).
[23]M. B. Porter, "The KRAKEN normal mode program," Technical Report No. SM-245, SACLANT Undersea Research Center, La Spezia, Italy, 1991.
[24]R. H. Bourke and A. S. McLaren, "Contour mapping of Arctic Basin ice draft and roughness parameters," J. Geophys. Res., [Oceans] **97**, 17715–17728 (1992).
[25]A. N. Gavrilov and P. N. Mikhalevsky, "Mode-coupling effects in acoustic thermometry of the Arctic Ocean," in *Inverse Problems in Underwater Acoustics*, edited by M. Taroudakis and G. Makrakis (Springer-Verlag, New York, 2001), pp. 105–125.
[26]International Bathymetric Chart of the Arctic Ocean (Editorial Board for IBCAO; digital data from http://www.ngdc.noaa.gov/mgg/bathymetry/arctic/arctic.html, last viewed in January 2004).
[27]W. H. Geddes, "Geoacoustic model of the Lincoln Sea," Technical Report No. 3-90, Geddes Geophysical Associates, 1990.
[28]J. L. Newton and B. J. Sotirin, "Boundary undercurrent and water mass changes in the Lincoln Sea," J. Geophys. Res. **102**, 3393–3403 (1997).

Geoacoustic inversion of broadband data by matched beam processing

Yongmin Jiang[a] and N. Ross Chapman[b]
School of Earth and Ocean Sciences, University of Victoria, P.O. Box 3055, Victoria, British Columbia V8W 3P6, Canada

Harry A. DeFerrari
Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, 4600 Rickenbacker Cswy, Miami, Florida 33149

(Received 13 December 2005; revised 10 March 2006; accepted 14 March 2006)

This paper describes results of geoacoustic inversion using broadband signals from an experiment carried out at a site near the South Florida Ocean Measurement Centre in the Florida Straits. M-sequence coded pulse trains at different center carrier frequencies from 100 to 3200 Hz were recorded at a vertical line array at a distance around 10 km. Geoacoustic inversion was carried out to determine the feasibility of inverting the environmental parameters from this long-range propagation experiment. The received signal at lower frequencies below 400 Hz consisted of a dominant water column signal and a secondary arrival delayed by 0.4 s. The secondary signal was spatially filtered by beamforming the array data, and the beam data were inverted by matched beam processing in the time domain, combined with an adaptive simplex simulated annealing algorithm. The estimated values of compressional wave speed and density were in good agreement with ground truth values from sediment cores. The inverted shear wave speed appears to be a sensitive parameter and consistent with compressional wave speed. As a cross check, range and water depth were also included as inversion parameters, and the inversion results were close to the known values within small uncertainties. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195114]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 3707–3716

I. INTRODUCTION

In ocean acoustics, knowledge of sea bottom properties plays an important role in predicting sonar system performance, especially in shallow water and littoral environments where sound propagation is dominated by multiple bottom interacting paths. Since the early 1990s, there has been considerable research effort in developing inversion techniques for estimating the parameters of geoacoustic models of the ocean bottom from acoustic field data.[1–10] One of the most effective techniques is matched field inversion (MFI), which is based on matched field processing (MFP) of data obtained at an array of sensors. In addition to the conventional approach using complex pressure data of spectral component, other approaches include applications using of amplitude and phase information in a time series[11,12] (matched wave form processing). Overall, MFI is in widespread use, and has proved to be very successful in applications with low-frequency ($<$500 Hz) experimental data.

Matched field inversion is usually implemented as an optimization problem that searches a multidimensional geoacoustic model parameter space to determine the model that provides the best fit to the acoustic field data. The geoacoustic models generally consist of profiles of sound speed, density, and attenuation in the sea bottom. Matched field inversions are highly sensitive to the sound speed, but are less sensitive to density and attenuation.[13,14] The recent benchmarking exercises to calibrate the performance of MFI techniques have shown that reliable estimates of attenuation are very difficult to obtain from conventional short range experimental geometries with vertical line arrays.

This paper describes geoacoustic inversion of data acquired in an acoustic propagation experiment designed to investigate signal fluctuations and coherence at a site near the South Florida Ocean Measurement Centre in the Florida Straits.[15] In this experiment, M-sequence coded signals were transmitted at frequencies from 100 to 3200 Hz over a range of 10 km to a sparsely populated vertical array. Signal propagation was parallel to the coast along the contour of the continental slope at a water depth of 155 m. Previous analysis of the data showed that signal propagation was dominated by relatively shallow angle paths ($<\pm 10°$) with multiple bottom interactions.[16] A unique feature of the received signal is a strong coherent secondary arrival that is seen in the low frequency data ($<$800 Hz).

The long-range geometry and sparse array design presented significant challenges for an inversion based on conventional MFP, which makes use of the spatial phase relationships of signal spectral components across the array. Our initial attempts using conventional MFI were not successful in generating a geoacoustic model that could reproduce the strong secondary signal that was observed in the data.[17] Those initial inversions were based on the hypothesis that the secondary arrival was due to the subbottom structure. In contrast, the inversions reported here are based on a different

[a]Electronic mail: minj@uvic.ca
[b]Electronic mail: chapman@uvic.ca

hypothesis that the secondary signal is related to three-dimensional propagation in the local continental shelf environment. The secondary signal is viewed as an interfering event that should be removed from the data.

Our approach in this inversion focuses on the low frequency (100–200 Hz) M-sequence data, and makes use of both the spatial coherence across the array and also the information in the time domain, from the temporal relationships in the multipath signals. The method is based on matched beam processing (MBP), a technique introduced by Yang[18,19] for range tracking and localization in shallow water. The inversion matches the sparse array beam response in the time domain with replica beam time series that are calculated for acoustic fields based on candidate geoacoustic models. A global optimization algorithm, differential evolution, was employed to determine the array weighting parameters for improved beamforming performance, as described later in this paper. The matched beam cost function was used in the adaptive simplex simulated annealing algorithm (ASSA[4]) to generate optimal estimates of the model parameters. This approach allows the use of data from the full array aperture, and provides the option of spatial filtering of the beam response in order to remove the interfering secondary arrival.

The remainder of the paper is organized in the following way. The experimental information and characteristics of signal propagation in the Florida Straits site are described in Sec. II. The matched beam inversion method is then described, and the inversion results are discussed and interpreted in Sec. III. The paper is summarized in Sec. IV.

II. DESCRIPTION OF THE EXPERIMENT

A. Experimental site and geometric setting

The data used for geoacoustic inversion were obtained in an acoustic propagation experiment carried out near the South Florida Ocean Measurement Centre in the Florida Straits.[16] Figure 1 depicts the location of the experimental site and the relative positions of the components of the experimental system. The source was deployed at 26°8.39 N, 80°3.20 W, and the vertical line array at 26°2.92 N, 80°3.769 W. The research ship used a dynamic positioning system connected to a differential global positioning system for precise deployment of the source and receiving arrays, within about 4 m in North/South direction and within 10 m in East/West direction. Consequently, the range between the ships that deployed the source and the receiver was assumed to be accurately determined at 10.14 km with an error of about ±0.02 km.

Two environmental arrays were deployed at ranges of 2.5 and 7.5 km from the source, approximately in the plane of the source and the vertical hydrophone array. The bathymetry data were downloaded from the National Geophysical Data Center.[20] In the right panel of Fig. 1, the grid size of the bathymetry data is 3 s, and the depth contour scale is 5 m. From Fig. 1, the water depths at the source, the receiving array, and the two environment arrays were between 150 and 160 m.

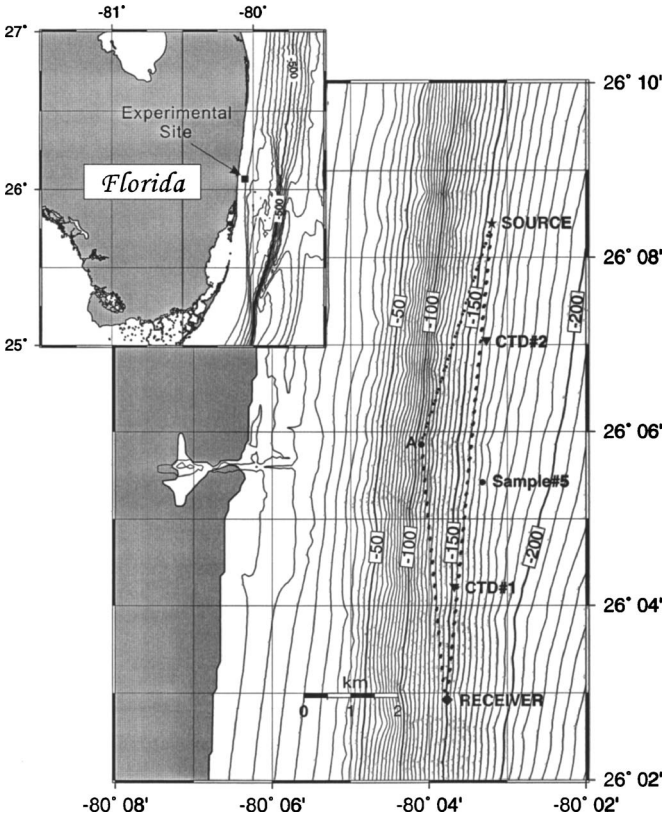


FIG. 1. Experimental site showing the locations of the source, VLA, and the two environmental arrays. Depth contours are 5 m apart in the right panel.

The source consisted of several sets of transducers that transmitted M-sequence signals at carrier frequencies from 100 to 3200 Hz. Source depths were derived from a pressure sensor on the source system. The 100 Hz source was at a depth of 107 m, and the higher frequency sources were slightly deeper at 112 m.

Although the vertical line array consisted of 32 hydrophones with frequency band response from 10 Hz to 25 kHz, the separation of the hydrophones on the array was sparse, and about half the channels were low gain. Consequently only 13 of the array hydrophones were used in this inversion. The effective aperture of the sparse VLA spanned a distance from about 39.8 to 136.7 m in the water column.

B. Transmitted signal

The sound source transmitted M-sequence coded pulses at center frequencies from 100 to 3200 Hz in octave steps. The pulse width was around 2.55 s, as determined by the duration of M-sequence period at each frequency. The bandwidth of the signal at each frequency was 25% of the center frequency. The source transmitted continuously for 1 h at each frequency and cycled through the entire band over a 6 h period. For example, the first hour transmitted an M-sequence coded pulse with carrier frequency of 100 Hz, the next hour was for the 200 Hz signal, then an hour for 400 Hz and so on.

The signals received at the array were averaged for 1 min, and then compressed to get the impulse response at each receiver using an ultrafast Hadamard transform method

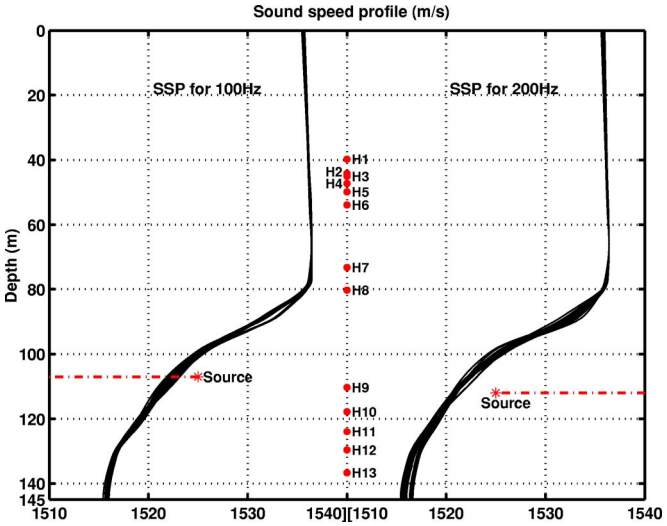


FIG. 2. (Color online) Sound speed profiles (SSPs) measured during the time periods of the data analyzed in this paper. The dots in the middle of the plot are the positions of the array hydrophones.

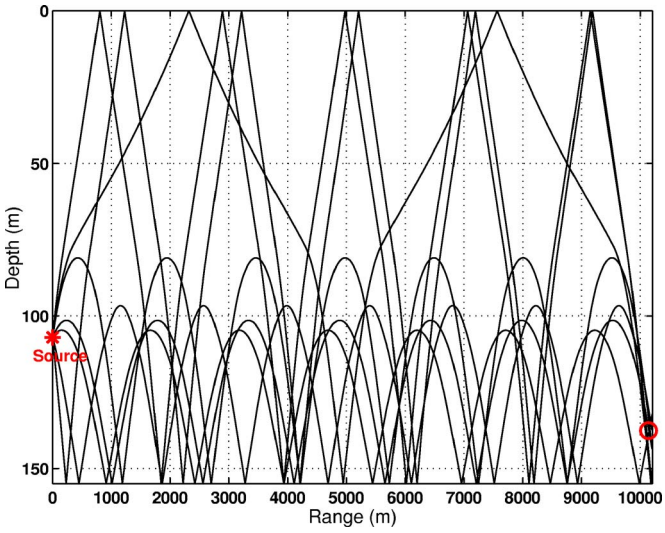


FIG. 3. (Color online) RBR and SRBR ray paths for deep source with one of the SSPs during 100-Hz signal was transmitted, as shown in the left portion of Fig. 2.

developed by Birdsall and Metzger.[21] As a result, a band-limited impulse response of the ocean channel was derived at each receiver for each minute during the source transmission period at each frequency. The band-limited impulse responses were the "raw signals" that were used in the inversion.

C. Water column sound speed profile

The water sound speed profile data were collected simultaneously from the two environmental arrays along the propagation path. Each of the environmental arrays was equipped with a temperature-pressure recorder, a conductivity-temperature recorder, and 10 temperature recorders that spanned the water column at depths ranging from 33 to 132 m in 11 m steps. Sound speed profiles were derived from 3-min temperature samples and the averaged temperature-salinity relationship, which provided one sound speed profile (SSP) from each environmental array every 3 min. The data used in this paper were selected from two consecutive 1 h periods when the temperature time series from the two environment arrays were highly correlated, so that the sound speed could be considered as range independent.

A typical sound speed profile at the site during the experiment period consisted of an iso-speed layer within the upper 40–50 m, overlying a thermocline that generated strongly downward refracting propagation at deeper depths. Figure 2 shows the sets of SSPs that were derived from the environmental array data for the two time periods; on the left side the profiles for a 100-Hz transmission period, and on the right side for the subsequent 200-Hz transmission period. As can be inferred from the figure, the sound speed profile was very stable over the entire 2-h period.

D. Acoustic propagation characteristics

The acoustic field from the source, located in the thermocline in the deeper portion of the water column, consists of two components that can be described in terms of ray theory. Steep angle rays ($>8°$) propagate by surface-reflected/bottom-reflected (SRBR) paths that span the entire water column. These rays are received at all the array sensors throughout the water column, and the relative arrival times span about 150 ms. Shallow angle rays less than about 8° are trapped in the waveguide formed by the downward refracting deep thermocline, and they arrive as a strongly focused group of refracted/bottom reflected (RBR) rays. These rays have the shortest travel times, with relative arrival times within 10–20 ms, and the intensity of the group is 10–15 dB greater than that of the SRBR components.[18] The RBR signal is received only on the deeper hydrophones of the array that are within the waveguide at the bottom of the water column. Propagation for both SRBR and RBR rays is bottom limited with 8–10 bottom interactions along the propagation path, as indicated in Fig. 3.

Figure 4 shows the structure of the measured signal that was received over a 1-h period at each frequency for the hydrophone at the top of the vertical array, ~39.8 m below the sea surface. The data span a complete cycle of the sound transmission frequencies. Two distinct groups of multipath arrivals separated by ~0.4 s are observed at the lower frequencies, i.e., 100, 200, and 400 Hz. Both groups of arrivals are strongly coherent and stable.[18] The amplitude of the second group of arrivals attenuates as frequency increases, and the arrivals are not evident in the data at 800 Hz and higher. The depth variation of the signal in the water column is displayed in Fig. 5, which shows 1-min averaged compressed signals at 100 Hz for all the array hydrophones. The strong shallow angle RBR components are seen in the deeper channels (9–13), but the prominent secondary arrival is not observed at these hydrophones. Arrival times for the first group of arrivals are consistent with predicted times for the SRBR and RBR components, assuming propagation in the plane of the source and receiver.

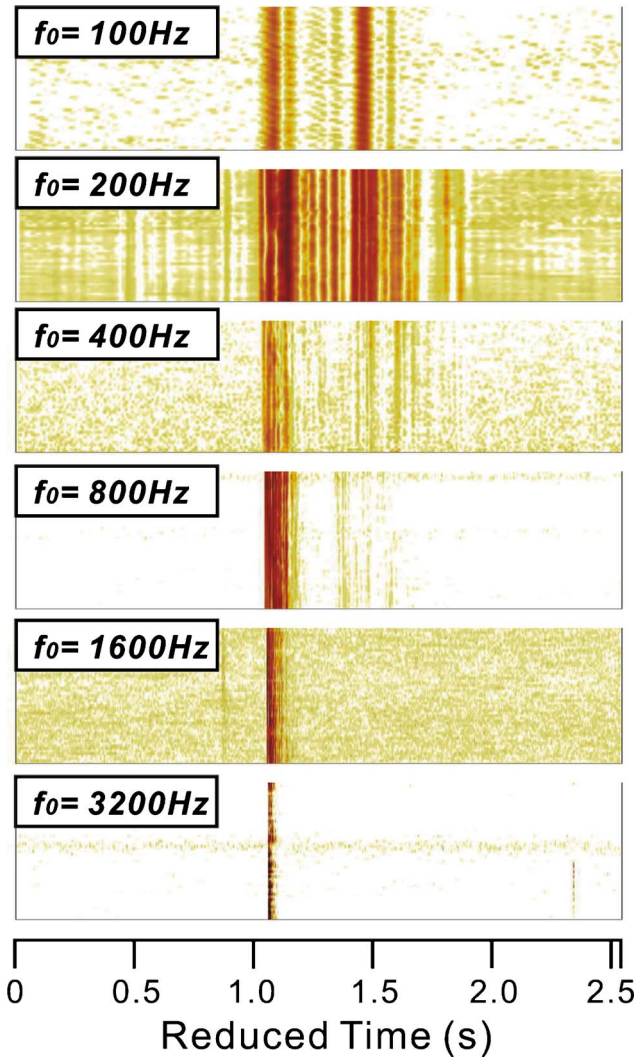


FIG. 4. (Color online) Band-limited impulse response signals at the six center frequencies, all from the hydrophone at 39.8 m below the sea surface. In each frequency panel, the *y* axis represents the ping number. 1 ping/min over the 1 h period. The *x* axis is the time scale of time series. The strong second group of arrivals are found in the 100-, 200-, and 400-Hz data sets, around 0.4 s after the first arrival.

E. Vertical directionality of the signal

In order to understand the structure of the acoustic field in the Florida Straits environment, the array data were processed using a beamforming technique applied in the time domain. Since the spacing of the sensors used in this array was uneven, conventional array shading methods could not be applied. Instead, the amplitude weighting factors for each beam were determined adaptively in a nonlinear inversion process that was designed to find the weighing vector that minimized the ratio of the highest side lobe level to the main beam level for each beam steering direction. A global optimization algorithm, differential evolution, was used to invert the weights. Steering angles were chosen so that the −3-dB points of the main lobes of adjacent beams were the same. The improvement in sidelobe suppression for the 0° beam at 100 Hz is shown in Fig. 6. The overall suppression obtained for the beams at 100 and 200 Hz was −9 and −7 dB, respectively, as shown in Fig. 7. Due to the sensor configuration, application of the beamforming process was limited to the two lowest frequencies.

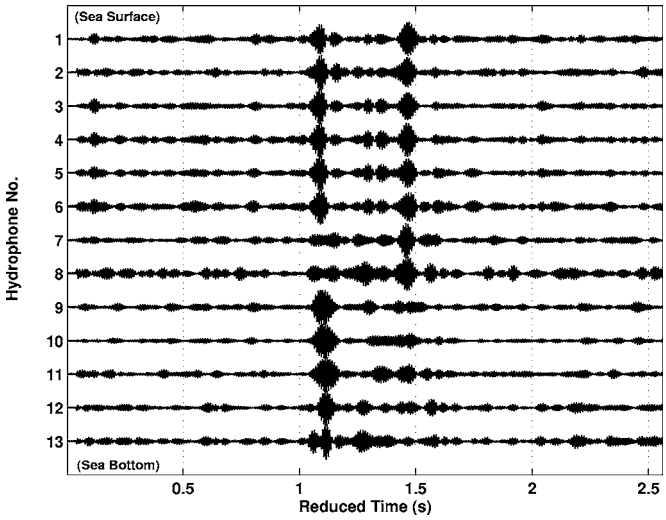


FIG. 5. Band-limited impulse response signals from the array hydrophones at 100 Hz. The hydrophone depths range from 39.8 to 136.7 m. Note that the very strong second group of arrivals are observed for hydrophones 1 to 8, all of which are above the thermocline.

Examples of the vertical arrival structure at the array are shown in the beam time series in Fig. 8 for 100 Hz (a) and 200 Hz (b). These results were typical for all the signals transmitted over the 1-h period for each frequency. It is clear from the figure that the second arrival consists of only higher angle components (>10°). Acoustic energy propagating at lower angles is only contained in the first arrival.

We present here two hypotheses to account for the strong secondary arrival. The first one assumes that the arrival is due to interactions with the subbottom; this implies that signal propagation is entirely in the source/receiver plane. However, our previous attempts in estimating a geoacoustic model based on this hypothesis were unsuccessful in

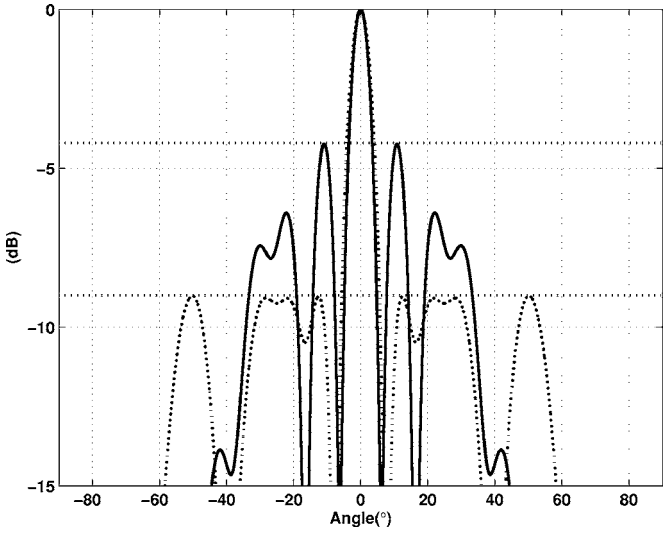


FIG. 6. Comparison of beam patterns of the unevenly spaced linear array before (solid line) and after (dashed line) weighting when steering angle is 0° at 100 Hz. Note that the largest sidelobe level is −4.2 dB before but −9.0 dB after weighting.

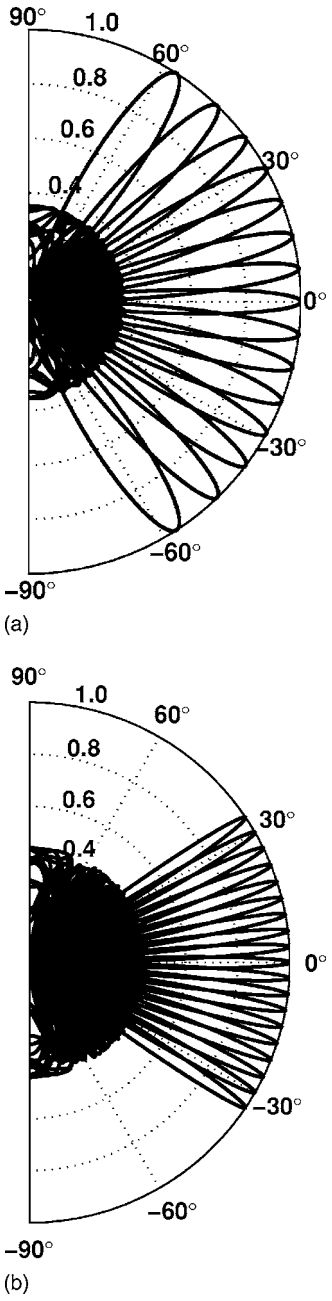


FIG. 7. Beam patterns for the unevenly spaced linear array with different steering angles at (a) 100 Hz and (b) 200 Hz, both after array shading. The beams steered to different directions overlap at the −3-dB point, the narrowest beam width (when steering angle is 0°) is 7.1° at 100 Hz and 3.5° at 200 Hz.

generating a realistic model that could account for the strong secondary arrival.[17] The second hypothesis makes the assumption that the secondary arrival is due to interactions with the complicated sloping bottom that generate horizontal refraction. This model is developed further in the next section.

III. GEOACOUSTIC INVERSION

A. Model of sound propagation

Our objective in this work is to apply the MBP method to invert the acoustic field data received at the array. The inversion is formulated as an optimization problem designed

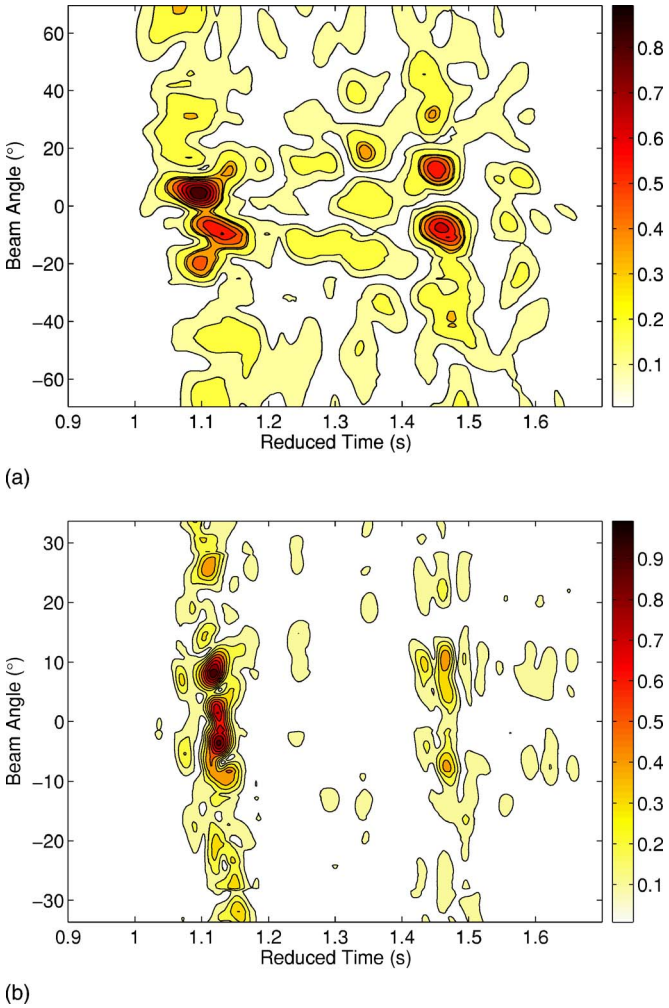


FIG. 8. (Color online) Vertical directionality of the signal at the array. (a) for 100 Hz; (b) for 200 Hz.

to obtain optimal seabed parameter estimates by searching the multidimensional geoacoustic model parameter space. As mentioned previously, the long range, cross slope geometry of this experiment presents significant challenges for an inversion based on conventional MFP. We make some remarks here to clarify the interpretation of the inversion results.

First, signals propagating in the cross-slope geometry are refracted horizontally due to the sloping bottom environment. For a uniform slope, the deviation of the propagation path from the source/receiver plane depends on the magnitude of the bottom slope and the vertical propagation angle of each propagating mode. The slope at the source/receiver contour in this experiment is less than one degree. However, the slope is not uniform shoreward of the experimental site, and increases significantly to around 3 degrees a few kilometers shoreward of the array (Fig. 1). In these conditions, accurate modeling of the signal at the array should account for effects of three-dimensional propagation.

Second, the information about the bottom contained in the data measured at the array represents the integrated effect of the interaction with the ocean bottom along the entire propagation path. The specific variability of the sea bottom materials along the path is not known in detail. However, sediment cores in the vicinity of the site indicate a consistent

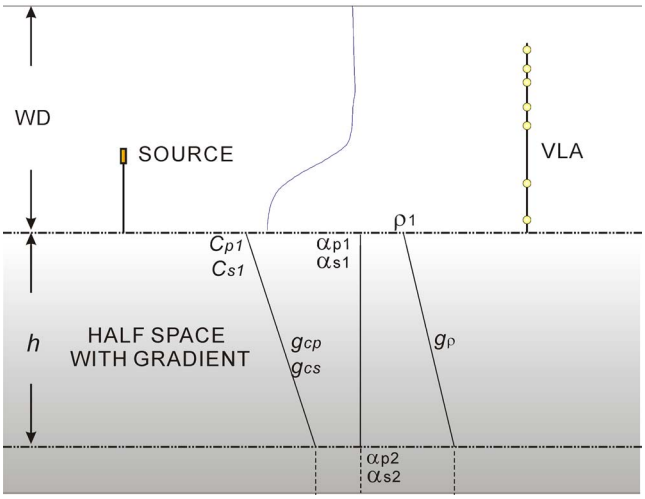


FIG. 9. (Color online) Geoacoustic model used in this paper's inversion study.

lithology of fine-grained sand overlying more consolidated mudstone or packstone within the first 2 or 3 m of the sea floor. Consequently, we assume a simple range-independent geoacoustic model for the site, and interpret the inversion results in terms of averaged model parameters over the propagation path.

Our model for sound propagation assumes that the first arrivals are propagated in the plane of the source and receiver. Since the gradient of the continental slope at the source is very small ($<1°$), this is likely a good assumption, particularly for the very small angle RBR modes. We also assume that the steep gradient portion of the slope shoreward of the array acts like a "wall" to refract signal components of shoreward radiated sound back to deeper waters, by the interaction with the steep slope. Qualitatively this model allows a secondary arrival, following the signal that is propagated in the plane of the source and the receiver, and is consistent with the vertical directionality observed in the data.

B. Inversion strategy

In view of these considerations, our strategy for the inversion follows the concept of matched beam processing. We apply the time delay beamforming algorithm to spatially filter the array data, and select only the low-angle beams less than 8°. This operation removes the secondary arrival. Our assumption is that the remaining low-angle signal components can be modeled using a two-dimensional propagation model.

The normal mode propagation model ORCA[22] was used to calculate the replica fields at the array for the range-independent geoacoustic model shown in Fig. 9. The model consisted of a thick gradient layer over a half space with parameter values equal to those at the bottom of the gradient layer. The parameters to be inverted were: p-wave speed, c_p, and gradient, g_{cp}; p-wave attenuation at top of the sediment α_{p1}; p-wave attenuation at the interface of the sediment and lower half space α_{p2}; density at the top of the sediment ρ; and density gradient g_ρ. In addition, geometrical parameters such as water depth, WD, range, R, and source depth, SD, were also included as parameters to be inverted. Since water depth was included as one of the parameters, related param-

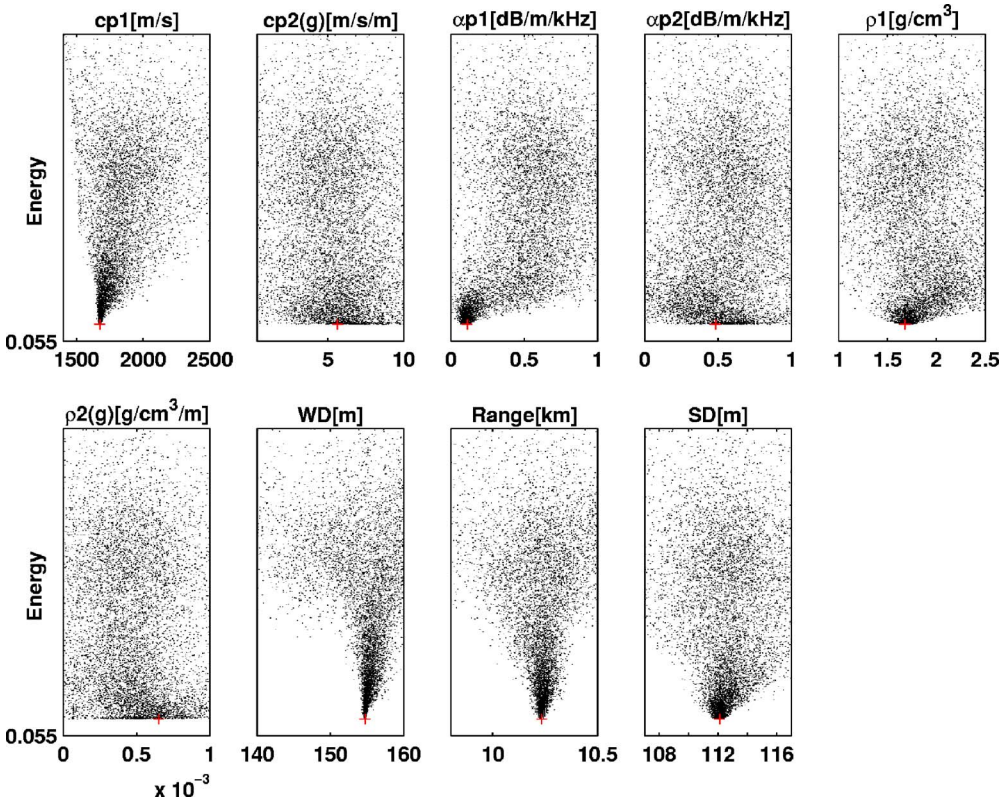


FIG. 10. (Color online) Scatter plots of the energy as a function of parameters to be inverted at 200 Hz.

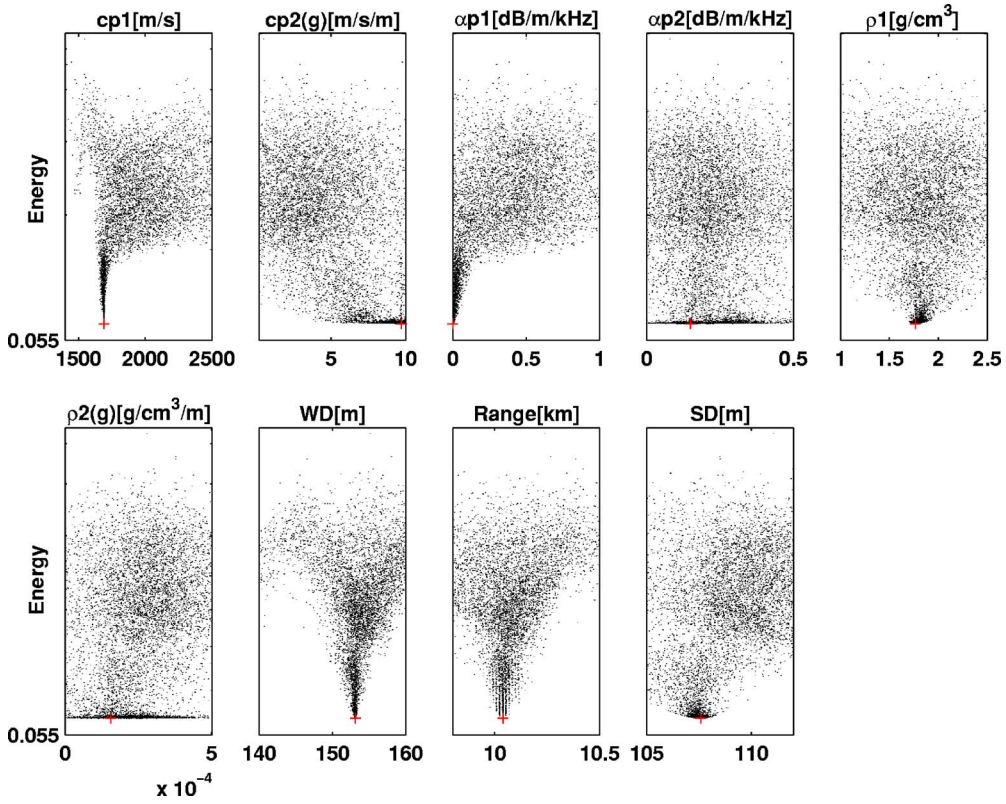


FIG. 11. (Color online) Scatter plots of the energy as a function of parameters to be inverted at 100 Hz.

eters such as sound speed in the water at the sea bottom and the hydrophone depths of the bottom moored array were adjusted accordingly. In order to explore shear wave effects from the data set, shear wave velocity c_s, c_s gradient g_{cs}, shear wave attenuation at top α_{s1} and shear wave attenuation at bottom α_{s2} were included later in the inversion study as well. In this case, there were ten geoacoustic parameters and three geometric parameters to be inverted.

The cost function of matched beam inversion used in this approach was one minus the average of the cross correlation of the envelopes of the measured and modeled beams:

$$E(m) = 1 - \frac{1}{N_b}\left[\sum_{i=1}^{N_b}\left(\frac{1}{T}\int_0^T B_i(m;t)D_i(t+\tau)dt\right)_{\tau=\tau 0}\right],$$

where m is the parameter model vector, N_b is the number of beams used in the inversion, T is the pulse length, $B(m;t)$ is the envelope of the modeled beam time series, $D(t)$ is the envelope of the measured beam signal, and τ is time delay. $B(m;t)$ and $D(t)$ are normalized, respectively, so that the autocorrelation is equal to 1 when τ equals 0.

The optimization method was adaptive simplex simulated annealing (ASSA[4]), a hybrid optimization algorithm that combines a global search method, simulated annealing (SA), and a local gradient search algorithm, downhill simplex. The annealing schedule was determined by several trial runs with different annealing schedules prior to the inversion. A typical annealing schedule for the 100-Hz data inversion was T_0=0.3 and β around 0.995, where T_0 is the starting temperature and β is a constant which controls the cooling speed.

C. Inversion result

Inversions were carried out for 20 of the 58 signal transmissions in the hour periods for each frequency, i.e., an inversion for each of the 20 SSPs during the hour. Results of one of the ASSA inversions for 200 Hz are shown in the scatter plots in Fig. 10. These inversions were done using the real axis option in ORCA that uses a perturbation approach to approximate the attenuation. The most sensitive parameters, as indicated by the narrow "tornado-like" distributions in one-dimensional slices in the model parameter space, are sea floor sediment sound speed, the water depth, range, and source depth. Sea floor sediment attenuation and density are less sensitive, and all other parameters are insensitive. The final energy was 0.055, indicating over 94% match with the measured beam time series. Inversions carried out to estimate shear wave parameters used the complex axis option in ORCA, and required substantially longer computing times

TABLE I. Summary of "true" values, parameter means, and standard deviations of the Inversion results of 100- and 200-Hz data set.

Parameter	"True"value	100 Hz	200 Hz
c_p (m/s)	1676	1683.3±10.5	1686.8±11.0
c_s (m/s)	···	250.1±73.1	
α_p (dB/m/kHz)	···	0.0013±0.0045	0.123±0.054
ρ (g/cm^3)	1.779	1.795±0.327	1.520±0.133
WD (m)	155	153.7±1.3	153.5±0.7
R (km)	10.141	10.086±0.054	10.060±0.094
SD (m)	107/112	107.1±2.3	109.9±1.5

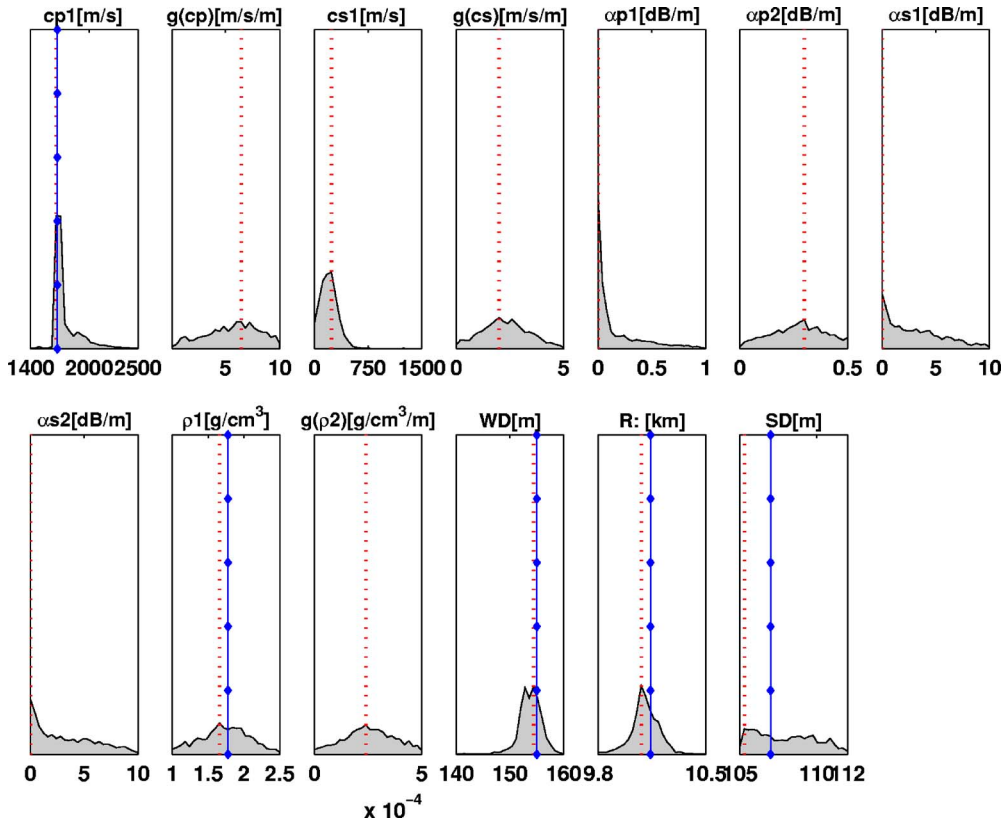


FIG. 12. (Color online) Accumulated distribution of 20 inversion results of 100-Hz data set from Trans. No.5. Dashed lines indicate where the models are most concentrated, and diamond lines indicate the "true" values.

(about five times longer). However, estimated values of the sensitive parameters were not significantly changed.

A similar example of the inversions at 100 Hz is shown in Fig. 11. The same hierarchy of sensitivity is observed. The estimated values are summarized in Table I. The quoted values represent the means and standard deviations of the averaged values for all the ASSA inversions at each frequency. Another display of the averaged result presented in Fig. 12 shows the cumulative stacks of all the inversions at 100 -Hz for selected geoacoustic parameters. Although these plots are qualitatively similar to the displays of *a posteriori* probability distributions from Bayesian inversions, they do not have precise statistical significance. However, they provide an impression of the variability of the estimates over the 1-h period, and an indication of the sensitivity. The inversion is sensitive to the sea floor parameters, but there is not reliable information in the data about the subbottom.

The sensitivity of the *s*-wave speed at the sea floor appears to be greater than the sensitivity of sediment density. The estimated value of ~250 m/s is consistent with the *p*-wave speed, but the result is not as convincing because of its high standard deviation, ±73.1 m/s.

D. Discussion

The matched beam inversions were very successful in estimating accurate values of the water depth, range, and source depth of both the high and low frequency sources in our study. The estimated values are in excellent agreement with independent measurements of the experimental arrangement that were made during the experiment, as shown in Table I. Notably, the correct value of source depth for the lower frequency source is obtained (about 5 m shallower than the high frequency source depth). Although the range is consistently overestimated by 60–80 m, the error is within reasonable limits for the true GPS value. The GPS data provide the ship to ship distance from the positions at deployment of the equipments; the actual positions of the source and receiver on the sea bed are not precisely known.

The most sensitive geoacoustic parameters in this inversion are those for the sea floor sediments, and the estimated values are consistent with the results from 2 to 3 m piston cores sampled in nearby sediments. Cores[23] near the propagation path between the source and VLA were composed of a very thin layer of mudstone (clay/silt) at the top, then a layer of grainstone (primarily very fine to fine sand), and Packstone in the lower layer. Over the top 2–3 m of the sediment, the density varies from 1.713 to 2.408 g/cm^3; and the *p*-wave speed measured at 10 MPa ranges from 1676 to 1924 m/s. The closest core to the transmission path is sample 5, as shown in the right panel of Fig. 1. It is only 310.6 m away from the source-VLA plane and thus was chosen to be the reference ground truth data. The density at a depth of 9 cm in the core was 1.779 g/cm^3, and *p*-wave speed at a depth of 22 cm in the core was 1676 m/s. For comparison, those two values, along with water depth, source-receiver range, and source depth, are plotted as diamond lines in Fig. 12.

The estimated values of attenuation of the sea floor sedi-

ment for the two frequencies are paradoxical. Regardless of the fact that p-wave attenuation was also a sensitive parameter in the inversion, the estimates are not consistent between the inversion results from the 100 and 200-Hz data sets. The estimate of p-wave attenuation from 100-Hz data is as low as 0.0013 dB/m at 1 kHz, whereas the estimate from 200 Hz is 0.123 dB/m at 1 kHz. (The lower search bound of this parameter for both cases is zero.) Although the result for 100 Hz is consistent with the recent empirical model of Zhou[24] for the nonlinear relationship between frequency and attenuation, the estimated value at 200 Hz is not. However, Zhou's model is based on sediment material for which the ratio of sediment to bottom water sound speed is 1.01–1.06, whereas the ratio for the sediments at the Florida Straits site was significantly greater, ~1.10. So, it is not clear that the model in its present form is relevant to our experiment.

Although the shallow angle beams are well matched in the inversion, we do not expect that the estimated geoacoustic model will accurately predict the strong secondary arrival observed in the signal at a single hydrophone. We have assumed that the secondary arrival is not related to interaction with deeper subbottom layers, but instead due to three-dimensional propagation. In this model, the source transmits omni-directionally, and the sound transmitted toward the coast interacts with the sloping bottom shoreward and returns to deeper water. Although the effect requires a three-dimensional propagation model for accurate description, this effort is beyond the scope of this paper. Rather, we demonstrate here qualitatively that the hypothesis is reasonable, using the Gaussian beam model BELLHOP[25] to simulate the time difference between the two paths. We assume that the first path propagates in the source-receiver plane, and the travel time for the first group of arrivals is t_1 seconds; the second path starts from the source but with an angle shoreward to the west, and is reflected by the slope back to the VLA. If the travel time for this path is t_2, then the point that gives $t_2-t_1=0.39$ s would be A(26°5.85N,80°4.10W), which is on the slope at the depth between 105- and 110-m contour, as shown in Fig. 1.

To demonstrate the composition of the signal, we show in Fig. 13 the comparison between the measured signal at a single sensor compared to the signal modeled by ORCA using the estimated geoacoustic model. The figure displays the composition of the wave form for the topmost array hydrophone (at the depth of 39.8 m underwater) in terms of the first 11 modes, which represent the propagating modes at 100 Hz. The geoacoustic and geometric parameters used in this impulse response prediction are the results from the 100-Hz inversion. The first line in Fig. 13 is the measured impulse response, showing the strong second group of arrivals. The wave forms under the simulated signal are the contributions of the lower order modes with the most distinguishable amplitudes. The lowest order RBR mode is not received at this sensor since the RBR mode propagates in the deep sound channel at the bottom of the waveguide and does not reach the hydrophones at shallower depth. As expected from an assumption of two-dimensional propagation, the propagating modes contribute only to the first arrival. Although the very low attenuation and relatively high sediment

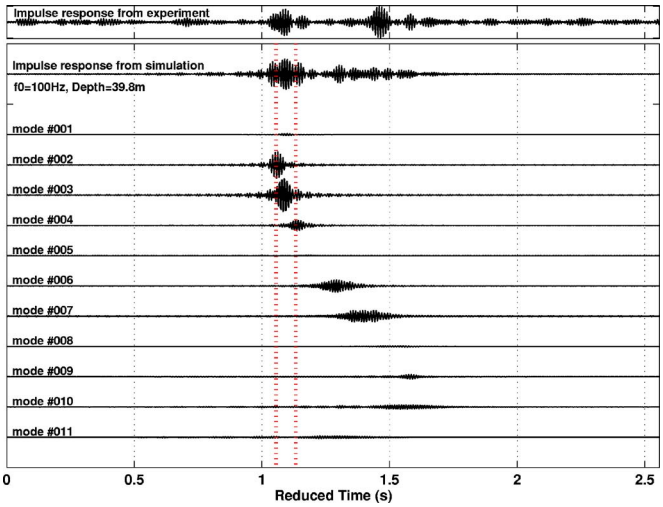


FIG. 13. (Color online) Comparison of 100-Hz real (first line in the plot) and predicted (second line in the plot) band-limited impulse response at depth of 39.8 m. The geoacoustic parameters used in the simulation are the mean values of inversion results from 100-Hz data. The wave forms under the predicted signal are the contribution of the modes with distinguishable amplitudes.

sound speed do give rise to later arrivals, they are very weak and do not accurately model the strong secondary arrival seen in the data.

The results of the inversion indicate that there is very little information in the data about deeper layers in the subbottom. The very weak sensitivity to subbottom model parameters below the sea floor is easily understood in terms of the experimental geometry, and the approach used in the inversion. The inversion makes use of only very low angle components of the acoustic field at the array, $<6-8°$. Since these angles are much less than the sea floor critical angle, the acoustic field is evanescent below the sea floor, and is not effective for sampling the deeper subbottom layers. The most reasonable geoacoustic parametrization in this scenario is a simple model that focuses on the sea floor parameters, and not on the fine structure of the sediment column. Although the sensitivity to the sea floor sediment parameters is high, the estimated values must be interpreted in terms of averages over the unknown variations along the propagation path.

IV. SUMMARY

A time domain matched beam geoacoustic inversion for an unevenly spaced linear array is described in this paper. Array weighting factors were inverted by a differential evolution algorithm. Time domain beamforming technique was used in the inversion procedure to eliminate a strong second group of arrivals that we assume are related to horizontal refraction by a sloping bottom in the local environment. The inversion used a cost function based on the correlation of the measured and modeled envelopes of the wave forms at selected directions after beamforming the array data.

The method was applied to low frequency broadband data that were collected in a bistatic long range propagation experiment at a site near the South Florida Ocean Measurement Centre. The inversions for the 100- and 200-Hz data provide estimates of the sea floor geoacoustic parameters that

agree well with the ground truth information. Also, the estimated values of the geometric parameters were consistent with the geometric setting and the knowledge of the water depth of this experimental site.

ACKNOWLEDGMENTS

This work is supported by Office of Naval Research. The authors would like to thank Dr. N. J. Williams from South Florida Ocean Measurement Center (SFOMC), for providing assistance in understanding the environmental data from the experiment.

[1]M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean bottom properties," J. Acoust. Soc. Am. **92**, 2770–2883 (1992).
[2]N. R. Chapman and C. E. Lindsay, "Matched field inversion for geoacoustic model parameters in shallow water," IEEE J. Ocean. Eng. **21**, 347–355 (1996).
[3]M. Musil, M. J. Wilmut, and N. R. Chapman, "A hybrid simplex genetic algorithm for estimating geoacoustic parameters using matched-field inversion," IEEE J. Ocean. Eng. **24**, 358–369 (1999).
[4]S. E. Dosso, M. J. Wilmut, and A. S. Lapinski, "An adaptive-hybrid algorithm for geoacoustic inversion," IEEE J. Ocean. Eng. **26**, 324–336 (2001).
[5]P. Gerstoft, "Inversion of seismo-acoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[6]P. Gerstoft and Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[7]L. Jaschke and N. R. Chapman, "Matched field inversion of broadband data using the freeze bath method," J. Acoust. Soc. Am. **106**, 1838–1851 (1999).
[8]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[9]S. E. Dosso and P. L. Nielsen, "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow water data," J. Acoust. Soc. Am. **111**, 143–159 (2002).
[10]R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," J. Acoust. Soc. Am. **117**, 626–637 (2005).
[11]Z.-H. Michalopoulou, "Matched-impulse-response processing for shallow-water localization and geoacoustic inversion," J. Acoust. Soc. Am. **108**, 2082–2090 (2000).
[12]Z.-H. Michalopoulou and M. Picarelli, "Gibbs sampling for time-delay and amplitude estimation in underwater acoustics," J. Acoust. Soc. Am. **117**, 799–808 (2005).
[13]A. Tolstoy, N. R. Chapman, and G. E. Brooke, "Workshop' 97: Benchmarking for geoacoustic inversion in shallow water," J. Comput. Acoust. **6**, 1–28 (1998).
[14]N. R. Chapman, S. Chin-Bing, D. King, and R. B. Evans, "Benchmarking geoacoustic inversion methods for range-dependent waveguides," IEEE J. Ocean. Eng. **28**, 320–330 (2003).
[15]H. B. Nguyen, H. A. DeFerrari, and N. J. Williams, "Ocean acoustic sensor installation at the South Florida ocean Measurement Center," IEEE J. Ocean. Eng. **27**, 235–244 (2002).
[16]H. A. DeFerrari, N. J. Williams, and H. B. Nguyen, "Variability, coherence and predictability of shallow water acoustic propagation in the straits of Florida," in *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer Academic, Dordrecht, 2000), pp. 245–254.
[17]N. R. Chapman and Y. Jiang, "Geoacoustic inversion of broadband data from the Florida Straits," in *High Frequency Ocean Acoustics*, edited by M. Porter, M. Siderius, and W. A. Kuperman (American Institute of Physics, New York, 2004), pp. 40–46.
[18]T. C. Yang and T. Yates, "Matched-beam processing: Application to a horizontal line array in shallow water," J. Acoust. Soc. Am. **104**, Pt. 1 1316–1330 (1998).
[19]T. C. Yang and T. Yates, "Matched-beam processing: Range tracking with vertical arrays in mismatched environments," J. Acoust. Soc. Am. **104**, 2174–2188 (1998).
[20]http://www.ngdc.noaa.gov/mgg/gdas/gd_designagrid.html.
[21]T. G. Birdsall and K. Metzger Jr., "Factor inverse matched filtering," J. Acoust. Soc. Am. **79**(1), 91–99 (1986).
[22]E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[23]D. F. McNeill, "Sedimentology, physical properties, and acoustic response of sediment from piston cores at the South Florida Test Facility," Data Report, Comparative Sedimentology Laboratory, Division of Marine Geology and Geophysics, Rosenstiel School of Marine and Atmospheric Science, University of Miami, April. 2001.
[24]J.-X. Zhou and X.-Z. Zhang, "Nonlinear frequency dependence of the effective acoustic attenuation from low-frequency field measurement in shallow water," J. Acoust. Soc. Am. **117**(4), 2494 (2005).
[25]M. B. Porter and Y.-C. Liu, "Finite element ray tracing," in *Theoretical and Computational Acoustics*, edited by D. Lee and M. H. Schultz (World Scientific, Singapore, 1994), Vol. **2**, pp. 947–953.

Temporal and cross-range coherence of sound traveling through shallow-water nonlinear internal wave packets

Timothy F. Duda
Applied Ocean Physics and Engineering Dept., MS 11, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 8 September 2005; revised 5 April 2006; accepted 6 April 2006)

Expressions governing coherence scales of sound passing through a moving packet of nonlinear internal waves in a continental shelf environment are presented. The expressions describe the temporal coherence scale at a point, and the horizontal coherence scale in a plane transverse to the acoustic path, respectively. Factors in the expressions are the wave packet propagation speed, wave packet propagation direction, the fractional distance from the packet to the source, and the spatial scale S of packet displacement required to cause acoustic field decorrelation. The scale S is determined by the details of coupled mode propagation within the packet and the waveguide. Here, S is evaluated as a function of frequency for one environment, providing numerical values for the coherence scales of this environment. Coherence scales derived from numerical simulation of coupled mode acoustic propagation through moving wave packets substantiate the expressions. [DOI: 10.1121/1.2200699]

PACS number(s): 43.30.Re, 43.30.Bp, 43.30.Ft [AIT] Pages: 3717–3725

I. INTRODUCTION

Recent theoretical, computational, and observational papers have demonstrated that nonlinear internal waves are responsible for fluctuations of sound (of a few hundred hertz) propagating a few tens of kilometers on continental shelves. Research into this problem accelerated after demonstration that anomalous transmission loss between 300 and 1100 Hz, observed in the Yellow Sea, was consistent with coupled-mode propagation through packets of nonlinear internal waves (Zhou, Zhang, and Rogers). That work raised the possibility of a resonance effect between the internal wave wavenumber and the difference wavenumber of interacting modes.

A subsequent paper showed that mode coupling within individual solitary waves could be described by distinct coupling events at each of the two sides of the wave (Preisig and Duda, 1997). That paper also showed that the vertical mode shape overlap at depths having strong lateral gradients of sound speed played a strong role in mode coupling, as expected, and demonstrated that waves with small lateral scale produce little coupling, regardless of their amplitude, because of canceling coupling events. A resonance effect between the internal wave horizontal wavenumber and the difference horizontal wavenumber of the incident (coupled-from) and transmitted (coupled-to) modes was not found to be responsible for strong mode coupling in the particular cases of waveguide geometry and frequency that were chosen for study. Instead, resonance and phasing of dominant modes propagating in the wave center (the trough for a nonlinear wave of depression) were found to be of central importance. This work shows that resonance between the internal wave scale and the mode interference scale, as discussed in Frank *et al.* (2004) and papers cited therein, must play a role, but that the mode pair or pairs responsible for the largest effects may be difficult to ascertain.

Further work showed that the mode coupling process is sensitive to the precise location of internal waves or wave packets (Duda and Preisig, 1999). This sensitivity to position, coupled with the propagating nature of ocean internal waves, creates a time dependence in the resultant acoustic fields. The effect is caused by the changing interrelations of model phases at the point where sound encounters a moving wave. Intuitively, this leads to the temporal decorrelation of signals transmitted from a fixed source to a fixed receiver that pass through moving internal waves (Duda and Preisig, 1999; Rouseff *et al.*, 2002). To apply a nomenclature to this process, let S (with $S>0$) be the internal wave (or packet) horizontal displacement scale along the acoustic propagation path that leads to a typical intensity or phase fluctuation at the receiver. That is, movement of the wave of distance S causes an acoustic fluctuation of a typical magnitude.

In the field, measurements of intensity time series and of temporal correlation functions, collected using fixed acoustic propagation paths containing moving internal waves, exhibit fluctuating behavior. Particular measured examples are the rapid swings in signal energy arriving at vertical line arrays over intervals of only a few minutes (Duda *et al.*, 2004; Duda and Preisig, 1999; Lynch *et al.*, 2003; Chiu *et al.*, 2004). Here, by the term signal energy we mean depth-integrated intensity, which is used in other papers on this topic (Duda and Preisig, 1999; Oba and Finette, 2002). Many aspects of the fluctuations essentially validate that wave motion causes the fluctuations (i.e., that moving waves are responsible for fluctuations and that the scale S exists). It is possible that S is not constant at any single location, changing, for example, over a fortnightly tidal cycle as nonlinear wave heights change. However, if the change is slow, then it is reasonable to model the process with a quasisteady scale S.

To further examine aspects of wave-induced acoustic fluctuations, two expressions are derived in this paper for the idealized situation of propagation through a single packet of

internal waves with long straight crests to a horizontal line of receivers. The first expression is for the temporal coherence scale T_c of the received signal as a function of three parameters: S, the wave (packet) speed, and the wave (packet) direction. This is little more than a definition, and stems directly from the existence of the scale S. The other is an expression for the horizontal signal coherence scale Y_c in a plane normal to the acoustic path, which is written as a function of four parameters: S, source receiver distance R, source to packet distance R_P, and the wave packet angle with respect to the acoustic path. As in previous work (Duda, 2004; Duda and Preisig, 1999; Preisig and Duda, 1997), sound propagating distances of order $o(30)$ km in a waveguide with $o(100)$ m water depth is considered.

The ultimate usefulness of the expressions may lie in the possibility of predicting one of the three coherence scales S, T_c, and Y_c after measuring another of them while simultaneously knowing properties of the internal waves. For example, signal energy temporal coherence scale measurements at a vertical line array, when combined with internal wave angle and phase speed measurements (as from satellite synthetic aperture radar), may yield the horizontal cross-range coherence scale.

Results of the expressions are compared with coherence functions computed using N-by-two-dimensional ($N\times 2$-D) simulations of three-dimensional acoustic propagation, where N is the number of azimuthal planes passing through the source for which outgoing sound is modeled using a parabolic wave equation. $N\times 2$-D simulations are sufficient for the situation of negligible azimuthal scattering, which has been found to be the case for sound traveling at angles from 0° to 70° with respect to the propagation direction of long-crested internal waves (Oba and Finette, 2002). Thus, we address 70° of the possible 90° of the angle between the wave crests and the acoustic path. This is the regime where mode coupling dominates over mode refraction. Three-dimensional acoustic effects, such as mode refraction, may occur at the remaining 20° and are covered in other papers (Badiey *et al.*, 2005).

Other water-column processes and seafloor structures may also cause spatial and temporal decorrelation. When this occurs, the results shown here for decorrelation caused by a single packet of waves between the source and receiver may represent upper bounds on correlation distances and times. In situations where a single packet is the primary cause of acoustic fluctuation, the results might allow a useful prediction of correlation scales.

The paper is organized as follows. The geometry of the problem is defined and the coherence scale expressions are developed in Sec. II. In Sec. III, coherence scale results from the expressions are compared with computational simulation results. In Sec. IV, the scale S, an important factor governing coherence scales T_c and Y_c, is evaluated as a function of frequency. In Sec. V we discuss nondimensional cross-range coherence scale. Results are summarized in Sec. VI.

II. COHERENCE SCALE EXPRESSIONS

In this section, the geometry of the problem is defined and geometric arguments are used to derive the relationships between T_c and S and between Y_c and S, respectively. A diagram of source, receiver, and wave packet positions is shown in Fig. 1(a). The x direction is defined to be parallel to the source-receiver path. The source to receiver distance is denoted R. The source to packet distance along the source-receiver path is denoted R_P. The angle between the wave packet propagation direction (aligned with the velocity vector $\mathbf{C_P}$) and the acoustic path is η, which has complementary angle ϕ. Thus, ϕ is the angle between the wave crests (or single wave crest, wave trough, or wave troughs) and the acoustic propagation path. The cross-range tangent vertical plane, which is normal to the acoustic path, and within which we will examine cross-range coherence, is denoted **P**. Horizontal lines within **P** are parallel to the unit vector in the y direction.

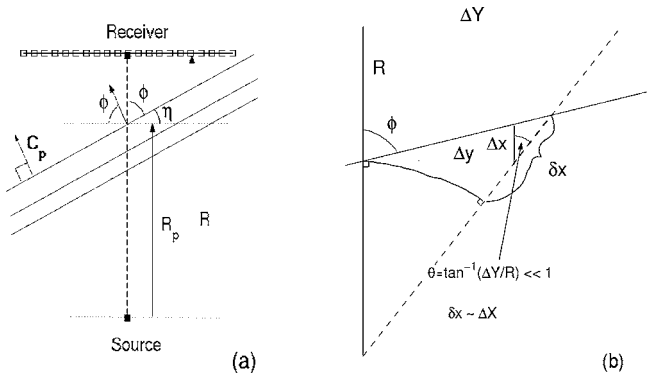


FIG. 1. (a) Definition sketch for wave/acoustic angle ϕ, source-receiver range R, source-wave packet range R_P, and cross-range tangent receiver plane. The angle η that is complementary to ϕ is also shown. The usual right-handed three-dimensional coordinates are used, with the x direction pointing upward parallel with the arrow indicating R, the y direction pointing to the left, and the z direction out of the page. Angle ϕ ranges from −90 to 90, with a positive example shown. The cross-range tangent plane **P** used for coherence studies rises out of the page from the receiver line (open squares), with position $y=0$ at the filled square at distance R from the source, (b) The right-hand side of (a) is reproduced. Line segments are defined for sound traveling along the dashed line from the source at the bottom of the triangle to a receiver in the tangent plane at cross-range distance ΔY at the upper right. The left-hand edge of the triangle (line segment R) connects the source with a receiver at $\Delta Y=0$. The line segments ΔY in the tangent plane, Δy, Δx, and δx are shown, along with angle ϕ. These line segments have lengths equal to their names, i.e.,R has length R. An internal wave trough at angle ϕ with respect to R is drawn. If $\Delta Y/R$ is sufficiently small, δx is approximately equal to Δx, which is easy to calculate.

The temporal coherence scale T_c follows directly from knowledge of S for the waveguide and internal wave forms under consideration, and of the characteristics of how those waves propagate. Allowing for the translation of the wave or packet at phase velocity C_{px} in the direction of acoustic propagation gives $T_c=|S/C_{px}|$. The phase velocity in the acoustic propagation direction can easily be computed for waves moving in any direction. If the waves are traveling at phase speed C_p in the direction η relative to the acoustics, as shown in Fig. 1(a), then $C_{px}=C_p/\cos\eta$. Substitution gives $T_c=|S\cos\eta/C_p|$. An equivalent expression is

$$T_c=|S\sin\phi/C_p|. \tag{1}$$

Figure 2 shows a contour plot of T_c with $C_p=0.8$ m/s. For given S and C_p, note that the longest time scales are asso-

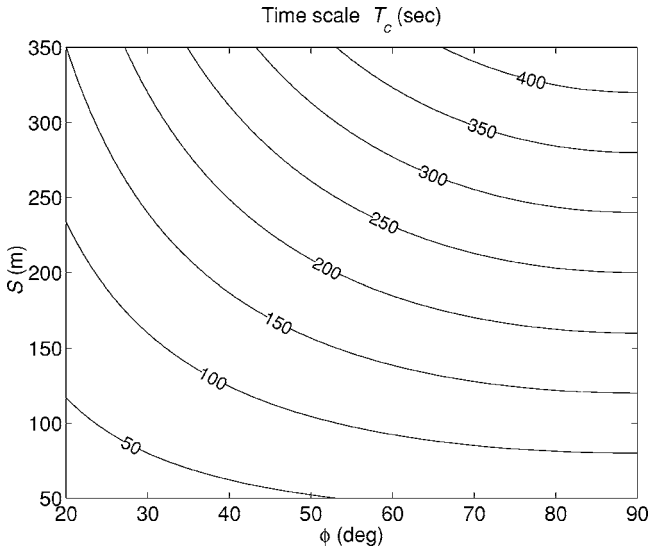


FIG. 2. Time scale $T_c(\phi,S)$ is contoured as a function of S in meters and ϕ in degrees for $\mathbf{C_p}=0.8$ m/s.

ciated with ϕ equal to 90°. This is because C_{px}/C_p has its minimum of unity at that angle. However, the use of a fixed value of S for waves of differing ϕ (but having an identical form) must be recognized as an implicit assumption or as an approximation. This is because the coupling process of sound in the x direction can be altered by the effective stretching of the wave scale for ϕ not equal to 90° (Oba and Finette, 2002; Preisig and Duda, 1997). Nonetheless, (1) describes the coherence time as a function of S, wave angle, and wave speed, despite the selection of a single C_p for the display. An alternative display of T_c/S contours with respect to C_p and ϕ would show other aspects of the behavior.

The T_c values given in Fig. 2 compare well with coherence times of 100 Hz bandwidth pulses measured during the active wave period of the South China Sea component of the Asian Seas International Acoustics Experiment. T_c values for two propagation paths, 32 km upslope and 21 km alongshelf, have been published (Chiu *et al.*, 2004; Duda *et al.*, 2004). These were computed from correlation functions of depth-averaged intensity using a threshold of 0.707. Published numbers range from approximately 40 to 220 s. The numbers can be recomputed using a decorrelation threshold of 0.5 in order to be consistent with the remainder of this paper. During the active wave period of 6–13 May 2001, the east (alongshelf) path coherence times, computed for one-day long averaging intervals, were 60 to 100 s, with most values near 70 s. For a wave along this path, good estimates of C_p and ϕ are 1.4 m/s and 55°, respectively. Using these values, S values obtained using (1) range from 100 to 180 m, with most values near 120 m. These measurements are entirely consistent with the assumptions and results of this paper. Additional coherence time estimates from this experiment ranging from 1 to 15 min, computed using energy in a narrow frequency band, appear in Mignerey and Orr (2004).

Next, an expression for the coherence scale Y_c along a cross-range line is derived. The expression pertains to the situation of a reception distant from a source such that $S/R \ll 1$, and such that R is greater than the cross-wave (aligned with $\mathbf{C_P}$) packet scale length. Adding an assumption that the lateral cross-range separations ΔY to be considered adhere to $\Delta Y/R \ll 1$, then ΔY on a line tangent to a circle at radius R can be approximated by $R\,\Delta\psi$, where ψ is azimuth and $\Delta\psi$ is azimuthal separation. (Such a line is contained in plane **P**.) Because the azimuth along which sound travels is a function of ΔY, the loss of sound coherence at increasing ΔY, with respect to sound at $\Delta Y=0$, where the tangent line touches the circle, is caused by sound at differing ΔY encountering the packet at differing R_P. If these differing R_P values surpass the value of S for sufficiently high ΔY, then coherence is small.

Y_C depends on the normalized range of the wave packet, $R_N=R_P/R$, which is between zero and one. This is evident by observing that the lateral separation at range R_P of sound encountering packets at azimuthal separation $\Delta\psi$ is $\Delta y = R_N\,\Delta Y = R_N R\,\Delta\psi$, so that near the limit of $R_N=0$ all sound reaching plane **P** at $\Delta Y < R$ will pass through a very restricted region of the packet.

Now, notice that Δy is associated with a perturbation of the packet distance from the source (measured with respect to the packet distance to source for $\Delta Y=0$) given by $\delta x \sim \Delta x = \Delta y \tan\eta$. An assumption used here is that the angle $\theta=\tan^{-1}(\Delta Y/R)$ is small, which is fully consistent with the assumption of small cross-range apertures with respect to R [Fig. 1(b)]. Now, to find the cross-range correlation length scale ΔY_c, equate the decoherence displacement scale S to δx. This converts $\delta x = R_N\,\Delta Y \tan\eta$ into $S=R_N Y_c \tan\eta$. Thus, $Y_c=S(R/R_P)/\tan\eta$, which can be written as

$$Y_c = S(R/R_P)\tan\phi. \tag{2}$$

For the chosen situation of $\Delta Y/R \ll 1$, the expression can be used to describe either an azimuthal (along an arc of equal distance from the source) or a cross-range (linear) correlation. Y_c/S is related in a simple way to R_N and $\tan\phi$.

Figure 3 shows the length scale Y_c normalized against S and contoured as a function of R/R_P and ϕ. Y_c is one to four times S for roughly half the diagram. It is very large for ϕ near 90°, exceeding the final contour indicating $Y_c=40S$. This is because the range at which sound encounters the packet, and at which mode coupling commences, is a weak function of sound propagation azimuth. At lower ϕ, the encounter range is a strong function of azimuth, causing sound to decorrelate at relatively small ΔY, and giving a shorter correlation scale.

III. COMPUTATIONAL STUDY OF CROSS-RANGE COHERENCE

Results of numerical simulations of sound traveling through internal wave packets are presented in this section. These are shown for two purposes: To verify the validity of the cross-range coherence scale expression (2), and to obtain estimates of the scale S. These simulations are similar in nature to those appearing in previous publications (Duda, 2004; Duda and Preisig, 1999). The previous simulations were of propagation along single paths, and the papers report quatities such as coherence time scales T_c, single-point intensity variances, depth-averaged intensity variances, and inten-

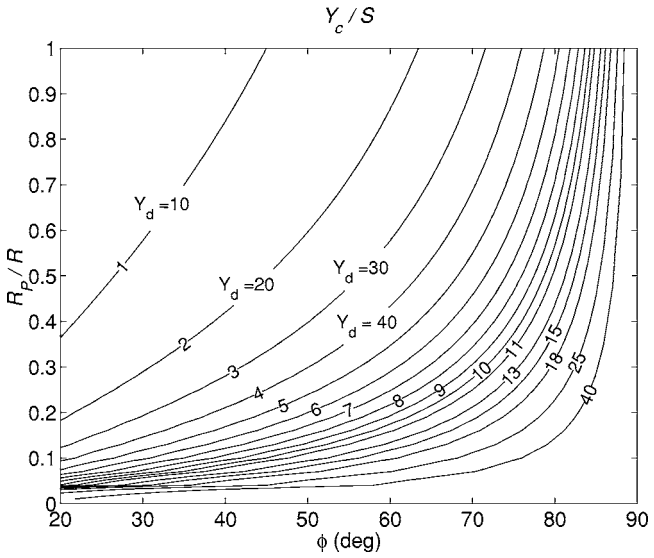


FIG. 3. Cross-range coherence scale $Y_c(\phi, R_P/R)$, divided by S for normalization, is contoured. The strictly numerical labels of the contours indicate Y_c/S. The other labels indicate nondimensional Y_d, where the representative value $S/\lambda = 10$ has been used (see Table III.)

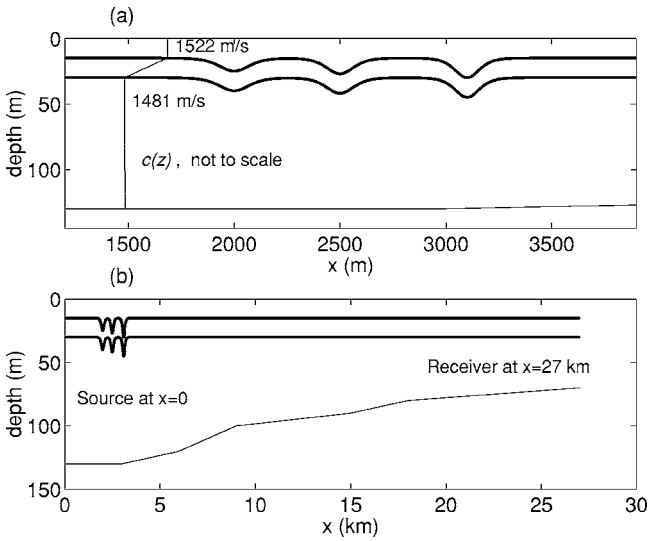


FIG. 4. (a) The background sound speed profile and the three-wave packet geometry used in the simulations are shown. The thick lines show the interfaces between the isovelocity upper layer, the gradient layer and the near-isovelocity lower layer. The waves are shown at packet position $R_P = 2000$ m. The wave shapes are shown in a plane normal to the wave troughs, which is parallel to their direction of propagation. The distances between the wave crests are 500 and 600 m. The waves have amplitudes a of 10, 12, and 15 m, and horizontal scales L of 95, 86, and 77 m, agreeing with solutions to the KdV equation for a density profile that is consistent with the sound speed profile, (b) The water depth between the source in 130 m deep water and the receiver plane in 70 m deep water is shown. The waves seen in panel (a) are shown again.

sity biases compared to a wave-free state. Mode-coupling behavior was also diagnosed. To examine cross-range coherence, the new simulations must have an expanded scope, with $N\times$2-D or 3-D computations taking the place of the single-plane computations. $N\times$2-D computations are chosen for use here, as explained in the introduction.

The results reported in this section are for monochromatic 200 Hz sound. A second set of simulations, reported in Sec. IV, are for monochromatic sound at selected frequencies between 100 and 900 Hz. The simulations are of gently upslope propagation from water having 130 m depth at the source to water having 70 m depth at the receiver plane **P**, 27 km from the source. The bathymetry is shown in Fig. 4. The source is 100 m below the surface unless specified to be 20 m, as it is in a few cases.

A. Description of the computations

A version of the RAM code originating at the Naval Research Laboratory is used. The input and output routines, and bottom parameter allocation within the depth grid, have been modified from the original code (RAM 1.1). The code uses methods developed by Collins and collaborators (Collins and Westwood, 1991). The following parameters are used: range step 0.5 m; depth grid interval 0.25 m; Pade expansion parameter 2. Attenuation in the seafloor is $\beta = 0.1$ db/wavelength to a depth of 6 m, then increases linearly to $\beta = 10$ db/wavelength at 50 m into the seabed, so that $\beta = 1$ at about 12 m depth. (This attenuation is unrealistically high and is the result of a coding error. Some simulations were repeated with the intended value of $\beta = 0.1$ for the top 50 m of the seabed, and the results were the same. The results are also relatively insensitive to the seafloor sound speed and seafloor density.)

The $N\times$2-D simulations of propagation through moving 3-D wave packet structures are accomplished by organizing N runs at a varying azimuth through wave structures, as shown in the plan view in Fig. 1, for each packet position R_P. Thus, an investigation of M values of R_P would require $N \times M$ runs. The runs terminate at a tangent plane at $R = 27$ km, with ΔY ranging from 0 to 1039 m, with associated azimuthal propagation ranges of 27 to 27.02 km. The azimuths are chosen such that the termination ranges are integer multiples of the range step of 0.5 m, resulting in an unequal ΔY sampling that causes no difficulty. The 27 km run (the shortest) has a packet at R_P, with packet positions for the other runs (i.e., those with $\Delta Y \neq 0$) computed using analytic geometry.

The speed and density profiles used in the simulations are similar to those used previously (Duda, 2004), with the exception that a higher subseafloor sound speed is used here. The sound-speed profile in the seabed is not a function of range. At the, seafloor the speed is always 1600 m/s regardless of the water depth, and has a gradient of 1 m/s for the top 50 m of the seafloor. At 50 m depth in the seabed, it then rises from 1650 to 2400 m/s to simulate the underlying bedrock.

Packets are composed of three waves, each of form $\zeta(x,t) = a\,\mathrm{sech}^2[(x-ct)/L]$, which are solutions of the Korteweg-de Vries (KdV) equation (Lamb and Yan, 1996; the Lee and Beardsley, 1974). The wave shapes with respect to range must be stretched by $1/\cos\eta$ within each of the propagation planes. An additional correction of stretching as a function of the azimuth, required for the wave shapes to be strictly correct, is not needed because the azimuthal domain is small.

Unlike earlier 2D simulations (Duda and Preisig, 1999; Preisig and Duda, 1997), these 3-D simulations require that a

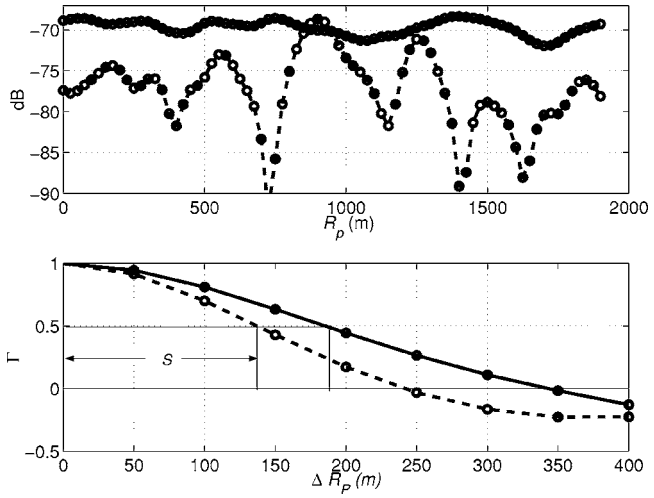


FIG. 5. (Lower panel) Correlation functions are shown for near bottom (dashed line) and depth-averaged (solid line) intensity at a 27 km receiver for ϕ=70° moving-packet 200 Hz simulations. The calculation is made for wave packets at ranges R_P=0 to 1900 m from the source, with a 25 m increment. Decorrelation displacement scales S estimated from the correlation functions are indicated. (Upper panel) The intensity fields from which the correlation functions are computed are shown. Two are shown: The depth-averaged water-column intensity (solid line) and the intensity at the bottom (dashed line). The scale is signal arriving from a 0 dB source (i.e., the additive inverse of transmission loss.)

and L be chosen for each wave such that ζ are solutions of KdV (i.e., properly scaled waves must be used, then $1/\cos\eta$ stretching applied, rather than using a variety of L for a given a to represent a variety of η). Appropriate wave parameters for a density profile consistent with our sound-speed profile are L:a pairs of 95:10, 86:12, and 77:15, with all dimensions in meters. Phase speeds for these waves are near $|\mathbf{C_P}|$ =0.73 m/s. Figure 4 shows water column (sound-speed profile) displacements in the acoustic propagation plane for an example wave packet at η=0° (ϕ=90°), so that the wave shapes are not stretched, as they would be for other values of η. The waves do not evolve over time in our simulations, unlike their true behavior. This means that acoustic coherence over long time scales cannot be addressed. This is not an important concern because the results shown here are consistent with the rapid loss of coherence, and adding wave packet evolution (time scale tens of minutes) would reduce coherence scales below the already small values shown here.

B. Results

First, estimates of S from the 200 Hz simulations will be examined, then estimates of Y_c will be presented. Figure 5 shows the depth-averaged intensity and near-bottom intensity at a single point (ΔY=0) as a function of R_P for a ϕ=70° simulation with the packet near the source. The cross-correlation function given by $\Gamma(\Delta R_P)=\langle F(R_P)F(R_P+\Delta R_P)\rangle$ is shown, with ΔR_P normalized by the acoustic wavelength. The angle brackets indicate ensemble average, computed in this paper by averaging results over small ranges of R_P. The variable F is either the depth-averaged or near-bottom intensity with the mean removed, normalized such that $\Gamma(0)=1$. The point where Γ drops to 0.5 is used here as the decorrelation scale. The correlation drops to one-half at the indicated S values of 190 and 140 m for the two measures of intensity, respectively.

TABLE I. Computed Y_c and S values from 200 Hz simulations are given for various wave angles ϕ and scaled packet ranges R_N=R_P/R. Y_c is reported in ratio form, Y_c/S. The theoretical ratio Y_c/S is also reported. Both S and Y_c are computed using the near-bottom field (for case 1, the dashed lines in Fig. 5). The notation nc means not computed. The value with the asterisk is extrapolated from a correlation function curve that has a value near 0.6 at the longest computed lag, above the threshold of 0.5.

Identifier	ϕ (°)	R_N	S (m)	Y_c/S simulated	Y_c/S using (2)
Case 1	70	0.04	140	nc	69
Case 2	70	0.18	118	10*	15
Case 3	70	0.50	77	5.7	5.5
Case 4	70	0.93	73	3.7	3.0
Case 5	50	0.50	78	2.5	2.4
Case 6	50	0.93	79	1.4	1.3

Table I gives S values for the simulation shown in Fig. 5 (case 1), and other cases numbered 2 through 6 covering over a variety of ϕ and R_N. In each case the correlation function used to compute the scale stems from an ensemble average taken over a small range of R_N. Therefore the R_N values in Table I are merely representative for each case. It would be preferable for S to be uniform over ϕ and R_N, which it is not, because then (1) and (2) would be most useful. S is reasonably uniform, however, except at small R_N.

To illustrate the lateral decorrelation effect parameterized by Y_c, representative images of computed acoustic fields in the cross-range tangent plane **P** are now shown. Figure 6 shows a snapshot of the intensity in this plane for a near-

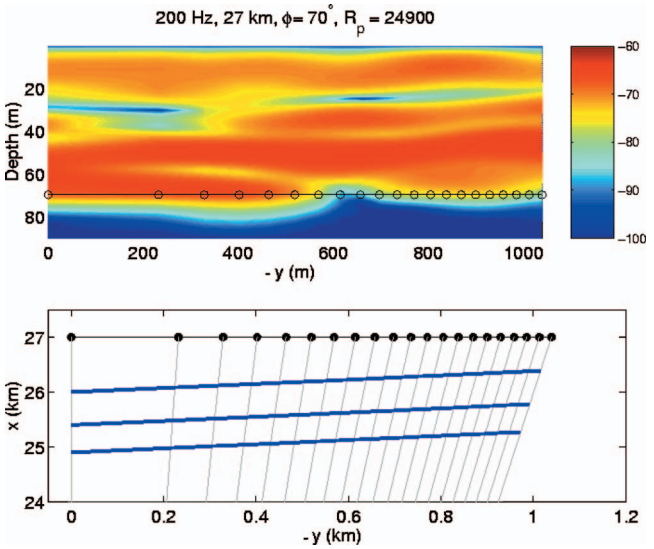


FIG. 6. (Upper panel) The intensity in the tangent plane **P** is contoured. The signal level in dB from a 0 dB source is shown. The left edge is 27 km from the source (200 Hz, at the standard 100 m depth); the right edge is 27.02 km from the source. Results with packet angle ϕ=70° at R_P=24.9 km are shown. The line of circles show the positions that bottom mounted receivers at the terminations of the N azimuthal propagation lines would have. (Lower panel) The positions of the wave troughs in the three-wave packet corresponding to the upper panel results are shown in this plan view. The acoustic source is at (0,0). The azimuthal propagation lines are shown. The intersection of the receiver plane **P** with the seafloor is shown with the black symbols.

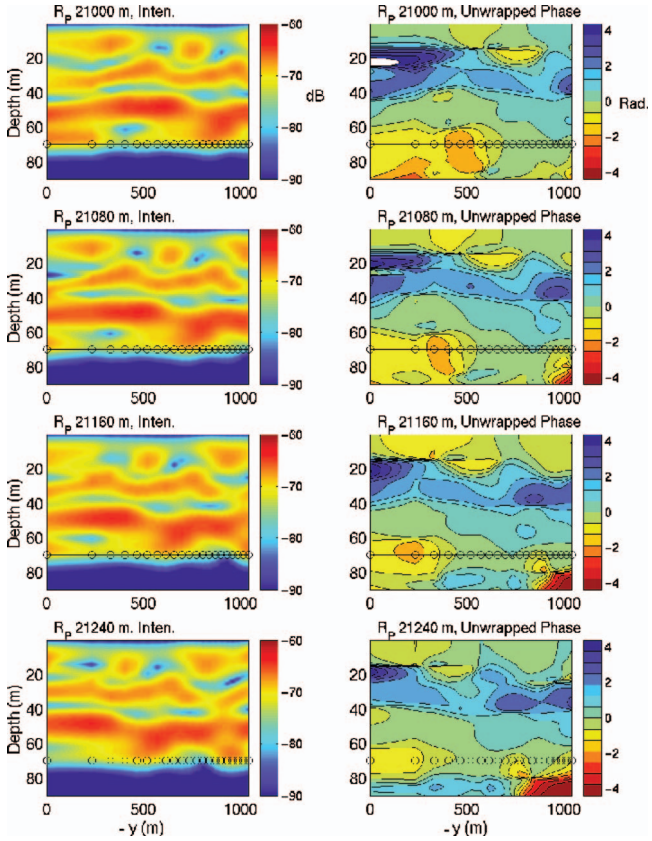


FIG. 7. Intensity and phase in the cross-range (tangent) plane **P** at range 27 km are shown for a $\phi=50°$ packet at four different R_P. The acoustic frequency is 200 Hz and the source depth is 20 m. The format is similar to the upper panel of the previous figure. The R_P are 21 000 to 21 240 in 80 m intervals, shown from top to bottom. For each of the four R_P, the intensity is contoured with a continuous color map at the left, and the unwrapped phase is contoured in a stepwise fashion at the right. The same section of **P** is shown in each panel. The viewpoint is such that sound would be propagating into the page. The fluctuation features that can be seen are typical for simulations over a range of frequencies, R_P, and ϕ. It is evident that the features move from right to left in the figure as the packet moves toward the receiver.

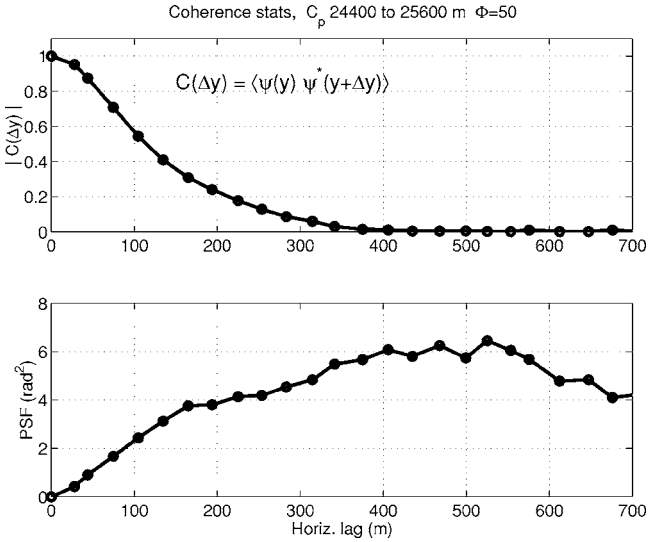


FIG. 8. (Upper panel) The magnitude of the complex cross-range lagged correlation function for a near-bottom 200 Hz acoustic field is shown for the case $\phi=50°$, $24.4<R_P<25.6$ km. The source depth is the standard 100 m. This is computed after normalizing the amplitudes across the array so that they average unity at each position, to account for the slightly different range to each position. This ensures that the value at zero lag is one. (Lower panel) Mean squared phase difference for the same data. The points shown in each panel are averages taken over R_P and over a few different lags falling into lag bins. The points shown are (average lag, average product) for each lag bin.

receiver wave packet. The figure is drawn such that sound would be propagating into the page. This snapshot is from case 4 (Table I). The intensity exhibits features with scale length Y_c of approximately 270 m. Next, Fig. 7 shows four sequential snapshots of fluctuating cross-range fields from simulated moving packets. The R_P increment is 80 m, and R_P is an increasing function of time, so that time would progress down the page. Intensity and phase are contoured on the left and right, respectively. Features are seen to move from right to left with increasing R_P or time. This is because the range of the packet from the source is a decreasing function of azimuth from right to left, as illustrated in Fig. 1. In this situation, acoustic field features, which are closely linked to the packet range from the source (Duda and Preisig, 1999), move from right to left as time progresses. Finally, note that cross-range features have a smaller scale length for this $\phi=50°$ case than for the $\phi=70°$ case of Fig. 6.

Table I includes Y_c/S computed directly from simulations for cases 3–6. These are computed from lagged correlation functions of the complex field, $C(\Delta Y)=\langle\Psi(\Delta Y)\Psi^*(Y=\Delta Y)\rangle$. (For case 1, $C(\Delta Y)=0.96$ at the longest lag, so the coherence scale is very long and was not computed; for case 2 with the asterisk, C was extrapolated beyond the maximum available lag of $\Delta Y=950$ m.) All of these simulation-based Y_c values are derived from near-bottom cross-range acoustic field properties, as explained in the next paragraph. Except for one value, these agree well with the values of Y_c/S computed from only R, R_P, and ϕ using (2), which are also listed in the table. The value that does not agree well is for case 2, which is a situation having waves very close to the source. In this situation, the computed correlation length is shorter than that given by the simple expression. A probable reason for this is that the curved line of constant range from source seen in Fig. 1(b) departs from the dashed line of length Δy, so that the approximation $\delta x=\Delta x$ breaks down, with $\delta x>\Delta x$.

Figure 8 shows a typical lagged correlation function of the complex field, C, from which the computational Y_c estimate is derived. Case 6 results are shown. The complex fields are tabulated as functions of cross-range horizontal lag and R_P, then normalized so that they have an average magnitude of one at each cross-range position. The cross-products are then computed, and finally the results are ensemble averaged to obtain meaningful averages at each lag. This process eliminates the (very small) effect of a weak intensity trend generated by the differing ranges to the various ΔY points. The lower panel of Fig. 8 shows the phase structure function (the mean squared phase difference $\mathrm{PSF}(\Delta y)=\langle[\varphi(y)-\varphi(y+\Delta y)]^2\rangle$) computed from the same dataset. The rapid increase of phase difference as a function of lag indicates that phase changes are responsible for much of the cross-range decorrelation within the complex field.

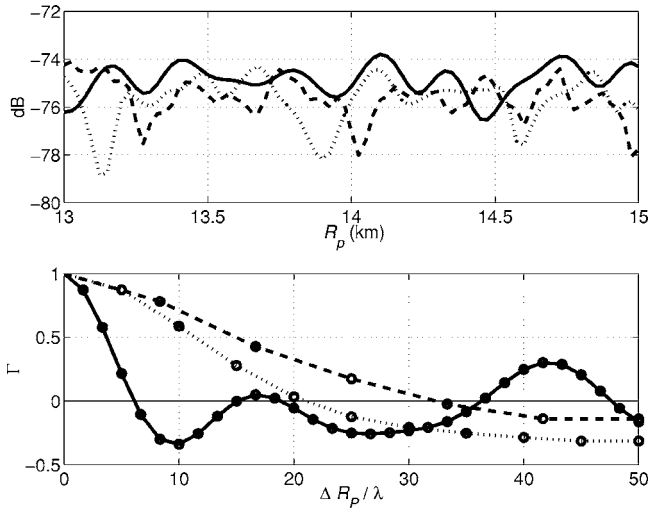


FIG. 9. (Upper panel) For moving wavepacket propagation runs having $\phi = 70°$ and source depth 20 m, depth-integrated intensity E within the water column is shown as a function of R_P for three frequencies: 100 Hz (solid); 300 Hz (dot); and 500 Hz (dash). (Lower panel) The cross-correlation function $\Gamma = \langle E_f(R_P) E_f(R_P + \Delta R_P) \rangle / \langle E_f^2 \rangle$ is shown, with ΔR_P normalized by the acoustic wavelength. E_F is E with the mean removed and normalized such that $\Gamma(0) = 1$.

IV. COMPUTATION OF *S* FOR A RANGE OF FREQUENCIES

The presence of the displacement decorrelation scale S in (1) and (2) means that it plays a critical role in determining the scales of temporal and cross-range coherence (T_c and Y_c), at least for the idealized cases of nonevolving and slowly evolving wave packets. This role was validated for 200 Hz sound in the previous section (Table I). Therefore, uncovering the frequency (f) dependence of S will help divulge the frequency dependence of T_c and Y_c. Here, this issue is given an initial exploratory treatment: Our simulations are extended to other frequencies and $S(f)$ is computed for $100 < f < 900$ Hz. In general, as pointed out previously, mode coupling in internal waves depends on many aspects of acoustic waveguide vertical structure, wave amplitudes, and horizontal scales (Duda and Preisig, 1999; Preisig and Duda, 1997), so analysis of the frequency dependence of the problem is beyond the scope of this paper.

Correlation functions $\Gamma(\Delta R_P)$ of depth-integrated intensity for moving packets at three frequencies (100, 300, and 500 Hz) are shown in the lower panel of Fig. 9. These can be compared with Fig. 5, although the packet-location lags in Fig. 9 are normalized by an acoustic wavelength. The series from which the correlation functions are computed are shown in the upper panel. The same geometry is used as in all previous simulations. The figure shows S to range from 4 to 15 times the acoustic wavelength. The intensity fluctuation variances, not tabulated here, are comparable to those resulting from previous simulations (Duda, 2004; Duda and Preisig, 1999).

Estimates of S derived from Γ curves such as those of Fig. 9 were obtained for frequencies from 100 to 900 Hz, and are given in Tables II and III. Table II shows values for the case of a 100 m source depth. Table III shows values for the case of a 20 m source depth. These estimates of S are given in dimensional units of length, and also in nondimensional form after normalization using the acoustic wavelength λ. The results shown in Fig. 9 are included in Table III. The 200 Hz near-bottom S value of 85 m in Table II is from a run similar but not identical to the run used to compute the directly comparable 200 Hz S value of 77 m in Table I, with the runs having differing R_P spacing and window size. These differing S demonstrate the practical confidence intervals on $S(\phi, R_N, f)$.

TABLE II. S is computed for various frequencies. The wave angle ϕ is 70° and the R_P range is from 13 to 15 km, so that $R_P/R \approx 0.5$. S is computed for near-bottom intensity and for depth-integrated intensity. The source depth is 100 m. λ is the acoustic wavelength.

Frequency (Hz)	S/λ nr. btm.	S/λ depth int.	S nr. btm. (m)	S depth int. (m)
100	6	4	87	66
200	11	9	85	66
300	7	8	36	42
400	11	14	40	52
500	10	16	31	49
600	11	13	27	33
700	10	18	21	39
800	12	13	23	24
900	11	23	19	39

The values of $S(70°, 0.5, f)$ for a 100 m source depth for the chosen waveguide and source-receiver parameters, given in Table II, are not strictly constant with respect to frequency in either dimensional or nondimensional form. The quantity closest to being constant is the nondimensional form S/λ computed for the near-bottom field, with a value of about $S/\lambda = 10$. The corresponding S decrease strongly with increasing frequency. On the other hand, when computed from the depth-integrated intensity, the dimensional S has a more constant value with respect to frequency than the nondimensional S/λ. However, S derived from the depth-integrated intensity decreases by approximately a factor of 2, as frequency increases from 100 to 900 Hz.

The same two statements can be made for the 20 m source depth case (Table III), namely, that the nondimensional near-bottom $S/\lambda \approx 10$, and that, for the depth-integrated field, dimensional S is a weaker function of fre-

TABLE III. S is computed for various frequencies. All parameters of the runs are identical to those reported in Table II, with the exception that the source depth is 20 m. The fluctuations of the near-bottom series are 4 to 6 dB rms. Fluctuations of the depth integrated series are near 1.0 dB rms.

Frequency (Hz)	S/λ nr. btm	S/λ depth int.	S nr. btm (m)	S depth int. (m)
100	5	4	73	55
200	16	8	117	59
300	9	11	43	57
400	9	11	34	40
500	8	15	24	45
600	7	17	18	43
700	11	19	23	40
800	9	27	17	50
900	13	31	22	51

quency than nondimensional S/λ. The first would be even more true in the absence of one outlier of near-bottom S/λ (the 200 Hz entry, which is 16).

The result that is most clear is that S/λ ranges from 4 to 31 in this frequency band. Variations within this range depend on how S is computed and on frequency in ways that are yet to be determined. If all the numbers in the two tables are used, nondimensional S/λ has a mean of about 12, with a standard deviation of 6.

V. NONDIMENSIONAL CROSS-RANGE COHERENCE SCALE

A nondimensional cross-range correlation scale can be obtained using the relationship between Y_c and S and using S/λ results from the preceding section. The scale can be nondimensionalized as follows: $Y_d = Y_c/\lambda = (Y_c/S)(S/\lambda)$. Figure 3 shows a contour plot of Y_c/S, and Tables II and III provide S/λ values.

For the 100 m source depth situation (Table II), the value of $S/\lambda = 10$ for the near-bottom intensity case allows a numerical conversion of Y_c/S into Y_d, with resultant Y_d values indicated in Fig. 3. The contours indicate that approximately half the possible situations show $Y_d < 50\lambda$. These are at the upper left. The shortest scales are for waves that are near the receiver having low angle ϕ (but not less than 20°, where the assumptions and the simulations both break down).

The situations with longer Y_d, which are those having high ϕ, having packets near the source, or having both, are equally likely to exist in nature. This means that field estimates of Y_d may take on many values ranging from a few wavelengths to hundreds of wavelengths, depending on the internal wave geometry, interval between packets, whether packets come from one direction or multiple directions, and so on.

The location of sources and receivers with respect to generation sites of nonlinear internal wave packets, in particular, may influence which part of Fig. 3 gives the relevant estimate of Y_c/S for a particular experiment, and thus the relevant Y_d. The location may also influence the acoustic propagation and mode coupling, such that S/λ may diverge from a constant, or such that it takes on a different value than 10 used in the figure. It is also possible that wave packets traveling in different directions might pass periodically between the source and the receiver. In this case, if each of the classes of waves (assign waves in each direction to a class) caused fluctuations of approximately equal variance, then the overall field-observed correlation scales would be determined by the waves with the shortest scales Y_c and T_c. Figure 3 suggests that these would be the waves of low ϕ close to the receiver, giving correlation lengths of 20λ or less.

For a given coastal area, it may be possible to reliably model S/λ at the frequency of interest. Subsequently, the source location, receiver location, and local waveguide geometry (obtained via *in situ* or remote techniques) can then be used to determine ϕ and R_N, yielding Y_c/S and Y_d.

VI. SUMMARY

We have used two simple models to relate fluctuation effects caused by the presence of internal waves or wave packets to temporal and spatial acoustic coherence scales. The first uses the internal wave velocity to convert the effect of internal wave displacement into a temporal coherence scale. The second uses the plan form geometry of the internal waves to convert the effects of variable internal wave distance from the source into a cross-range coherence scale. Computational simulations give results that agree with predictions from the simple models.

A common factor in the two coherence scale expressions is the internal wave horizontal displacement scale S associated with significant acoustic fluctuation. This scale is governed in part by mode shapes and wave numbers, and depends on details of the environment such as pyncocline geometry, water depth, source depth, and proximity of the waves to either the source or the receiver. Computational examples of how internal wave propagation (horizontal displacement, in the terminology used here) causes acoustic fluctuations can be found here and in other articles (Duda, 2004; Duda and Preisig, 1999; Rouseff *et al.*, 2002).

One of those papers (Duda and Preisig, 1999) contains an explanation of how input mode interference can control the scale S. Briefly, the time series of relative amplitudes of various modes on the receiver side of moving packets, which are tied to field fluctuations at the receiver in the absence of further significant coupling, are governed by the spatial beat patterns between the incident modes that couple significant energy into each of the modes. The dimensions of these patterns depend on the mode coupling coefficients, which in turn depend on the depth integral of the product of the coupled mode shapes. Potentially useful expressions governing fluctuations for any waveguide and frequency were included in the paper, but the coefficients and functions in the expressions (mode amplitudes and phases, mode shapes) are difficult to predict because of water column and bottom property uncertainty and variability.

Temporal coherence scales are estimated to be 30 to 120 s, based on S values of 40 to 100 m (Fig. 2). These correspond well to field estimates at 400 Hz (Duda *et al.* 2004) and at 300 and 500 Hz (Mignerey and Orr, 2004) Spatial scales are estimated to be from a few times the acoustic wavelength up to extremely long values, depending on the position of the waves in the source-receiver path and the angle of the waves with respect to that path.

The temporal coherence scale expression (1) contains the internal wave velocity $\mathbf{C_P}$. This can be obtained from satellite or aircraft remote sensing, pairs of moorings, ship observations, or combinations of these. Thus, the unknown in (1) for a particular site in the field will often be the important scale length S. Likewise, the unknown in the coherence scale expression (2) will also often be S. As indicated in Sec. IV, S often has values of order 10λ (order 50 m for chosen frequencies), but exhibits a broad range from 4λ to 31λ (17–120 m), with greater variation not ruled out. There-

fore, understanding the behavior of S is critical to the application of the results shown here to prediction of acoustic coherence in the field.

An important consideration is the fact that S may be a function of ϕ for a given wave packet. This is not allowed for here, so that the coherence scale formula (2) and Fig. 3 are in fact simplifications that do not reflect that possibility. They must therefore be applied carefully.

The results shown here are for monochromatic acoustic propagation. The problem of pulse spatial variability is more complicated. In a best-case scenario, the correlation behavior would be independent of frequency throughout the entire band of a broadband source (say, with a bandwidth/center frequency ratio of 1/4), and a pulse composed of a sum of frequency constituents might be expected to have behavior similar to that shown here. Alternatively, the changes in mode coupling (and the resultant field after propagation through a packet) that occur as a packet is displaced may be sufficiently frequency dependent for the results shown here to not apply directly to pulse coherence. Analyses of how the acoustic bandwidth affects coherence for signals expressible in terms of sums of Gaussian wave packets (multipath) are contained in Colosi and Baggeroer (2004) and Colosi *et al.* (2005) and may be useful to apply, or broadband simulations may be used.

One point about the numerical simulations must be noted. In this paper, characteristic values of S were computed from intensity fluctuations rather than from the complex field. This was done in order to be consistent with prior work. On the other hand, Y_c values were computed using the complex field correlation function. The complex field was chosen for this because of the role of the phase in array processing and signal processing applications. It is possible that Y_c values obtained using the complex field may in some cases be dominated by phase fluctuations, which are not considered in the S evaluations. This means that in some cases the comparison of S with Y_c may not precisely fit (2). This only affects the evaluation of (2) via simulation, and does not influence the applicability of (2).

Finally, modifications to the derivation can extend the work to nonbroadside incidence of the signals at the receiver. Such modifications would yield a generalized version of (2).

ACKNOWLEDGMENTS

The 2-D phase unwrapping code was provided by J. Colosi. This work was funded by grants from the Ocean Acoustics Program of the U.S. Office of Naval Research.

Badiey, M., Katsnelson, B. G., Lynch, J. F., Pereselkov, S., and Siegmann, W. L. (**2005**). "Measurement and modeling of three-dimensional sound intensity variations due to shallow-water internal waves," J. Acoust. Soc. Am. **117**, 613–625.

Chiu, C.-S., Ramp, S. R., Miller, C. W., Lynch, J. F., Duda, T. F., and Tang, T. Y. (**2004**). "Acoustic intensity fluctuations induced by South China Sea internal tides and solitons," IEEE J. Ocean. Eng. **29**, 1249–1263.

Collins, M. D., and Westwood, E. K., (**1991**). "A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed, and density," J. Acoust. Soc. Am. **89**, 1068–1075.

Colosi, J. A., and Baggeroer, A. B. (**2004**). "On the kinematics of broadband multipath scintillation and the approach to scintillation," J. Acoust. Soc. Am. **116**, 3515–3522.

Colosi, J. A., Baggeroer, A. B., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Dushaw, D. B., Howe, B. M., Mercer, J. A., Spindel, R. C., Birdsall, T. G., Metzger, K., and Forbes, A. M. G. (**2005**). "Analysis of multipath acoustic field variability and coherence in the finale of broadband basin-scale transmissions in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1538–1564.

Duda, T. F. (**2004**). "Acoustic mode coupling by nonlinear internal wave packets in a shelfbreak front area," IEEE J. Ocean. Eng. **29**, 118–125.

Duda, T. F., and Preisig, J. C. (**1999**). "A modeling study of acoustic propagation through moving shallow-water solitary wave packets," IEEE J. Ocean. Eng. **24**, 16–32.

Duda, T. F., Lynch, J. F., Newhall, A. E., Wu, L., and Chiu, C.-S. (**2004**). "Fluctuation of 400 Hz sound intensity in the 2001 ASIAEX South China Sea experiment," IEEE J. Ocean. Eng. **29**, 1264–1279.

Frank, S. D., Badiey, M., Lynch, J. F., and Siegmann, W. L. (**2004**). "Analysis and modeling of broadband airgun data influenced by nonlinear internal waves," J. Acoust. Soc. Am. **116**, 3404–3422.

Lamb, K. G., and Yan, L. (**1996**). "The evolution of internal wave undular bores: Comparisons of a fully nonlinear numerical model with weakly nonlinear theory," J. Phys. Oceanogr. **26**, 2712–2734.

Lee, C., and Beardsley, R. C. (**1974**). "The generation of long nonlinear internal waves in a weakly stratified shear flow," J. Geophys. Res. **79**, 453–462.

Lynch, J. F., Newhall, A. E., Sperry, B., Gawarkiewicz, G., Fredricks, A., Tyack, P., Chiu, C.-S., and Abbot, P. (**2003**). "Spatial and temporal variations in acoustic propagation characteristics at the New England Shelfbreak Front," IEEE J. Ocean. Eng. **28**, 129–150.

Mignerey, P. C. and Orr, M. H. (**2004**). "Observations of matched-field autocorrelation time in the South China Sea," IEEE J. Ocean. Eng. **29**, 1280–1291.

Oba, R., and Finette, S. (**2002**). "Acoustic propagation through anisotropic internal wave fields Transmission loss, cross-range coherence, and horizontal refraction," J. Acoust. Soc. Am. **111**, 769–784.

Preisig, J. C., and Duda, T. F. (**1997**). "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269.

Rouseff, D., Turgut, A., Wolf, S. N., Finette, S., Orr, M. H., Pasewark, B. H., Apel, J. R., Badiey, M., Chiu, C.-S., Headrick, R. H., Lynch, J. F., Kemp, J. N., Newhall, A. E., von der Heydt, K., and Tielbuerger, D. (**2002**). "Coherence of acoustic modes propagating through shallow water internal waves," J. Acoust. Soc. Am. **111**, 1655–1666.

Zhou, J., Zhang, X., and Rogers, P. H. (**1991**). "Resonant interaction of sound wave with internal solitons in the coastal zone," J. Acoust. Soc. Am. **90**, 2042–2054.

Dugong (*Dugong dugon*) vocalization patterns recorded by automatic underwater sound monitoring systems

Kotaro Ichikawa,[a] Chika Tsutsumi, and Nobuaki Arai
Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

Tomonari Akamatsu
National Research Institute of Fisheries Engineering, 314-0408 Ibaraki, Japan

Tomio Shinke
System Intech Co., Ltd., R&D Center, 424-8610 Shizuoka, Japan

Takeshi Hara
Japan Fisheries Resource Conservation Association, 104-0054 Tokyo, Japan

Kanjana Adulyanukosol
Phuket Marine Biological Center, 83000 Phuket, Thailand

(Received 29 September 2005; revised 6 April 2006; accepted 10 April 2006)

To quantitatively examine the diurnal, or tidal, effects on dugong behavior, we employed passive acoustic observation techniques to monitor the animals. Automatic underwater sound monitoring systems for dugongs (AUSOMS-D) were deployed on the sea floor at depths of about 5 m south of Talibong Island, Thailand. The AUSOMS-D recorded underwater sound in stereo at a sampling frequency of 44.1 kHz for more than 116 consecutive hours. Dugong calls were automatically detected by newly developed software with a detection rate of 36.1% and a false alarm rate of 2.9%. In total, 3453 calls were detected during the 164 h of recording. The autocorrelation of the call rate indicated an attendance cycle of about 24 or 25 h, and the most frequent vocalizations were observed from 0300 to 0500 h. The calculated bearings of the sound sources, i.e., dugongs, were used as an indicator to track the relative numbers of dugongs during the monitoring periods. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2201468]

PACS number(s): 43.30.Sf, 43.80.Ka, 43.30.Wi, 43.60.Vx [WWA] Pages: 3726–3733

I. INTRODUCTION

The dugong (*Dugong dugon*) is the only strictly marine mammal that is a benthic-feeding, seagrass specialist (Marsh *et al.*, 2002). Although the world's largest dugong population is found in Australian waters, the local populations in many other locations are becoming endangered (e.g., Marsh *et al.*, 2002; Preen, 1998; Preen, 2004). The lack of ecological and behavioral information on this species, however, has slowed conservation actions in many places. Behavioral and ecological information on local populations of dugongs is urgently needed for the implementation of appropriate protection measures.

Dugongs are usually found in shallow waters, spending 72% of their day in waters less than 3 m deep (Chilvers *et al.*, 2004), but little information is available on the diurnal, or tidal, effects on the presence, swimming direction, and minimum group size of the animal. Conventional behavioral observations have relied largely on visual observations from either airplanes or ships, or sometimes from land observatory platforms (Anderson, 1986; Anderson, 1995a; Marsh and Sinclair, 1989; Marsh *et al.*, 1999). However, visual surveys are problematic; the probability of finding dugongs depends on several factors, such as the experience level of the observers, weather conditions, and water clarity. An additional weak point of visual observation is that it cannot be performed at night.

Recently, passive acoustical monitoring techniques are being used for observing the behavior of marine mammals. In response to West Indian manatees (*Trichecus manatus latirostris*), which are in the-same order (Sirenia) as the dugong, being accidentally struck by boats, Niezrecki *et al.* (2003) devised a method to distinguish manatee calls from ambient noise, and 96% of manatee vocalizations were correctly identified, with a false alarm rate of 16%. Niezrecki *et al.* (2003) suggested a warning system to alert boaters to the presence of manatees, and could thereby avoid collisions. Wang *et al.* (2005) and Akamatsu *et al.* (2001) compared the detection of finless porpoises in China using visual and acoustic observations, and acoustic observations were determined to be highly reliable in detecting finless porpoises emitting ultrasonic pulses because of their frequent phonation (Akamatsu *et al.*, 2005).

Acoustic observation has several advantages over visual observation. It can be performed at a constant detection efficiency for a long period, and consecutive observations, including night monitoring, are useful for better understanding study animals. For subjects with easily distinguished calls in habitats without great geographical constraints, acoustic observation could be a powerful outdoor observational tool.

[a]Electronic mail: ichikawa@bre.soc.i.kyoto-u.ac.jp

 0001-4966/2006/119(6)/3726/8/$22.50

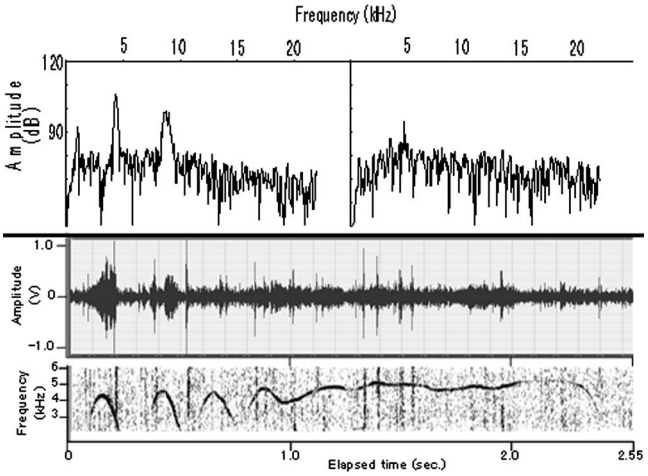


FIG. 1. A sonogram (bottom inset) and a waveform (middle inset) of dugong calls including three short-duration calls and a long-duration call, recorded in the study area. Spectrograms of a typical short-duration call (top left) and a typical long-duration calls (top right) are shown in the top inset.

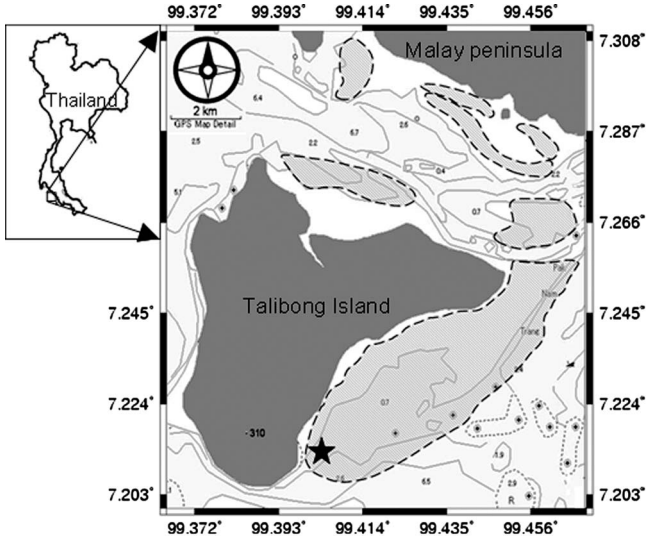


FIG. 2. A map of the study site around the southern part of Talibong Island, Trang, Thailand. AUSOMS-D, represented by the star, was deployed on the sea floor at depths of about 5 m. Shaded areas outlined by break lines indicate seagrass distribution, surveyed by Nakanishi *et al.* (2006). Most of the dugongs were sighted in/around the seagrass beds (Nakanishi *et al.*, 2005; Hines *et al.*, 2005).

Dugong vocalizations have been classified into three groups (Anderson *et al.*, 1995b): chirp-squeaks, barks, and trills. Chirp-squeaks are short, frequency-modulated signals, with two to five harmonics within the 18-kHz band. Barks appear to be broadband signals covering the 500- to 22000-Hz range with a median frequency of about 1200 Hz and no frequency modulation. The bark duration ranges from 0.03 to 0.12 s. Trills, heard as a rising cascade of notes, last 100–2200 ms, with an average and standard deviation of 850±269 ms. (Anderson *et al.*, 1995b). Anderson *et al.* (1995b) also suggested the function of each call type as follows: "Chirp-squeaks may serve as announcements of territory ownership, have a ranging function in territorial defense, and serve as signatures distinguishing neighbors from possible mate or challengers. Barks seem appropriate as warning sounds, indicating readiness to attack. Trills may indicate nonaggressive intentions" Dugong calls in southern Thailand have shown a slight difference in the acoustical characteristics that were described in the Anderson *et al.* (1995b) report, and have been roughly categorized into two groups: short-duration and long-duration calls (Ichikawa *et al.*, 2003). Out of more than 1400 dugong calls, we could not record any sound that was likely to correspond to the barks in the Anderson *et al.* (1995b) study. Short-duration calls, which were the most frequent in the study area, had an average frequency of 4521±1615 Hz and an average duration of 126±87 ms (n=704). Long-duration calls, which seemed to correspond to the trills in the Anderson *et al.*, 1995b study, were characterized by a frequency of 4152±1111 Hz and a duration of 1737±1049 ms (n=74). The highest vocalization frequency was once every 2 s. Frequencies of once in 20 to 30 s were common and suitable for the application of passive acoustical observation tools (Ichikawa *et al.*, 2003). Figure 1 shows a sonogram of three short-duration calls and a long-duration call, together with a time series waveform view and the spectrograms.

In the present study, we applied passive acoustic observation techniques to observe the diurnal, or tidal, effects on dugong behavior and to detect changes in the number of individuals in the monitored area, with the goal of characterizing the ecological aspects of dugongs in nearshore areas.

II. MATERIALS AND METHODS

A. Research site and the monitored area

Our recording experiments were conducted south of Talibong Island, Trang, Thailand, (07°12.908′N, 99°24.071′E; Fig. 2). The sea floor around Talibong Island is covered by approximately 7 km^2 of seagrass beds, the largest in Thailand, and mainly composed of *Halophila ovalis*, for which dugongs show a preference (Hines, 2002).

The seafloor is almost flat with a rather gentle slope, and few obstacles exist to obstruct the path of a sound wave in the focal area. The largest dugong population in Thailand is also found in this area (Adulyanukosol, 1999). A population of 123 dugong individuals, about 62% of the total estimated population (about 200 individuals) along the Andaman coast of Thailand, were counted by aerial surveys conducted in the Trang province (around Talibong Island and the neighboring Muk Island) in 2000 and 2001, and all dugongs sighted were in ≤5-m depth lowest low water (Hines *et al.*, 2005).

B. Recording device and its installation

Automatic underwater sound monitoring systems for dugongs version 1.0 (AUSOMS-D; System Intech Co., Ltd., Tokyo, Japan; see details in Shinke *et al.*, 2004) were used for passive acoustical monitoring. These systems were designed based on the acoustical characteristics of dugong calls described in Anderson and Birtles (1978) and Ichikawa *et al.* (2003). The AUSOMS-D consisted of a pair of hydrophones located 2 m apart for calculating the bearings of the sound sources. The electric circuits and batteries were housed in a pressure-resistant case [Fig. 3(a)], and the hydrophones were

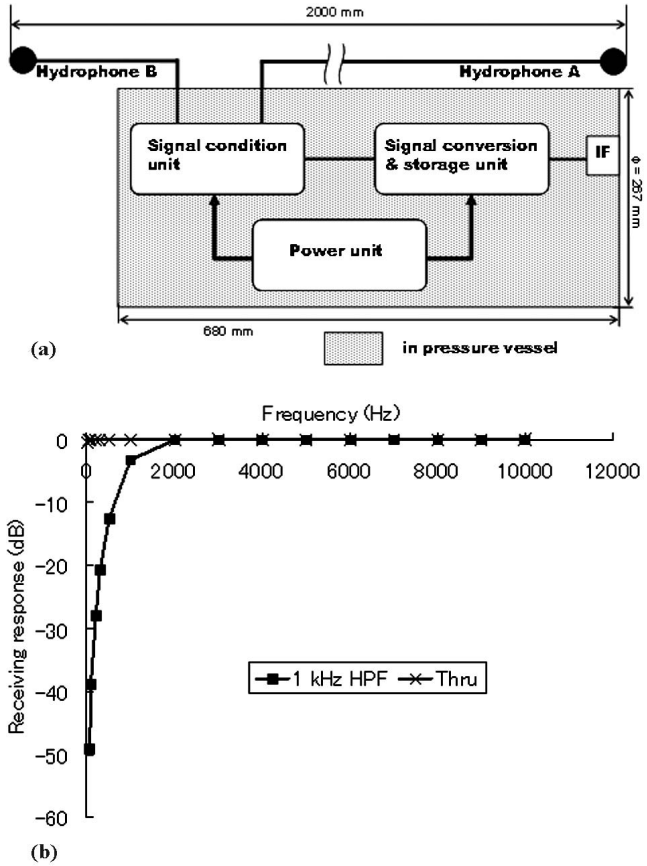


FIG. 3. (a) A block diagram of AUSOMS-D. The diameter, length, and weight (in air) of the system was 267 mm, 680 mm, and about 50 kg, respectively. (b) A complete system sensitivity/response curve with hydrophones of −190 dB, 1 kHz high-pass filter and 70 dB gain control. The system showed a flat frequency response over 2–10 kHz.

connected to a stereo preamplifier. Each hydrophone had flat frequency responses within 2 dB between 1 and 10 kHz. The hydrophone sensitivity was −190 dB (*re* 0 dB 1 V/μPa at 10 kHz). High-pass filters could be altered to thru, 200 Hz, 1 kHz, 4 kHz. The recording level could be adjusted to 70, 80, or 90 dB in advance of recording by an attenuator. Figure 3(b) shows the complete system response curve with the hydrophones of −190 dB sensitivity, 1 kHz high-pass filter, and 70 dB gain control. Digitized signals were recorded on an 80 GB removable hard disk drive in uncompressed.wav format with a time stamp. The longest recording duration was over 116 h. The power supply system with batteries and dc/dc converter were also housed inside the case. The sampling frequency of the A/D converter was 44.1 kHz and the dynamic range was 74–120 dB (*re*1 μPa) with a 16-bit resolution. The time synchronizations and recording schedules were set by the user's personal computer.

AUSOMS-D was deployed on the sea floor off the southern part of Talibong Island (07°12.786′ N, 99°24.114′ E, Fig. 2). Two sandbags were put together with the AUSOMS-D so that the instrument would not slide around and an underwater compass was used for the instrument orientation. Data were recorded for up to five consecutive days from 24 to 26 February 2004 and 28 February to 4 March 2004. The AUSOMS-D was installed on 24 February

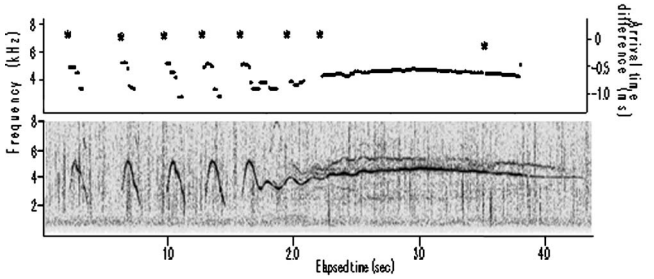


FIG. 4. An example of the results of the automated detection. Sonograms of three short-duration calls and a long-duration call are shown in the lower inset and the result of the filtering in the upper, with asterisks indicating the star-time versus the arrival time difference in a millisecond of each detection (a vertical axis on the upper left). The long-duration call was detected as a complex of four shorter-duration calls.

2004, retrieved on 26 February 2004, and the data were read out on 27 February 2004 then reinstalled in the same position on 28 February 2004. We used an anchored buoy to make sure the AUSOMS-D was positioned correctly.

To calibrate the bearing calculations, dugong calls were transmitted near the AUSOMS-D from a research vessel by an underwater speaker. The vessel trajectories measured by both GPS and the AUSOMS-D were compared.

C. Post-processing of the underwater sound datasets

Dugong calls were automatically detected from the ambient noise using newly developed software in MATLAB. The mean frequency range of the dugong vocalization was 4521±1615 Hz for the short-duration calls and 4152±1111 Hz for the long-duration calls (Ichikawa *et al.*, 2003). In order to reduce any low frequency sound component mainly produced by ship traffic, the sound signals were fed into a 1-kHz high-pass filter. To briefly describe the detection algorithm, the sound data was segmented into time windows. Then the dominant frequency component of each time window was calculated by taking the Fast Fourier Transform. After the summation processing, if the dominant frequency components of the neighboring time windows exceed the spectrum-level threshold, and continue and persist for more than length thresholds, the components was considered as a dugong call. The following describes the algorithm in detail (an examples is shown in Fig. 4).

(1) Underwater sound data were segmented into 1.8 ms time windows.
(2) To calculate the power spectrum density for each time window, a fast Fourier transform (FFT, 1024 points) was applied to every 80 time points. The spectrum density was named submatrix $\{F_{mn}\}$ ($m=1,2,3,\ldots,N, n=1,2,\ldots,512$), where N is the number of the time window. Each F_m had a 512-term frequency range from 0 to 22 kHz.
(3) Then, 25-moving summation processing was used to achieve a new matrix $\{S_{mn}\}$ ($m=1,2,3,\ldots,N$, $n=1,2,\ldots,512$) defined from F_{kn} by taking the summation of each corresponding frequency component of the 25-term successive submatrix $\{F_{kn}\}(k=m,m+1,m+2,\ldots,m+24,n=1,2,\ldots,512)$ [Eq. (1)],

TABLE I. Threshold settings and the detection efficiency.

Spectrum level (dB rms re $1\mu Pa^2/Hz$)	Duration (ms)	Frequency modulation (Hz)	Detection rate (%)	False alarm rate (%)	Number of calls
87.75	60	100	36.1	2.9	355

$$S_{mn} = \sum_{k=m}^{m+24} F_{kn}. \quad (1)$$

Equation (1) was repeated for $n=1$ to $n=512$ to obtain matrix $\{S_{mn}\}$. Given that dugong calls are narrowly frequency modulated at around 4 kHz over a duration of 60 ms (Anderson and Birtles, 1978; Ichikawa *et al.*, 2003), the 25-moving summation processing should have improved the signal-to-noise ratio of dugong calls by 5 dB, through increasing the relative amplitude of the narrow-band frequency component likely to be a dugong call and by decreasing other random components [Eqs. (2.1)–(2.3)].

$$\sum_{i=1}^{25} s_i = 25 \cdot s, \quad (2.1)$$

$$\sqrt{\sum_{i=1}^{25} n_i^2} = \sqrt{25} \cdot n, \quad (2.2)$$

$$\mathrm{SNR}' = \frac{25 \cdot s}{25 \cdot n} = \sqrt{25} \cdot \frac{s}{n} = 5 \cdot \mathrm{SNR}', \quad (2.3)$$

where s and n is the signal and the noise component, respectively, and SNR' is the improved signal-to-noise ratio (SNR) after the summation processing (Minami, 1986; Braganza and Prabhanada, 1995).

(4) The frequency A_m of the maximum spectrum density level within each S_m was determined.
(5) The spectrum density levels of A_m over a detection threshold level were used for the analyses. Comparing the values of successive A_m and A_{m+1}, A_{m+1} was filtered to be within ±100 Hz of A_m. If a successive series of $\{A_m\}$ that satisfied the above conditions lasted for more than the minimum allowable duration, the series was considered to be the dugong's biological sound; the frequency modulation of dugong chirp-squeaks and trills are reported to be gentle and to change little between neighboring 1.8-ms time windows. The detection threshold for the spectrum density level and the minimum allowable duration were altered to evaluate the detection probability.

D. Manual reduction of ship noise and remaining biologic noise

Despite the low false detection rate of automatic detection, broadband noise from fishing boats and pulse noise thought to be created by snapping shrimp masked dugong calls. In all cases, large-level masking resulted in an unnatural number of detected calls, e.g., more than 500 calls per 5-min interval. Doubtful detections, indicated by more than 10 calls per 5-min interval, were replayed to determine whether the detections were noise. Thus, ship noise and shrimp noise were manually removed from further analyses.

E. Bearing calculation and accuracy evaluation

Sound source bearings were calculated based on the time differences of sounds at stereo hydrophones obtained by the cross-correlation between the two channels [Eq. (3)]:

$$\theta = \cos^{-1}\left(\frac{c \times t}{d}\right), \quad (3)$$

where θ is the bearing of the sound source, c is underwater sound speed, t is the arrival time difference, and d is the distance between the stereo hydrophones.

F. Tide level

Tide levels were taken from the annual tide tables for Thai waters, which contained hourly recorded heights of water predicted in meters above the lowest low water. These data were published by the Hydrographic Department of the Royal Thai Navy.

III. RESULTS

A. Detection rate and bearing accuracy

Over 100 GB of underwater sound data were obtained in the recording experiment. To assess the detection probability, automatically detected calls and human detected calls were compared, and different sets of detection parameters, such as the spectrum density level duration, were tested. In total, 355 dugong calls were detected by a human analyst within 6 min from 2116 to 2122 h on 2 March. The best overall detection performance was provided by the parameter set specified in Table I.

The acoustical filter, with the filtering parameters described in Table I, detected 172 possible dugong calls out of the sound dataset identical to the one heard by the human analyst. The filter correctly detected 36.1% (= 128/77355 calls) of the human-detected calls and 2.9% (= 5/172 detections) of the filtered detections were the false alarms. There were 18 calls with a relatively long duration [avg. 1266.4±814.0 ms (SD)] that were divided into some pieces and counted as a series of some shorter-duration calls (see the example in Fig. 4). In total, 39 detections were confirmed to be such split parts of a call. The low detection rate was partly because of the high-spectrum level threshold used to prevent unreliable detection in/low signal-to-noise ratio situations, for if we exclude the data less than 100 dB at the dominant frequency, the detection rate increased to be about

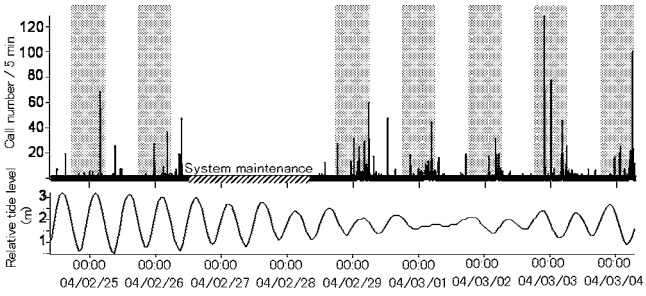


FIG. 5. A time series of change in the number of calls per 5-min interval (the upper) and the relative tide levels (the lower). The shaded bars indicate the nighttime from 1800 to 0600.

72.6% (= 98/135 calls). All of the underwater sound datasets were analyzed with the detection parameters found in Table I. Bearing calibration results indicated an average error in degrees of 9.43±9.61° (SD; n=117).

B. Frequency and bearing analyses

In total, 11 836 possible calls were detected during the 164 h recording trial from 1000 h on 24 February 2004 to 1000 h on 26 February 2004 (48 h; the first period) and from 1000 h on 28 February 2004 to 0600 h on 4 March 2004 (116 h; the latter period). We did not record underwater sound on 26 and 27 February because of system maintenance. Of the total detections, 3453 were defined as dugong calls, while 8383 ship and biologic noise components were removed from the analyses. A total of 664 calls (13.8 calls h^{-1}) were detected during the first period and 2789 (24.0 calls h^{-1}) occurred during the latter period. Figure 5 shows the time series change in the number of dugong calls per 5-min interval. Time series change of the relative tide level is also shown in the upper inset of Fig. 5. The recording period coincided with the last half of a spring tide and the first quarter of a neap tide. The tidal level variation was significantly greater in the first period than in the latter period (t test, $P<0.01$; 2.55±0.25 m, n=4, and 1.04±0.53 m, n=12, respectively). The tidal cycle was 12 h for the first period and 13 h for the latter.

In the first period, the autocorrelation of the dugong vocalization showed periodic cycles of 5.25 and 24.25 h (Fig. 6, upper), whereas in the latter period the vocalization rate showed a 25.58-h cycle (Fig. 6, lower).

Hourly variation in the number of detected calls and ship noise component is shown in Fig. 7. We divided the time of day into 8 periods, 0000–0300, 0300–0600, 0600–0900, 0900–1200, 1200–1500, 1500–1800, 1800–2100, and 2100–2400 h, and compared the average numbers of calls per period. Vocalizations were most frequent from 0300 to 0600 h, both in the first and the latter period (Scheffe's pairwise comparison test, $P<0.01$), accounting for 34.03 (226/664) and 47.40% (1322/2789) of the total calls per period, respectively.

Although the periodic cycle of vocalization rate was nearly identical, dugong vocalization was infrequent during the first period relative to the latter period. Figure 8 shows the distribution of the sound source direction from the AUSOMS-D, with 0° and 180° corresponding to the offshore

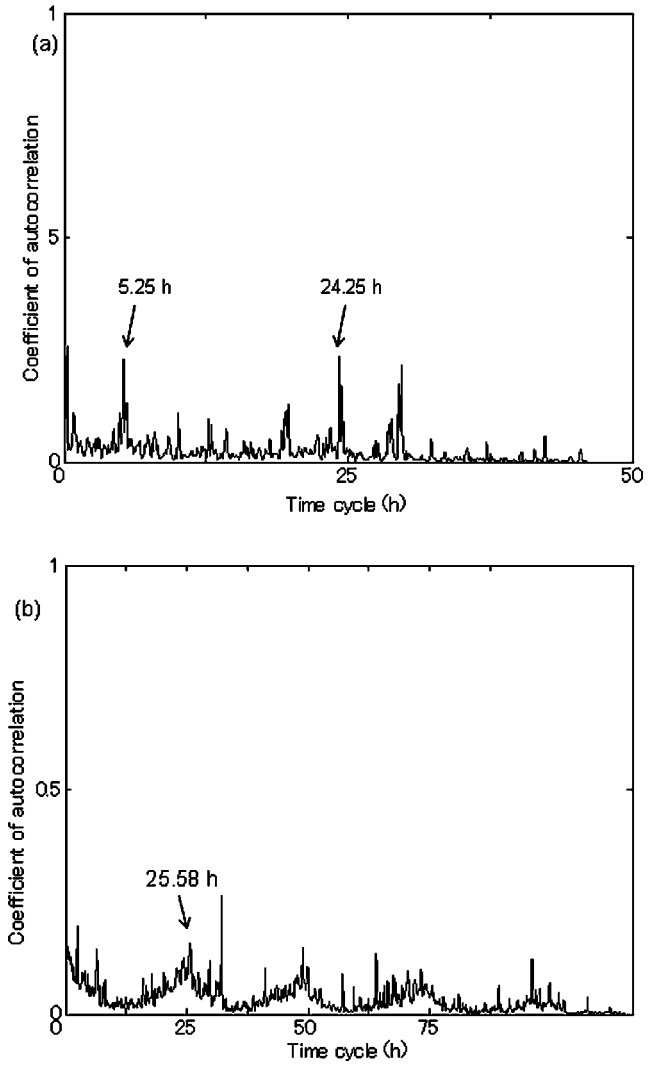


FIG. 6. Autocorrelation of the dugong vocalization rate. The number of calls 5-min interval showed about 24-h time cycle in both periods.

and inshore directions from the AUSOMS-D. Time 0 was the lowest (a), (d) or highest tide (b), (e), or the accumulated data from other time periods (c), (f). In the first period, during the low tide hours, one hour before and after the lowest tide, about 60.8% of dugong calls were vocalized from the inshore direction [the source directions were less than 90° from the AUSOMS-D; Fig. 8(a)], whereas about 55.6% and

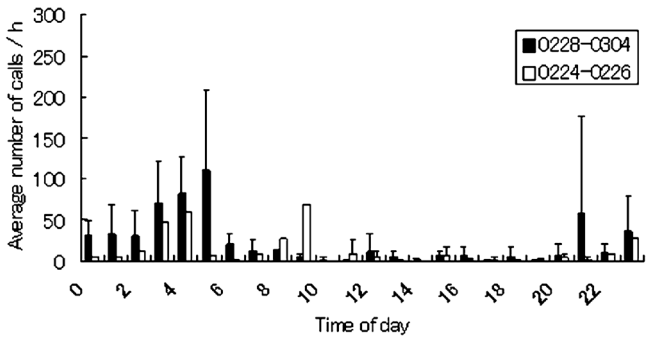


FIG. 7. Histogram of the average number of calls per hour. The horizontal axis represents time of day and the vertical axis shows the average number of calls.

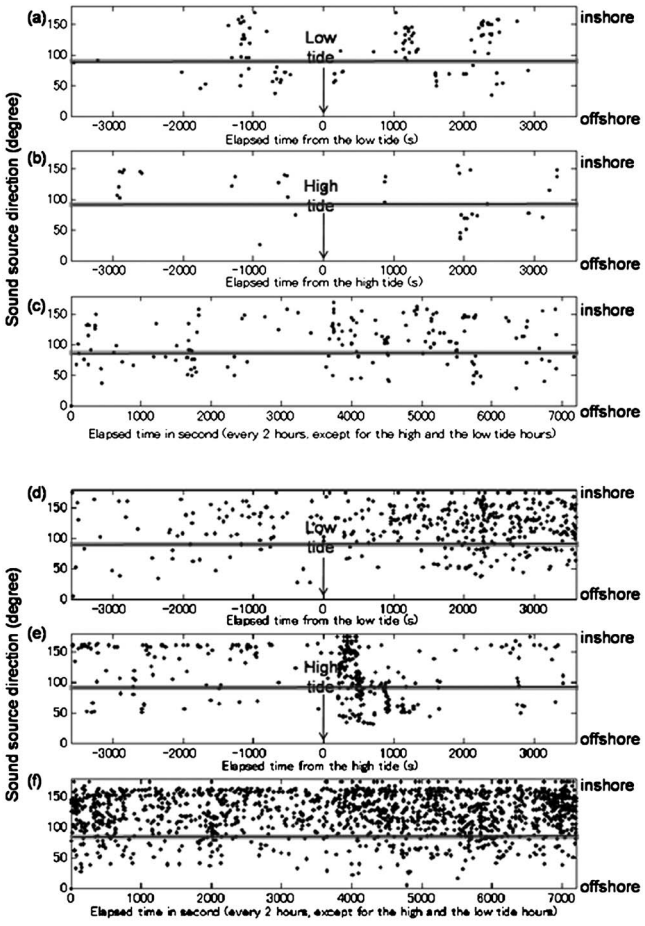


FIG. 8. Distribution of the sound source directions of dugong calls observed 1 h before and after the low tide (a),(d) and the high tide hours (b),(e), and the accumulated data from other time periods (c),(f). Lightly shaded bars in each of (a)–(f) was placed at 90°, with the width indicating a mean bearing error of about 9.43°.

68.8% were vocalized from the inshore direction during the high tide hours [Fig. 8(b)] and other periods [Fig. 8(c)], respectively. However, in the latter period, about 81.2%, 69.3%, and 84.6% of calls were vocalized from the inshore direction during low tide hours [Fig. 8(d)], high tides [Fig. 8(e)], and during other time periods [Fig. 8(f)], respectively.

Series of detected calls occurring within intervals shorter than 1 s were extracted. The variance in degrees in the sound source directions within each extracted series was then calculated and considered an index for the number of simultaneously existing dugong individuals, i.e., an "acoustical snapshot," hereafter referred to as the momentary population index or MPI. If the series were composed of calls emitted by more than one individual, the MPI would be greater. There were 51 series in the first period and 308 in the latter, with an average MPI of 35.4±20.91 (SE) and 157.08±36.48 (SE), respectively, a significant difference (Welch's t test, $P<0.01$), indicating that more individuals aggregated in the monitored area during the latter period than in the first period.

IV. DISCUSSION

In this study we provided the first long-term recording of dugong vocalizations using stationary acoustical monitoring systems. The monitoring period covered 10 days, or 164 effective monitoring hours (48 and 116 consecutive hours). The dugong vocalization detection performance of the AUSOMS-D was confirmed. The detection and the false alarm rates were 36.1% and 2.9%, respectively. Niezrecki *et al.* (2003) had applied automatic detection to manatee calls, which are reported to have similar acoustical characteristics to those of dugongs (Anderson, 1995), and concluded that the signal strength of a manatee vocalization and the sound level of the background noise were the two key factors affecting the performance of any manatee warning system. Both factors are directly related to the signal-to-noise ratio of a manatee signal (Niezrecki *et al.*, 2003). Likewise, the detection of dugong vocalizations would be affected by the signal-to-noise ratio of a dugong call. The next logical extension of the automatic detection system was to investigate dugong sound levels as a function of distance from the source. The average error of the bearing calculations resulted in 9.43±9.61° (SD, $n=117$).

The time cycle of the vocalization frequency itself was likely driven by circatidal rhythm or circadian rhythm, or both, because the autocorrelation of the vocalization rates showed 24.5 and 25.6-h cycles through each period (Fig. 6). However, there was no drift that exactly followed a 25-h cycle, as shown in Fig. 7. If the attendance cycle of vocalizing dugongs in the monitored area were only driven by the tidal shift, vocalization peaks should have occurred, even in daytime, during the spring and neap tide periods. These results suggest that the periodic cycle of dugong vocalization frequency was driven, not only by the tidal cycle, but also by the diel cycle of about 25 and 24 h. The number of detected calls also varied between the two monitoring periods; many more calls were detected in the latter period, when the tidal range was much less than in first period. The dugong vocalization source bearings indicated more frequent use of inshore areas (from the AUSOMS-D) than offshore areas. A significant difference was found between the first and latter periods in the average percentage of calls vocalized from the inshore direction (Welch's t test, $n=3$ for each period, $P=0.027$). Dugongs vocalized more frequently inshore of AUSOMS-D in the latter period, during the neap tide [Figs. 8(a)–8(f)]. The variance of the sound source direction within call series of 1-s interval (MPI) suggested that group aggregation occurred in the latter period. The small tidal range would have facilitated the aggregation within a monitored area.

Several visual surveys revealed the evidence of dugongs' feeding in intertidal flats that were physically inaccessible during low tides (e.g., Adulyanukosol, 1999; Mukai *et al.*, 1999; Nakoaka and Aioi, 1999; Tsutsumi *et al.*, submitted). Thus, in shallow areas, they would been under strong time constraints regulated by the tidal range. Many studies have suggested a correlation between dugong activity and tidal shift. For example, Anderson and Birtles (1978) reported that dugong feeding at Shoalwater Bay (Queensland, Australia) was strongly regulated by tidal patterns. They also reported that diurnal inshore feeding was normal, even without frequent anthropogenic disturbance, and that the restriction to nocturnal feeding was a response to human

disturbance rather than a normal pattern. Hines (2002) performed several daytime aerial surveys around Talibong Island and reported that more individuals were found around the hours of high tide than at low tide. Nocturnal feeding behavior was observed by Dung (2003) in Vietnamese waters and Mukai *et al.* (1999) in Thai waters. Anderson (1986) observed dugong feeding behavior throughout the day, and at night whenever conditions allowed.

The visual sightings during the daytime seems to conflict with the acoustic recordings during the nighttime. Here arise two potential interpretations for this confliction. One is that the vocalization rates within each individual increased. The other is that there were much more individuals staying around the monitored area. In both cases, it is well ascertained fact that there had been more than one dugongs making calls for over a long hours in the dark. In such situations, sound signals are good medium for information transmission. If the dugong calls were used as a tool for interindividual information exchange, it is very important to investigate the functions of the different call types. Long-term observation, improvement in the accuracy of the detection software, and additional deployment of the recording systems would be needed to reveal the function of the dugong calls.

V. CONCLUSIONS

In summary, the dugongs in the waters to the south of Talibong Island, Thailand, visited the inshore areas near the seagrass primarily during the night. The attendance of the animals showed a 24- to 25-h periodicity. More dugongs utilized the inshore side of the monitored are during the neap tide [MPI and Figs. 8(a)–8(f)]. Long-term, highly accurate monitoring and additional deployment of the AUSOMS-D would greatly benefit the understanding of dugong ecology, in matters such as the interindividual contacts made by vocal exchanges.

Assuming that the above hypothesis is correct, effective measures to prevent dugong bycatch would include regulating boat operation in the very early morning hours to avoid collisions, closing the mouth of set nets from midnight to early morning or checking the set nets early in the morning to rescue animals in case of bycatch, and operating a warning system that would alert anglers to the presence of dugongs.

ACKNOWLEDGMENTS

We thank the National Research Council of Thailand, the members of the Phuket Marine Biological Center, METOCEAN CO., LTD., System Intech Co., Ltd., the National Research Institute of Fisheries Engineering, the Japan Fisheries Resource Conservation Association, and Biosphere informatics for support and cooperation. This study was funded in part by the cabinet office, a Grant-in-Aid for Scientific Research No. (16255011 and No. 1665808), the Information Research Center for Development of Knowledge Society Infrastructure, the Ministry of Education, Culture, Sports, Science and Technology, and the Bio-oriented Technology Research Advancement Institution (BRAIN) Promotion of Basic Research Activities for Innovative Biosciences of Japan.

Adulyanukosol, K. (**1999**). "Dugong, dolphin and whale in Thai waters," *Proceedings of the 1st Korea-Thailand Joint Workshop on Comparison of Coastal Environment*, Korea-, Thailand, pp. 5–15.

Akamatsu, T., Wang, D., Wang, K., and Wei, Z. (**2001**). "Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China," J. Acoust. Soc. Am. **109**, 1723–1727.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005**). "Biosonar behaviour of free-ranging porpoises," Proc. R. Soc. London, Ser. B, **272**, 797–801.

Anderson, P. K. (**1986**). "Dugongs of Shark Bay, Australia–Seasonal Migration, Water Temperature, and Forage," Natl. Geogr. **2**, 473–490.

Anderson, P. K. (**1995a**). "Scarring and photo identification of dugongs (Dugong dugon) in Shark Bay, Western Australia," Aquatic Mammals **21**, pp. 205–211.

Anderson, P. K., and Birtles, A. (**1978**). "Behavior and ecology of the dugong, dugong dugon (Sirenia): Observations in shoalwater and Cleveland Bays, Queensland," Aust. Wildl. Res. **5**, 1–23.

Anderson, P. K., and Barclay, R. M. R. (**1995b**). "Acoustic signals of solitary dugongs: Physical characteristics and behavioral correlates," J. Mammal. **76**, pp. 1226–1237.

Braganza, O, N. P., and Prabhanada, B. S. (**1995**). "Filtering noise from slowly varying signals taking advantage of a fast ADC in data acquisition," Meas. Sci. Technol. **6**, 329–331.

Chilvers, B. L., Delean, S., Gales, N. J., Holley, D. K., Lawler, I. R., Marsh, H., Preen, A. R. (**2004**). "Diving behavior of dugongs, Dugong dugon," J. Exp. Mar. Biol. Ecol. **304**, 203–224.

Dung, P. H. (**2003**). "The primary assessment on the Dugong population in Viet Nam," *Proceedings of the 4th SEASTAR 2000 Workshop*. pp. 64–71.

Hines, E. (**2002**). "Conservation of the Dugong (Dugong dugon) along the Andaman Coast of Thailand: An example of the integration of conservation and biology in endangered species research," Ph.D. dissertation, Department of Geography, University of Victoria, Victoria, BC Canada.

Hines, E., Adulyanukosol, K., and Duffus, D., A. (**2005**). "Dugong (Dugong Dugon) Abundance along the Andaman Coast of Thailand," Marine Mammal Sci. **21**, 536–549.

Ichikawa, K., Akamatsu, T., Shinke, T., Arai, N., Hara, T., and Adulyanukosol, K. (**2003**). "Acoustical analyses on the calls of dugong," *Proceedings of the 4th SEASTAR 2000 Workshop*, pp. 72–76.

Marsh, H., and Sinclair, D. F. (**1989**). "An experimental evaluation of Dugong and sea turtle aerial survey techniques," Aust. Wildl. Res. **16**, 639–650.

Marsh, H., Eros, C., Corkeron, P., and Breen, B. (**1999**). "A conservation strategy for dugongs: implications of Australia research," Z. Gerontol. Geriatr. **50**, 979–990.

Marsh, H., Penrose, H., Eros, C., and Hugues, J. (**2002**). "Dugong status report and action plans for countries and territories," UNEP Early Warning and Assessment Report Series Vol. 1, p. 162.

Minami, S. (**1986**). *Kagakukeisoku No Tameno Hakei De-Tashori* (CQ Publishing Co. Ltd., City), pp. 105–106 (in Japanese).

Mukai, H., Aioi, K., Lewmanomont, K., Matsumasa, M., Nakaoka, M., Nojima, S., Supanwanid, C., Suzuki, T., and Toyohara, T. (**1999**). "Dugong grazing on Halophila beds in Haad Chao Mai National Park, Trang Province, Thailand—How many dugongs can survive?," Effects of grazing and disturbance by dugongs and turtles on tropical seagrass ecosystems, pp. 239–254.

Nakaoka, M., and Aioi, K. (**1999**). "Growth of the seagrass Halophila ovalis at the dugong trails compared to existing within-patch variation in a Thailand intertidal flat," Effects of grazing and disturbance by dugongs and turtles on tropical seagrass ecosystems, pp. 255–267.

Nakanishi, Y., Hosoya, S., Nakanishi, Y., Arai, N., and Adulyanukosol, K. (**2005**). "The distribution of dugong trenches in the seagrass beds of Libong Island, Thailand," Journal of Advanced Marine Science and Technology Society **11**, 53–57 (in Japanese).

Niezrecki, C., Phillips, R., and Meyer, M. (**2003**). "Acoustic detection of manatee vocalizations," J. Acoust. Soc. Am. **114**, 1640–1647.

Preen, A. (**1998**). "Marine protected areas and dugong conservation along Australia's Indian Ocean coast," Environ. Manage. (N.Y.) **22**, 173–181.

Preen, A. (**2004**). "Distribution, abundance and conservation status of dugongs and dolphins in the southern and western Arabian Gulf," Biol. Conserv. **118**, 205–218.

Shinke, T., Shimizu, H., Ichikawa, K., Arai, N., Matsuda, A., and Akamatsu,

T. (**2004**). "Development of automatic underwater sound monitoring system version 1," *Proceedings of The 2004FY Annual Meeting of the Marine Acoustics Society of Japan*, pp. 33–36.
Tsutsumi, C., Ichikawa, K., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosol, K. "Feeding behavior of wild dugongs monitored by a passive acoustical method using an underwater recording system" J. Acoust. Soc. Am.(submitted).
Wang, K., Wang, D., Akamatsu, T., Li, S., and Xiao, J. (**2005**). "A passive acoustical monitoring method applied to observation and group size estimation of finless porpoises," J. Acoust. Soc. Am. **118**, 1180–1185.

Design of an underwater Tonpilz transducer with 2-2 mode piezocomposite materials

Yongrae Roh and Xiaoguang Lu
School of Mechanical Engineering, Kyungpook National University, 1370 Sankyukdong, Bukgu, Daegu 702-701, Korea

(Received 15 June 2005; revised 20 February 2006; accepted 23 March 2006)

The performance of a transducer is determined by the properties of constituent materials and the effects of many structural parameters. In this study, the use of 2-2 piezocomposite materials in an underwater Tonpilz transducer was investigated. Through finite element analyses, the relationship between the piezocomposite material properties and the performance of the transducer was investigated, i.e., operation frequency, bandwidth, and sound pressure. Based on the analysis result, the geometry of the Tonpilz transducer that could provide the highest sound pressure for a given electric field amplitude while satisfying requirements such as bandwidth and operation frequency was optimized. The optimization result was compared with that of a traditional piezoceramic transducer to confirm the superiority of the piezocomposite transducer; a Tonpilz transducer made of 2-2 mode piezocomposite plates could provide a higher effective coupling factor and thus a wider bandwidth at a desired operation frequency with a smaller size than a traditional piezoceramic transducer. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197788]

PACS number(s): 43.30.Yj, 43.38.Fx, 43.58.Fm [WMC] Pages: 3734–3740

I. INTRODUCTION

A Tonpilz transducer is a time-tested design of relatively low frequency and high power. It is made of a piezoceramic ring stack between a head mass and a tail mass, prestressed by a central bolt.[1,2] Traditionally, PZT has been used as the active material in the transducer. The principle of the Tonpilz transducer is that the piezoceramic inside the transducer induces an extensional displacement for high-power acoustic waves in the poling direction of the piezoceramic when exposed to electrical drive.[3,4] In general, the three most important performance parameters of the transducer are its operation frequency, bandwidth, and transmitted sound pressure level.[5] Operation frequency in this paper is defined as the frequency at which the transducer shows the maximum sound pressure, and bandwidth is the −3-dB bandwidth of the transducer. The high sound pressure level is normally achieved by the following two methods. The first is the use of a good active material in the transducer. Pure piezoelectric ceramics have been widely used for the Tonpilz transducer. As is well known, because of their high acoustic impedance and low electromechanical coupling factor, piezoceramics have a certain limitation in the achievable sound pressure level.[6] The second method is to optimize the geometry of the transducer. The performance of an acoustic transducer is influenced by many structural variables and, in general, the influences of these variables are not linearly independent of each other.[7,8] A proper design scheme should be employed to optimize the transducer geometry.

This paper is to investigate the optimal geometry of a Tonpilz transducer made of piezocomposites that can improve the performance of piezoceramics. The piezocomposite transducer was designed to transmit acoustic waves of the highest possible pressure level for a given electric field amplitude at a given operation frequency while having the bandwidth larger than a predetermined value. Piezocomposite material is an important update of the piezoceramics in terms of acoustic properties. Piezocomposites are comprised of a piezoceramic, as a filler, and a piezoelectrically inactive polymer or epoxy as a matrix of the structure. Piezocomposite materials utilize the high piezoelectricity of the ceramic and the low acoustic impedance of the compliant polymer to maximize a predefined figure of merit of the application envisaged.[9] For simple two-phase solids, there are as many as ten different patterns of connectivity that may be taken up by the phases. A series of papers by Newnham showed that the connectivity between the phases in this composite is a key feature in determining its overall properties.[10,11] In this study, we used a 2-2 mode piezocomposite material that shows low density and low acoustic impedance and that can generate a large surface displacement with a moderate driving voltage.[12] Also it can improve the bandwidth without lowering the sensitivity.[12,13] The optimal transducer geometry was determined through finite element analysis (FEA) and statistical multiple regression analysis of the FEA results. The design scheme can reflect not only the individual, but also all the cross-coupled effects of the structural variables. In comparison with conventional analytic and finite element methods, the present method can determine the detailed geometry of a transducer with great efficiency and rapidity. The performance of the designed piezocomposite transducer was compared with that of a traditional piezoceramic transducer to confirm the superiority of the piezocomposite transducer.

II. 2-2 PIEZOCOMPOSITES MODELING AND ANALYSIS

Figure 1 is the FEA model of the 2-2 mode piezocomposites. For simplicity, a 2-D model was constructed with the center being symmetric and the right side boundary being

 0001-4966/2006/119(6)/3734/7/$22.50

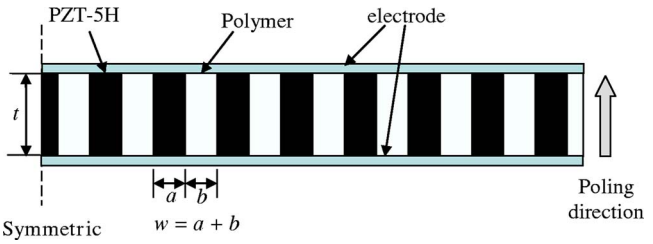


FIG. 1. (Color online) Schematic model of the 2-2 mode piezocomposites for thickness mode vibration analysis; w=0.64 mm and t=1.2 mm in this model.

free to allow free strain along the thickness.[14] The model was to analyze the thickness mode vibration of a piezocomposite plate because the thickness mode is the main strain mode utilized in a Tonpilz transducer. The modeling and analysis were carried out with a commercial FEA package, ANSYS. The piezoceramic in the composite was PZT-5H whose material constants were referred to Ref. 15. Material properties of the polymer analyzed in this paper are given in Table I. Higher electromechanical coupling factor and lower acoustic impedance have been made possible either by controlling the volume fraction of PZT or by changing the ratio of the width of the PZT and polymer ($w=a+b$ in Fig. 1) to the plate thickness (t).[16–18] In Fig. 1, w is 0.64 mm and t is 1.2 mm as sample values. With the model in Fig. 1, firstly, we investigated the effects of the volume fraction of PZT from 10% to 100% by controlling the "a" and "b" in the figure for a given aspect ratio. Electromechanical properties of the piezocomposite plates can be calculated through harmonic analyses, and Fig. 2 is the analyzed results. In the figure, the volume fraction of 1.0 corresponds to a pure PZT plate. The 2-2 piezocomposite turned out to have (a) a higher thickness mode electromechanical coupling factor in thickness, k_t, (b) a lower density, and (c) lower acoustic impedance than pure PZT. These results are agreeable to reported works.[19] According to the results in Fig. 2, the piezocomposite has the highest electromechanical coupling factor, k_t, when the volume fraction of PZT is 0.6. Next, we investigated the effects of the aspect ratio by controlling the value of w/t from 0.1 to 1.0 at the PZT volume fraction 0.6. According to the result in Fig. 3, the piezocomposite plate has the highest k_t when the aspect ration is 0.7. The sharp increment of the k_t at the aspect ratio of 0.7 is considered due to the coupling of thickness mode and lateral mode vibration of the piezocomposite plate. As the ratio of w/t increases, the coupling factor of the thickness mode gradually decreases, while the coupling factor of the lateral mode gradually increases due to the modes coupling.[16]

TABLE I. Material properties of the polymer composing the piezocomposites.

Density	Young's modulus	Permittivity	Rayleigh stiffness damping
ρ (kg/m^3)	E (10^9 N/m^2)	ε (10^{-11}C/m)	β (10^{-10} s)
1097	3.5	3.44	70.0

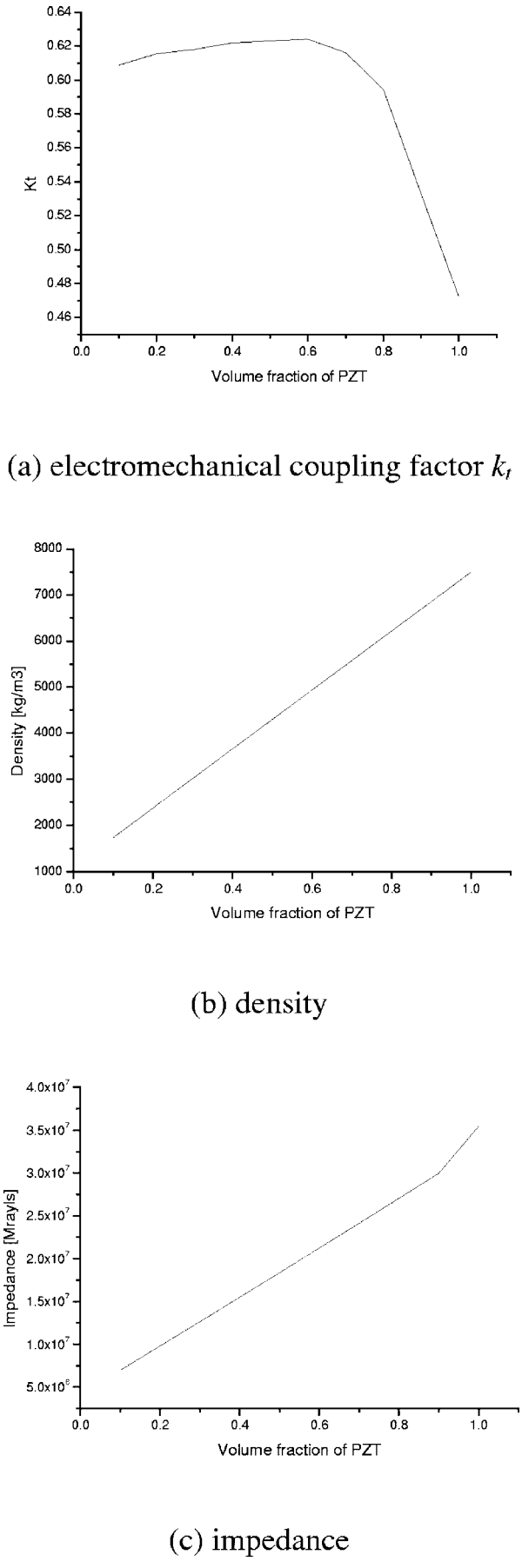


FIG. 2. Variation of the electromechanical coupling factor k_t, density, and impedance of the piezocomposite in relation to the volume fraction of PZT.

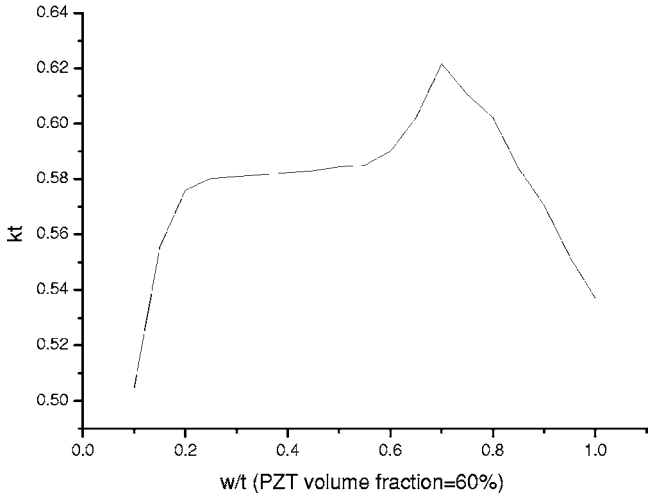


FIG. 3. The relation between coupling factor k_t and w/t at PZT volume fraction of 0.6.

Following the results of all the harmonic analyses, the optimal geometry of the 2-2 mode piezocomposite plate was determined to have the PZT volume fraction of 0.6 and the aspect ratio of 0.7.

III. TONPILZ TRANSDUCER MODELING AND ANALYSIS

The 2-2 mode piezocomposite designed in Sec. II was installed into a two-dimensional finite element model of a Tonpilz transducer. Figure 4 is the FEA model constructed to analyze the underwater performance of the transducer. The 2-D model is axisymmetric along the central axis of the Tonpilz transducer, hence 3-D in effect. To prevent reflection of an acoustic wave by fluid medium, no reflection boundary conditions were applied to all the outside fluid boundaries.

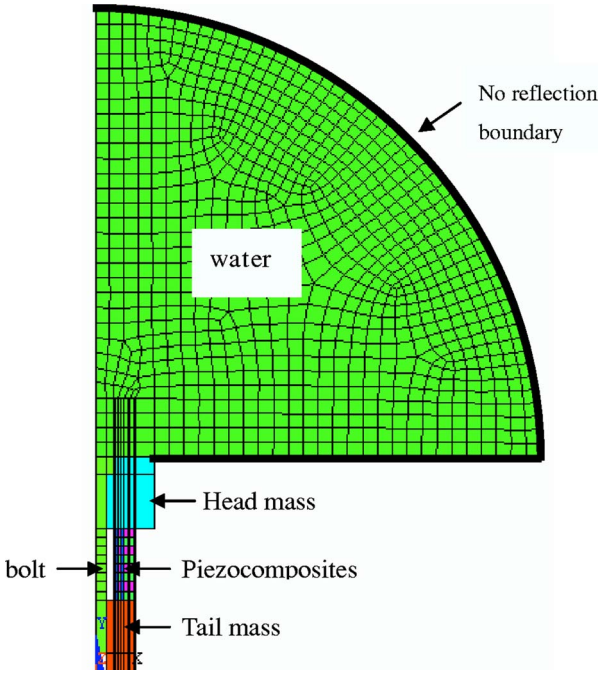


FIG. 4. Finite element of the underwater Tonpilz transducer made of 2-2 mode piezocomposites.

TABLE II. Basic geometry of the Tonpilz transducer.

	Thickness (mm)	Innerradius (mm)	Outerradius (mm)
Piezocomposite stack	2.0×8 layers	4.0	9.0
Tail mass	16.0	2.5	9.0
Head mass	16.0	2.5	13.2
Bolt	44.0 (length)		2.5

The transducer is composed of a head mass, a tail mass, a bolt, and a stack of eight piezocomposite plates. The basic geometry of the transducer was arbitrarily set as shown in Table II. Table III shows properties of the materials composing the transducer. The process to determine the optimal geometry of the transducer was conducted following the routine in Fig. 5. Through preliminary calculations and literature survey, the most influential structural variables were established: thickness of 2-2 piezocomposite plate, thickness of the tail mass, thickness of the head mass, and radius of the head mass. The performance factors analyzed are the sound pressure, the operation frequency, and the −3-dB bandwidth of the transducer. The goal of this design was to maximize the sound pressure of the transducer for a given electric field amplitude while achieving a desired operation frequency and −3-dB bandwidth.

Through harmonic analyses, we investigated the change of the performance factors of the transducer in relation to the variation of the major structural variables. The variation ranges of the structural variables were arbitrarily set wide enough (+50% and −50% of the initial values) to include all the possible modifications of the variables for practical applications as shown in Table IV. We commenced with a model having the basic geometry as in Table II and the 2-2 mode piezocomposite designed in Sec. II. Structural variables were varied one by one over the range indicated in Table IV while all other variables were fixed to the basic geometry, which made 17 cases in total. Figure 6 is a data sample of the harmonic analysis that shows the variation of the transducer performance in relation to the change of thickness of the 2-2 piezocomposites. We can determine the operation frequency and the maximum sound pressure for each thickness of the piezocomposites. Electrical voltage applied to each piezocomposite plate was arbitrarily set to be 40 V/mm all through the FEA. The absolute value of the sound pressure in Fig. 6 has no practical meaning, but is good for comparison only. Results of the analyses of Fig. 6 are summarized in Fig. 7. An increase of the thickness of the

TABLE III. Material properties of the transducer.

	Material	Young's modulus (GPa)	Density (kg/m^3)	Poisson's ratio
Tail mass	Steel	206	7,955	0.30
Head mass	Aluminum	70.3	2,770	0.33
Bolt	Stainless steel	185	7,920	0.3

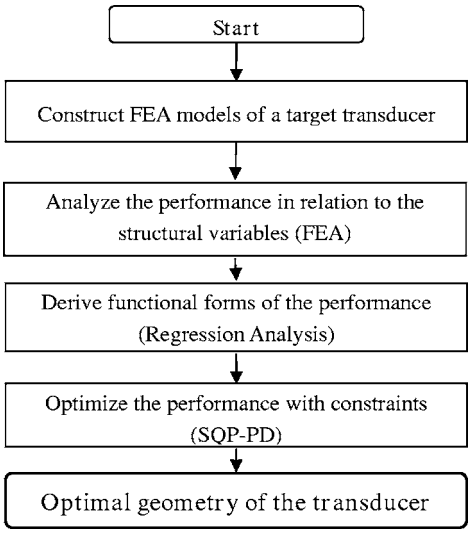


FIG. 5. (Color online) Optimal design algorithm for the transducer.

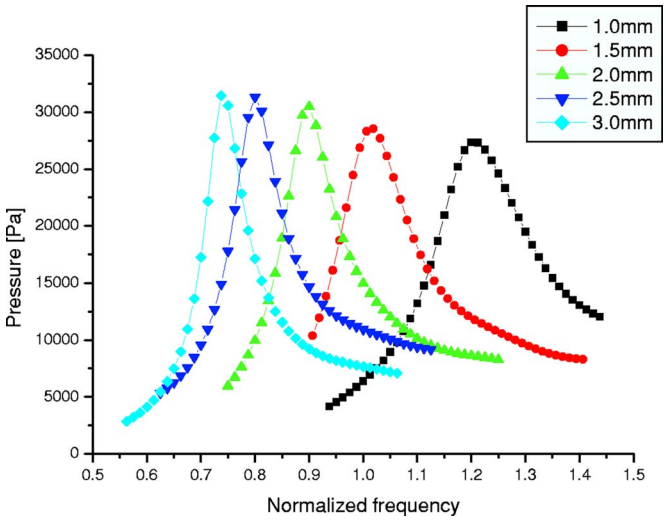


FIG. 6. (Color online) FE harmonic analysis result showing the relation between the transducer performance and the thickness of the piezocomposite plate.

piezocomposite plate causes the decrease in the operation frequency and the increase in the maximum sound pressure of the transducer.

Similar analyses were conducted to find the trends of the operation frequency and the maximum sound pressure change in relation to the other structural variables. Figure 8 shows the influence of the tail mass thickness on the operation frequency and maximum sound pressure. Both performance factors were observed to decrease in proportion to the thickness of the tail mass. According to Fig. 9, both performance factors were observed to decrease in proportion to the thickness of the head mass. In Fig. 10, the operation frequency decreased consistently while the maximum sound pressure showed a peak at a certain value of the head mass radius.

Some of the variables showed consistent influence on the three performance factors, either positive or negative, while some others showed contradictory influences. In addition, the effects of the variables were not independent of, but cross-coupled with, each other. The purpose of this study was to determine the optimal geometry of the piezocomposite Tonpilz transducer that could provide the highest sound pressure for a given electric field amplitude at a desired operation frequency f_0, while the −3-dB bandwidth is wide enough to meet the given requirement. Hence, there should be a clever way to trade off the effects of the structural variables to achieve the optimal performance, thus we employed the statistical multiple regression analysis technique and the optimization technique for the purpose.

TABLE IV. Variation range of the structural variables of the Tonpilz transducer.

Structural variables	Variation range (mm)
thickness of each 2-2 piezocomposite plate	1.0–3.0
thickness of the tail mass	8.0–24.0
thickness of the head mass	8.0–24.0
radius of the head mass	9.0–17.4

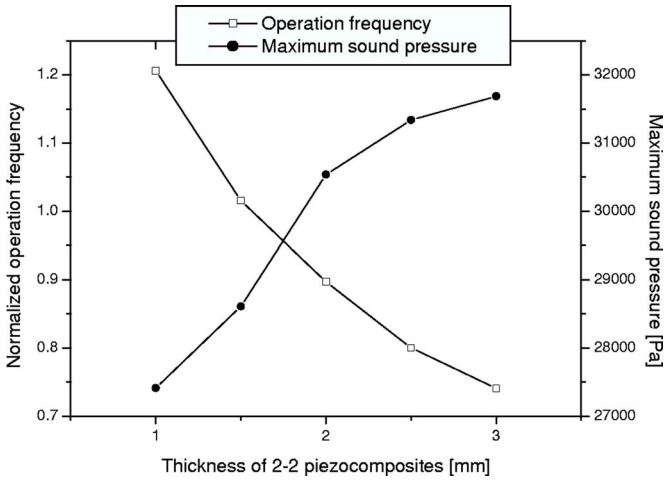


FIG. 7. Normalized operation frequency and maximum sound pressure versus thickness of the piezocomposite plate.

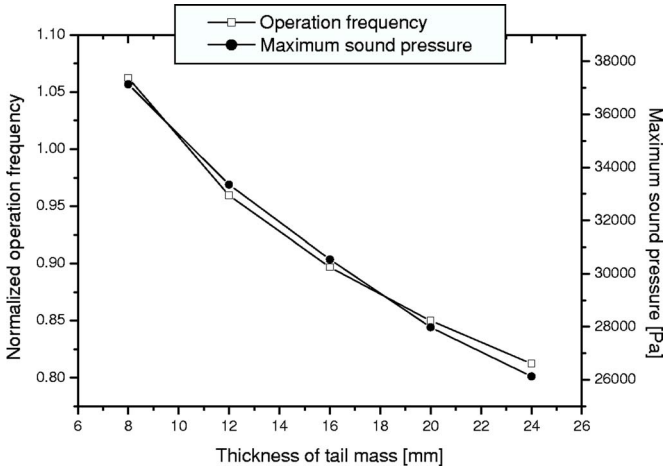


FIG. 8. Normalized operation frequency and maximum sound pressure versus thickness of the tail mass.

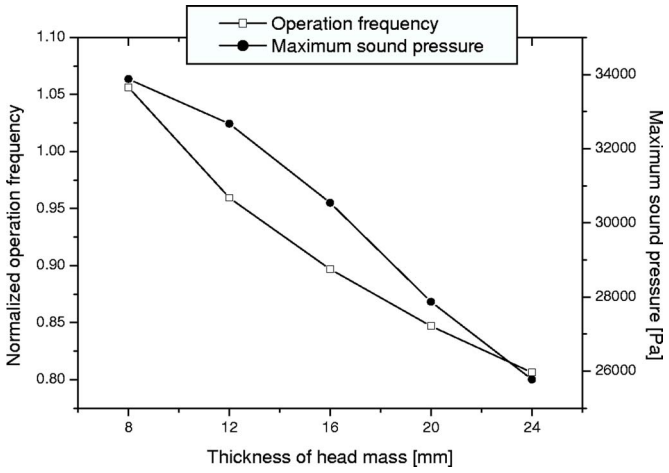


FIG. 9. Normalized operation frequency and maximum sound pressure versus thickness of the head mass.

IV. OPTIMIZATION OF THE TONPILZ TRANSDUCER STRUCTURE

The statistical multiple regression analysis method validated in our previous work[20] was employed for optimizing the transducer geometry in this study.

In Sec. III, we analyzed 17 different combinations of the structural variables to find the variation trends of the performance of the transducer. However, 17 cases were not enough to compose the transducer's performance as a function of the structural variables. We used the design method of 2^n experiments, where n is the number of structural variables under consideration. Table V shows the normalization factors established for each of the structural variables in this study. The normalization factors 1 and −1 correspond to the maximum and the minimum values of each of the structural variables, respectively. The normalization factor of 0 corresponds to the basic geometry of the transducer. The combinations of the variables corresponding to the cases are shown in Table V numbered $2^4=16$. However, through preliminary calculations, we found that those 16 sets of data were insufficient for accurate formulation of the transducer performance. So, in addition to those extreme combinations, we analyzed the intermediate cases that corresponded to the

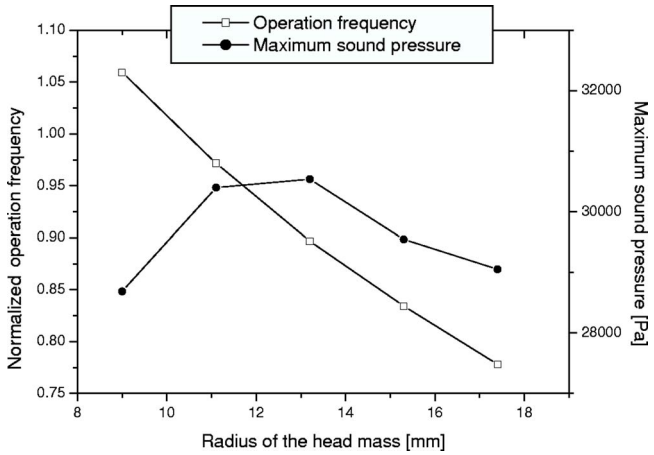


FIG. 10. Normalized operation frequency and maximum sound pressure versus radius of the head mass.

TABLE V. Normalization factors of the structural variables.

Normalization factors	Piezocomposite stack thickness (mm)	Tailmass thickness (mm)	Head mass thickness (mm)	Head mass radius (mm)
−1	8.0	8.0	8.0	9.0
−0.5	12.0	12.0	12.0	11.1
0	16.0	16.0	16.0	13.2
0.5	20.0	20.0	20.0	15.3
1	24.0	24.0	24.0	17.4

normalization factors of ±0.5 for each of the structural variables. The number of additional combinations of variables with normalization factors of ±0.5 is also $2^4=16$. In total, we did the analysis for 49 (17+16+16) different combinations of the structural variables.

For statistical analysis of the FEA results to formulate the transducer's performance, we developed a second-order multiple regression model that could consider the cross-coupled effects of the structural variables by product terms. A second-order regression model took the form of Eq. (1) where we had to find proper regression coefficients for the function to coincide with the analysis data:

$$y = a_1x_1^2 + \cdots + a_4x_4^2 + b_1x_1 + \cdots + b_4x_4 + c_1x_1x_2 + \cdots + c_6x_3x_4 + d_1x_1x_2x_3 + \cdots + d_4x_2x_3x_4 + e_1x_1x_2x_3x_4 + e_2, \quad (1)$$

where, a_1-e_2 are the 20 regression coefficients, and x_1-x_4 are the four structural variables in Table V. The statistical analysis was made using the commercial software SAS. Through the analysis, the regression equation for the operation frequency f_r, the sound pressure p_r, and the −3-dB bandwidth Δf were determined as Eqs. (2)–(4), respectively,

$$f_r = 2203.4x_1^2 + 1097.5x_2^2 + 897.5x_3^2 + 521.1x_4^2 - 7330.7x_1 - 3862.9x_2 - 3904.8x_3 - 4972.6x_4 + 1161.9x_1x_2 + 1479.4x_1x_3 + 1724.7x_1x_4 + 178.4x_2x_3 + 93.8x_2x_4 + 1945.4x_3x_4 - 71.1x_1x_2x_3 - 32.8x_1x_2x_4 - 1483.5x_1x_3x_4 - 332.4x_2x_3x_4 + 330.7x_1x_2x_3x_4 + 28\,684.7, \quad (2)$$

$$p_r = -1021.9x_1^2 + 1337.3x_2^2 - 457.9x_3^2 - 1357.5x_4^2 + 1386.8x_1 - 5088.5x_2 - 4295.96x_3 - 384.8x_4 + 702.3x_1x_2 + 181.7x_1x_3 + 928.3x_1x_4 + 639.8x_2x_3 - 830.5x_2x_4 + 3184.0x_3x_4 - 385.7x_1x_2x_3 + 317.7x_1x_2x_4 - 1114.3x_1x_3x_4 - 845.7x_2x_3x_4 + 100.5x_1x_2x_3x_4 + 30\,129.3, \quad (3)$$

$$\Delta f = 440.1x_1^2 + 16.5x_2^2 + 757.7x_3^2 - 458.0x_4^2 - 1395.2x_1 - 327.4x_2 - 1382.3x_3 + 223.6x_4 - 91.8x_1x_2 + 625.8x_1x_3 + 310.3x_1x_4 - 23.7x_2x_3 + 213.4x_2x_4$$

TABLE VI. Result of the optimization of the piezocomposite transducer.

Structural variables and performance	ANSYS validation
Each piezocomposite plate thickness (x_1)	2.50 mm
Tail mass thickness (x_2)	8.00 mm
Head mass thickness (x_3)	12.00 mm
Head mass radius (x_4)	13.20 mm
Normalized operation frequency	f_0
Sound pressure per unit electric voltage	36.2 Pa/V
Normalized −3-dB bandwidth	13.3%

TABLE VII. Result of the optimization of the PZT transducer.

Structural variables and performance	ANSYS validation
Each PZT plate thickness (x_1)	3.00 mm
Tail mass thickness (x_2)	8.00 mm
Head mass thickness (x_3)	17.28 mm
Head mass radius (x_4)	13.2 mm
Normalized operation frequency	f_0
Sound pressure per unit electric voltage	28.8 Pa/V
Normalized −3-dB bandwidth	11.2%

$$- 157.7x_3x_4 + 347.5x_1x_2x_3 - 426.2x_1x_2x_4 - 325.4x_1x_3x_4 - 292.1x_2x_3x_4 + 395.5x_1x_2x_3x_4 + 3055.6. \quad (4)$$

The coefficient of determination of the formulated functions describes the accuracy of the determined regression model in representing the data to be fitted. For Eqs. (2)–(4), the coefficients of determination were found to be 0.998, 0.975, and 0.978, respectively, which confirmed that the equations well described the variation of the transducer performance in relation to the structural variables.

The purpose of this study was to find the geometry of the transducer that could provide the highest sound pressure for a given electric field amplitude at a desired operation frequency f_0 with a bandwidth larger than a given requirement. Equation (5) is a target function to achieve the highest sound pressure for a given electric field and Eq. (6) is corresponding constraint conditions. The optimization was to find the combination of the structural variables that could minimize the target function while satisfying the two constraint equations. The minimization process was carried out with the SQP-PD method.[20]

Target function

$$T(x_1,x_2,x_3,x_4) = \left|\frac{1}{p_r}\right| \quad (5)$$

is subject to

$$f_r = f_0, \quad (6)$$

$$\Delta f \geqslant 10.9\,\% .$$

Results of the optimal design are summarized in Table VI. The dimensions in Table VI indicate the transducer geometry that can provide the highest sound pressure level for a given electric field amplitude while satisfying the operation frequency and the bandwidth requirements. The designed transducer showed the maximum sound pressure of 36.2 Pa/V at f_0 and had a bandwidth of 13.3% larger than the requirement.

For comparison, the same procedure was applied to the optimal design of a piezoceramic Tonpilz transducer. The piezoceramic transducer had pure PZT-5H as the active material but the other structures were the same as the piezocomposite transducer. Variation ranges of the structural variables were as those in Table IV. Results of the design are in Table VII. The piezoceramic transducer was a little longer than the piezocomposite transducer in Table VI to achieve the operation frequency f_0. However, the maximum sound pressure and the bandwidth were smaller and narrower than those of the piezocomposite transducer. The transmitting voltage response (TVR) spectra of the two transducers compared in Fig. 11 manifest the difference. These results illustrate the superiority of the piezocomposite transducer in terms of the acoustic impedance matching and the bandwidth. The piezocomposite transducer should have a lower source power due to a smaller volume of the inherent piezoelectric material. The higher TVR of the piezocomposite transducer observed nevertheless in Fig. 11 is due to lower acoustic impedance and a higher effective coupling factor of the piezocomposite material as shown in Fig. 2. Therefore the Tonpilz transducer made of 2-2 mode piezocomposite plates was concluded to provide a higher effective coupling factor and thus a wider bandwidth at a desired operation frequency with a smaller size than a traditional piezoceramic transducer.

V. SUMMARY

The performance of an acoustic transducer is determined by both the use of various active materials and the effects of many structural variables. In this study, we investigated the use of 2-2 mode piezocomposite materials in an underwater Tonpilz transducer. Through finite element analyses, we investigated the relationship between the piezocomposite material properties and the performance of the transducer, i.e., operation frequency, bandwidth, and sound pressure. Based

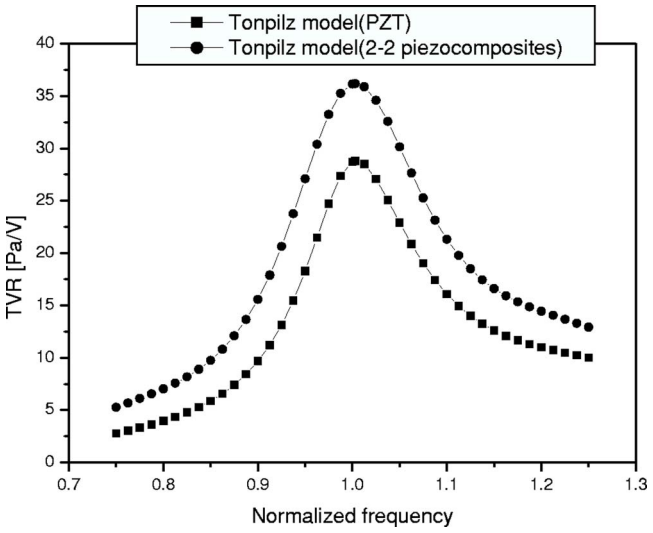


FIG. 11. Transmitting voltage responses in linear units of the Tonpilz transducer designed with PZT and 2-2 mode piezocomposites.

on the result, we optimized the geometry of the Tonpilz transducer that could provide the highest sound pressure for a given electric field amplitude while satisfying bandwidth and operation frequency constraints. The optimization result was compared with that of a traditional piezoceramic transducer to confirm the superiority of the piezocomposite transducer. The designed transducer made of 2-2 mode piezocomposite plates turned out to provide a higher effective coupling factor and thus a wider bandwidth at a desired operation frequency, with smaller size, than a traditional piezoceramic transducer.

[1]Q. S. Yao and L. Bjørnø, "Broad Tonpilz underwater acoustic transducers based on multimode optimization," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1060–1066 (1997).

[2]C. Desilets, G. Wojcik, L. Nikodym, and K. Mesterton, "Analyses and measurements of acoustically matched, air-coupled Tonpilz transducers," Proc. IEEE Ultrasonic Sym., (1999), pp. 1045–1048.

[3]K. R. Dhilsha, G. Markandeyulu, B. V. P. S. Rao, and K. V. S. R. Rao, "Design and fabrication of a low frequency giant magnetostrictive transducer," J. Alloys Compd. **258**, 53–55 (1997).

[4]R. J. Meyer, J. T. C. Montgomery, and W. J. Hughes, "Tonpilz transducer designed using single crystal piezoelectrics," Oceans '02 MTS/IEEE (2002), Vol. **4**, pp. 2328–2333.

[5]D. Boucher, "Trends and problems in low frequency sonar projects design," in *Power Sonics and Ultrasonic Transducers Design*, edited by B. Hamonic and J. N. Decarpigny (Springer-Verlag, Berlin, 1987), Chap. 5, pp. 100–120.

[6]M. G. Silk, *Ultrasonic Transducers for Nondestructive Testing* (Atomic Energy Research Establishment, Harwell, UK, 1984), Chap. 2, pp. 4–24.

[7]P. Dufourcq, J. Adda, M. Letiche, and E. Sernit, "Transducers for great depths," in *Power Transducers for Sonics and Ultrasonics*, edited by B. Hamonic, O. B. Wilson, and J. N. Decarpigny (Springer-Verlag, Berlin, 1987), Chap. 6, pp. 121–133.

[8]R. Hanel, A. Mues, and R. Sobotta, "Description of ultrasound transducers through wave parameters, "Ultrasonics **34**, 159–162 (1996).

[9]K. Han and Y. R. Roh, "The performance of a 1-3 mode piezocomposite ultrasonic transducer in relation to the properties of its polymer matrix," Sens. Actuators, A **75**, 176–185 (1999).

[10]R. E. Newnham, D. P. Skinner, and L. E. Cross, "Connectivity and piezoelectric–pyroelectric composites," Mater. Res. Bull. **13**, 525–536 (1978).

[11]R. E. Newnham, "Ferroelectric composites," Jpn. J. Appl. Phys., Part 2 **24**, 16–17 (1985).

[12]Q. M. Zhang, J. Chen, H. Wang, and L. E. Cross, "A new transverse piezoelectric mode 2-2 piezocomposite for underwater transducer applications," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 774–781 (1995).

[13]N. Lamberti, F. M. Espinosa, A. Iula, and M. Pappalardo, "Array transducers based on 2-2 piezocomposites," Proc. IEEE Ultrasonics Symposium (1996), pp. 963–966.

[14]W. K. Qi and W. W. Cao, "Finite element analysis of periodic and random 2-2 piezocomposite transducers with finite dimensions," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1168–1171 (1997).

[15]B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. 1.

[16]X. C. Geng and Q. M. Zhang, "Evaluation of piezocomposites for ultrasonic transducer applications—influence of the unit sell dimensions and the properties of constituents on the performance of 2-2 piezocomposites," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 857–872 (1997).

[17]Q. M. Zhang, W. W. Cao, J. Zhao, and L. E. Cross, "Piezoelectric performance of piezoceramic-polymer composites with 2-2 connectivity-a combined theoretical and experimental study," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 556–564 (1994).

[18]Y. Shui and Q. Xue, "The piezoelectric coupling of 2-2 composite transducer," Proc. IEEE Ultrasonics Symposium (1995), pp. 929–932.

[19]S. Sanchez and F. R. M. Espinosa, "Modeling (2-2) piezocomposites partially sliced in the polymer phase," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 287–296 (1997).

[20]K. Kang and Y. R. Roh, "Optimization of structural variables of a flextensional transducer by the statistical multiple regression analysis method," J. Acoust. Soc. Am. **114**, 1454–1461 (2003).

Finite element modeling of torsional wave modes along pipes with absorbing materials

Michel Castaings[a] and Christophe Bacon
Laboratoire de Mécanique Physique, Université Bordeaux 1, UMR CNRS 5469, 351 cours Libération, 33405 Talence, France

(Received 16 November 2005; revised 16 March 2006; accepted 28 March 2006)

This paper describes the implementation of equations of dynamic equilibrium in a finite element (FE) code for modeling, in axisymmetry, the propagation of torsional wave modes along metallic pipes coupled to solid elements. Materials constituting either pipes and/or surrounding elements can be absorbing media, the absorption being caused by either their viscoelasticity or scattering on their internal structure (or both). Complex moduli are used as input data to equations of dynamic equilibrium, which are solved in the frequency domain rather than in the temporal domain. Their real and imaginary parts represent material elasticity and damping, respectively. A new definition of efficient and easy-to-implement absorbing regions that suppress undesired reflections from boundaries is proposed. The resolution of equations in the frequency domain, together with the use of absorbing regions, lead to significant reductions in the number of mesh elements and also in the number of iterations required for describing problems of propagation and scattering. Through two examples, the model is validated by successful comparisons of numerical predictions with experimental data. Then, a third example is presented to illustrate the importance of properly modeling waves damping when using FE models for setting-up or optimizing NDT techniques. [DOI: 10.1121/1.2197807]

PACS number(s): 43.35.Cg [YHB] Pages: 3741–3751

I. INTRODUCTION

Guided ultrasonic waves are used more and more for pipe testing because they allow rapid and reliable inspection to be performed, and they permit inaccessible areas to be interrogated. For instance, they can be excited from easy-to-access locations on the pipe, propagate many meters along the pipe, and return echoes indicating the presence of defects in insulated, coated or buried zones.[1–3] Among the different types of modes that can propagate along pipe walls, the torsional mode has many advantages:[4,5] In particular it has high sensitivity to cracks and suffers less attenuation than flexural or longitudinal modes since it is uncoupled to either internal or external fluids, thus leaking energy out of the pipe only if this is coupled to solid media. However this situation of solid contacts often occurs in pipes buried in soil, running through concrete walls, coated with epoxy-based materials, etc. Damping of the torsional waves modes can therefore come from viscoelasticity properties of the pipe itself and/or scattering by its internal structure, but also from energy leakage into surrounding solids, the leaky waves being also sometimes attenuated by viscoelasticity and/or scattering in these media.

In order to set and optimize (NDT) techniques of pipes, based on the propagation and diffraction of guided wave modes, numerical models based on the finite element (FE) or boundary element (BE) methods have been widely used.[6] These methods however rarely take into account damping in pipes themselves or due to absorbing elements coupled to pipes. Spring-dashpot elements are sometimes associated with time marching routines for FE modeling of attenuating waves in viscoelastic materials,[7,8] but the appropriate input values used to define the spring and dashpot coefficients are not commonly known or measured. In contrast, it is possible to measure wave attenuation versus frequency and to deduce variations of complex material moduli in given frequency domains,[9] and whether these moduli are frequency dependent or not, it is then straightforward to use them as input data in a frequency model. It may be possible to establish corresponding relaxation-time models,[10,11] but this can be quite difficult too.

This paper presents the use of a commercially available FE code[8] for modeling the propagation and diffraction of torsional wave modes along metallic pipes coupled to solid elements. Equations of dynamic equilibrium are implemented in the FE code and solved in the frequency domain rather than in the temporal domain, thus allowing complex moduli to be used as input data for simply and properly modeling the dynamic behavior of viscoelastic materials, as well as attenuation due to scattering by the internal structure of materials. Complex moduli also allow the definition of absorbing regions so that the model can represent the behavior of a structure which is infinitely long, while still using a small number of mesh elements. A new definition of absorbing regions is proposed which is simple to implement yet specific to pure modes propagating in homogeneous media.

Improvements in terms of efficiency are discussed in comparison to other types of absorbing regions presented in the literature.[12–16] The interest of working in the frequency domain is also reinforced by the significant reduction in the number of iterations required for describing problems of

[a]Author to whom correspondence should be addressed. Electronic mail: m.castaings@lmp.u-bordeaux1.fr

wave propagation and scattering. Validation of the model is first done by predicting the phase velocity and attenuation versus frequency of the fundamental torsional mode propagating along a copper pipe filled with bitumen, and comparing these results to published experimental data.[17] Then, the reflection of the fundamental torsional mode from an epoxy blockage cast inside an aluminium pipe is simulated, and compared to experimental results.[8] Finally, the necessity of taking into account damping in materials is demonstrated by simulating a NDT technique, which consists of detecting hidden corrosion in cast-iron pipes running through concrete walls.

II. FINITE ELEMENT MODELING

The software[8] used here is a numerical tool for solving partial differential equations (PDE) by a finite element method. Numerous physical models are implemented in the libraries provided with the software. However, if the considered model is not implemented, it must be written in a specific formalism involving coefficients that must be established via an identification procedure. The corresponding PDE problem in the most general coefficient formalism is

$$\begin{cases} d_a \dfrac{\partial u}{\partial t} - \nabla \cdot (c \nabla u + \alpha u - \gamma) + \beta \cdot \nabla u + a u = f & \text{in } \Omega \\ n \cdot (c \nabla u + \alpha u - \gamma) + q u = g & \text{on } \partial\Omega \\ h u = r & \text{on } \partial\Omega \end{cases}, \tag{1}$$

where u is a time-dependent variable to be determined in the domain called Ω of boundary $\partial\Omega$, and t is the time for nonstationary problems. The outward unit vector on $\partial\Omega$ is denoted by $\mathbf{n}$. The first equation is the PDE that must be satisfied inside the domain Ω. The second and the third ones are the Neumann condition and the Dirichlet condition, respectively, defined at the boundary $\partial\Omega$. The symbol ∇ is the vector differential operator defined as

$$\nabla = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2} \right), \tag{2}$$

where x_1 and x_2 are the orthogonal spatial coordinates for a two-dimensional problem. The other parameters are coefficients that depend on the problem.

The dimension of the variable u depends on the number of dependent scalar variables of the considered problem. In our case, u is a scalar with $u=u_\theta$. The identification of the coefficients $d_a, c, \alpha, g, \beta, a, q, g, h, r$ depends on the physical problem considered. The partial derivative equations in the domain and the Neumann and/or Dirichlet equations on the boundary must be written in order to identify all the coefficients of the FE formalism.

A. Axisymmetry formalism

We consider a dynamic problem where structural components are made of linearly viscoelastic materials. As demonstrated in Ref. 15, and due to the Boltzmann's principle of superposition,[19,20] an efficient way to treat this kind of problem is to solve equations in the frequency domain, by using Fourier transforms of variables, instead of solving in the time domain.

The cylindrical geometry of the structures considered in this study leads to the use of cylindrical coordinates r, θ, z. It is supposed that the cylindrical axes are orthotropic axes. We assume an axisymmetric problem ($\partial/\partial\theta=0$) where the Fourier transforms of the displacement components $\tilde{u}_r$ and $\tilde{u}_z$ are zero. Then the Fourier transforms of equilibrium equations in domain Ω are given by

$$C_{44} \frac{\partial^2 \tilde{u}_\theta}{\partial z^2} + C_{55} \frac{\partial^2 \tilde{u}_\theta}{\partial r^2} + \frac{C_{55}}{r} \frac{\partial \tilde{u}_\theta}{\partial r} - \frac{C_{55}}{r^2} \tilde{u}_\theta = -\rho \omega^2 \tilde{u}_\theta, \tag{3}$$

where $\tilde{u}_\theta$ is the Fourier transform of the displacement u_θ, ρ the density, ω the angular frequency related to the frequency f by $\omega=2\pi f$, and $C_{kk}=C'_{kk}+iC''_{kk}(k=4 \text{ or } 5)$ are shear components of the complex stiffness matrix ($i^2=-1$), with the following notation: axes 1, 2, and 3 correspond to axes of symmetry of the orthotropic material r, z, and θ, respectively. These moduli may depend on frequency. Their real and imaginary parts represent the material elasticity and damping, respectively. If the material is isotropic, then $C_{44}=C_{55}=\mu$, where μ is the Lamé shear modulus.

The components of the stress tensor $\bar{\bar{\sigma}}$ must be calculated for the Neumann conditions They are given in the frequency domain by

$$\tilde{\sigma}_{r\theta} = C_{55} \left(\frac{\partial \tilde{u}_\theta}{\partial r} - \frac{\tilde{u}_\theta}{r} \right), \quad \tilde{\sigma}_{\theta z} = C_{44} \frac{\partial \tilde{u}_\theta}{\partial z}. \tag{4}$$

The other components or equations are null. For Neumann boundary conditions, the stress vector is $\mathrm{T}=\bar{\bar{\sigma}} \cdot \mathbf{n}$. For a boundary surface normal to the r axis, $\mathbf{n}=(1,0)$, the stress vector has one component

$$\tilde{T}_\theta = \tilde{\sigma}_{r\theta} = C_{55} \left(\frac{\partial \tilde{u}_\theta}{\partial r} - \frac{\tilde{u}_\theta}{r} \right). \tag{5}$$

For a boundary surface normal to the z axis, $\mathbf{n}=(0,1)$, the stress vector has one component

$$\tilde{T}_\theta = \tilde{\sigma}_{\theta_z} = C_{44} \frac{\partial \tilde{u}_\theta}{\partial z}. \tag{6}$$

To avoid divisions by zero in Eq. (3) when $r=0$, Eqs. (3), (5), and (6) are multiplied by r^2. These equations then become

$$C_{44} r^2 \frac{\partial^2 \tilde{u}_\theta}{\partial z^2} + C_{55} r^2 \frac{\partial^2 \tilde{u}_\theta}{\partial r^2} + C_{55} r \frac{\partial \tilde{u}_\theta}{\partial r} - C_{55} \tilde{u}_\theta + \rho \omega^2 r^2 \tilde{u}_\theta = 0 \text{ in } \Omega, \tag{7}$$

$$C_{55} r^2 \frac{\partial \tilde{u}_\theta}{\partial r} - C_{55} r \tilde{u}_\theta = \tilde{T}_\theta r^2 \quad \text{for } \mathbf{n}=(1,0) \text{ on } \partial\Omega, \tag{8}$$

$$C_{44} r^2 \frac{\partial \tilde{u}_\theta}{\partial z} = \tilde{T}_\theta r^2 \quad \text{for } \mathbf{n}=(0,1) \text{ on } \partial\Omega. \tag{9}$$

Equations (7)–(9) are finally used to identify the coefficients of the FE formalism with Eqs. (1). For the axisymmetric cases considered in this study, there is one non-null vari-

able $\tilde{u}_\theta$ and the problem is a two-dimensional problem in the plane $(r,z)=(x_1,x_2)$. In these conditions, the coefficients become

$$c=\begin{pmatrix} C_{55}r^2 & 0 \\ 0 & C_{44}r^2 \end{pmatrix}, \tag{10}$$

$$\alpha=\begin{vmatrix} C_{55}r \\ 0 \end{vmatrix}, \quad \gamma=\begin{vmatrix} 0 \\ 0 \end{vmatrix}, \quad \text{and} \quad q=0,$$

$$\beta=\begin{vmatrix} 0 \\ 0 \end{vmatrix}, \quad a=-\rho\omega^2 r^2, \quad d_a=0, \quad \text{and} \quad f=0. \tag{11}$$

The term g represents the quantity $r^2\tilde{T}_\theta$ where $\tilde{T}_\theta$ is the stress vector imposed on the boundary. In the case of a stress-free boundary $g=0$. If Dirichlet conditions are imposed on a boundary (displacement imposed in the θ direction), the coefficient h is 1 and r takes the value of the imposed displacement.

B. Absorbing regions

1. Absorbing regions of type 1

Viscoelastic absorbing regions have been previously defined for avoiding undesired reflections from ends of meshed solid domains.[15,16] In two-dimensional problems, they consist in areas having gradually increasing viscoelastic properties, so that incoming waves do not encounter brutal changes in the acoustic impedance while being more and more attenuated as they propagate deeply into the region. They have been used so far in the case of waves propagating along plates, with the Cartesian system. In the case of torsional wave modes (of any order) propagating along pipes made of absorbing, orthotropic materials, both moduli $C_{44}=C'_{44}+iC''_{44}$ and $C_{55}=C'_{55}+iC''_{55}$ represent the only material properties to be considered in the equations of motion (7)–(9), with C'_{kk} and $C''_{kk}(k=4 \text{ or } 5)$ representing the material stiffness and viscoelasticity, respectively. The following function then describes the gradual increase applied to the imaginary part of these moduli:

$$C_{kk_a}=C'_{kk}\left[1+i\left(\frac{|z-z_a|}{L_a}\right)^3\right]+iC''_{kk}, \quad k=4 \text{ or } 5, \tag{12}$$

where z is the position along the structure, z_a and L_a are the starting position and the length along the z axis, respectively, of the absorbing region. The absolute value of the term $z-z_a$ allows the definition of regions that absorb waves traveling in the negative z direction. The use of a cubic function,[16] rather than an exponential one,[15] to increase the material viscoelasticity improves the efficiency of absorbing regions because it ensures much more progressive changes in the acoustic impedance. These absorbers are quite efficient since they usually reflect less than 0.1% of the energy of the incident guided wave modes, but their length must be at least three times the largest wavelength of all propagating modes existing in the frequency range of consideration.[21] Contrary to perfectly matched layers,[13] they also absorb inverse modes, defined as waves having phase and group velocities with opposite signs, and which may be produced by scattering of classical modes on inhomogeneities in guides.

2. Absorbing regions of type 2

In the case of the fundamental, zero-order, torsional wave $T(0,1)$ propagating along a uniform guide, made of a single orthotropic material, the mode is in fact produced by the shear bulk wave that propagates along the guide, with a velocity depending on the C_{55} modulus only.[22,23] A short calculation then shows that the complex acoustic impedance produced by this mode can be classically defined by

$$Z_{\text{guide}}=\frac{\tilde{\sigma}_{z\theta}}{\tilde{v}_\theta}=\frac{\tilde{\sigma}_{z\theta}}{i\omega\tilde{u}_\theta}=\sqrt{\rho(C'_{55}+iC''_{55})}, \tag{13}$$

where ρ is the mass density, ω is the angular frequency, $\tilde{\sigma}_{z\theta}$ is the shear stress produced by the wave, $\tilde{u}_\theta$ and $\tilde{v}_\theta$ are the particle displacement and velocity, respectively, in the frequency domain.

If the mechanical properties of the absorbing region are described by the complex modulus $C_{55a}=C'_{55a}+iC''_{55a}$, it is then possible to adjust its mass density ρ_a, so that the acoustic impedance in the absorber is exactly the same as that in the guide

$$\rho(C'_{55}+iC''_{55})=\rho_a(C'_{55}+iC''_{55a}). \tag{14}$$

Therefore, it is no longer necessary to gradually increase the imaginary part C''_{55a} of the modulus C_{55a} to avoid brutal changes in the acoustic impedance. A large value is then given to this imaginary part for all positions along the absorbing region, for example, 80% of the real part C'_{55a}, and C'_{55a} is chosen to be equal to C'_{55}. This simulates a material having a very high level of viscoelasticity. In this condition, the modulus C_{55_a} in the absorber in defined by

$$C_{55a}=C'_{55}(1+i0.8)=\beta(C'_{55}+iC''_{55})=\beta C_{55} \tag{15}$$

with $\beta=(1+i0.8)/(1+itg\delta)$ and $tg\,\delta=C''_{55}/C'_{55}$. Consequently, the mass density in the absorber is chosen to be equal to $\rho_a=\rho/\beta$, so that the acoustic impedance is unchanged all along the guide and the absorbing region

$$Z_{\text{absorber}}=\sqrt{\rho_a C_{55a}}=\sqrt{\frac{\rho}{\beta}\beta C_{55}}=\sqrt{\rho C_{55}}=Z_{\text{guide}}. \tag{16}$$

The interest of this definition is double compared to that of gradually viscoelastic absorbing regions: First, it is easier to implement in FE codes, especially for those which do not support spatially varying functions as input data, and which would require creating several adjacent regions with different levels of viscoelasticity.[14] Second the length of the region can be significantly reduced (for example, from three to one times the largest wavelength) as long as the imaginary part of the modulus C_{55a} is chosen to be high enough to cause very strong attenuation. This reduction in the length of the absorbing regions can be particularly advantageous for future three-dimensional (3D) modeling when computing resources will be in greatest demand. Moreover, this kind of absorber is well suited for modeling semi-infinite fluid or solid media when only one type of wave exists (longitudinal or shear bulk waves) since only one material property is then involved when modeling such cases. However, they cannot be used for absorbing Lamb-type modes that result from the

superposition of two fundamental bulk waves (in 2D), one longitudinal wave and one shear wave, that reflect between the surfaces of the guide under different directions and with different velocities that depend on several material parameters. Indeed, the acoustic impedance of Lamb modes is much more complicated than that of a pure fundamental mode existing in a single-material guide, thus preventing a perfect acoustic match between the guide and the absorbing region by appropriate adjustment of the mass density. This type of absorber is also not appropriate for attenuating a fundamental, zero-order mode that would propagate along a region of a pipe made of several layers, since the acoustic impedance cannot be a simple function of one density and of one material modulus, as in Eq. (13), but is necessarily a combination of densities and moduli of the various materials. For this reason, both types of absorbers will be used in the numerical simulations presented in Sec. III, absorbers of type 1 (AR1) with gradually increasing imaginary parts of moduli when the end of the pipe is made of more than one material, and absorbers of type 2 (AR2) with a fixed imaginary part of the modulus C_{55a} and an adjusted mass density, when the end of the pipe is made of a single material.

III. EXAMPLES

In the three examples considered in this study, the propagation and/or scattering of the fundamental, zero-order, torsional wave mode $T(0,1)$ is investigated. The excitation is a frequency-dependent stress locally applied at one edge of the structure, other edges of the structures being free of stress. The frequency spectrum is initially calculated by Fourier transforming the temporal signal that would classically be used as an excitation in time-marching FE codes. One advantage of this frequency domain approach is that less than 100 frequency components are usually sufficient to describe the same information as that contained in temporal signals composed of thousands of time points, especially in the case of toneburstlike excitations commonly used in NDT measurements, since these have rarely large frequency spectrum. The number of solutions required to predict the responses of the structures is therefore considerably reduced. Moreover, as described in Sec. II A, writing the equations of equilibrium in the frequency domain allows complex moduli to be used as input data so as to model properly the damping of the waves due to material viscoelasticity and/or material heterogeneity. In this way, the absorbing regions can also be defined as explained in Sec. II B. The response of a structure to a frequency excitation is given by complex displacements and stresses predicted at every node of the mesh. These output data can directly be plotted in different forms using a graphic interface, or post-processed in order to quantify specific parameters. For example, a direct spatial Fourier transform is applied to quantify the amplitude, phase velocity, or attenuation versus frequency of forward- or backward-traveling waves.[15] Also, a temporal inverse Fourier transform can be used to obtain temporal waveforms at specific locations in the structure.

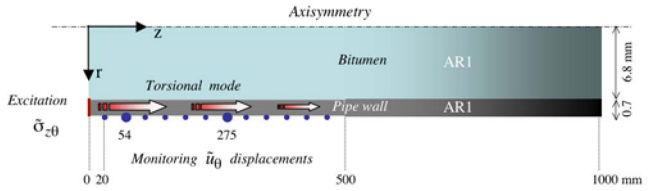


FIG. 1. (Color online) Schematic of copper pipe filled with bitumen, plus excitation, monitoring, and absorbing zones defined for FE simulations.

A. Propagation and attenuation of a torsional mode in a copper pipe filled with bitumen

In this first example, the purpose is to study the propagation of the torsional wave mode $T(0,1)$ along a copper pipe filled with bitumen (Fig. 1), i.e., to quantify the phase velocity and the attenuation of the mode for the various frequencies of the excitation. No scatterer is present in the structure and therefore forward-traveling waves only are simulated. This problem could have been modeled with standard methods based on propagator matrices, for example, the global matrix method,[24] but the interest in using the FE method is that further modeling of discrete scattering problems is possible, as shown through the other examples investigated later in this paper.

1. Input data

Copper is an isotropic material and its properties are given in Table I. Those of bitumen are defined by a complex Lamé modulus $\mu'_{\text{bit}}=\mu_{\text{bit}}+i\mu''_{\text{bit}}$, representing it as a viscoelastic, isotropic material. This modulus has been estimated from measured phase velocity C_s and attenuation A_s of the shear bulk wave in bitumen, shown in Fig. 2(a), according to the following relation:

$$\mu_{\text{bit}} = \rho_{\text{bit}} \frac{\omega^2}{k^2}, \tag{17}$$

where $\omega=2\pi f$ is the angular frequency, ρ_{bit} is the density, $k=k'+ik''$ is the complex wave number, with $k'=\omega/C_s$ and $k''=\text{Log}(10^{-A_s/20})$.

The phase velocity C_s and the attenuation A_s have in fact been deduced from those of the torsional mode $T(0,1)$ propagating along a copper pipe filled with bitumen, and measured for a set of frequencies comprised between 10 and 80 kHz, as described in Ref. 17. The set of values thus obtained for the complex modulus μ_{bit} are presented in Fig. 2(b). They finally have been fitted with quadratic functions of frequency, which will be used as input functions in the FE modeling

$$\begin{aligned} \mu'_{\text{bit}} &= 0.02652 + 4.4754f - 33.363f^2, \\ \mu''_{\text{bit}} &= -0.03084 + 3.9141f - 35.042f^2. \end{aligned} \tag{18}$$

TABLE I. Copper material properties.

	ρ (g/cm)^3	C_s (m/s)	A_s (dB/m)	μ (GPa)
Copper	8.9	2240	0	44.65

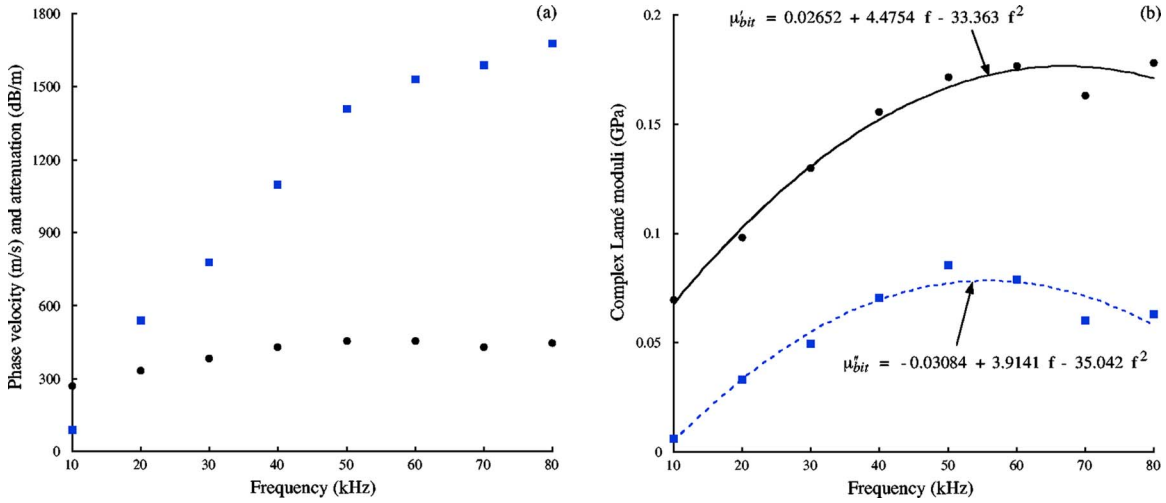


FIG. 2. (Color online) (a) Measured phase velocity (•) and attenuation (■) of shear bulk wave in bitumen from [17] and (b) values of real (•) and imaginary (■) parts of bitumen Lamé modulus derived from data in (a), plus fitting curves (lines).

The excitation is a uniform-amplitude stress $\tilde{\sigma}_{z\theta}=1$ MPa applied through the pipe wall, at the left-hand end of the pipe. The signal of excitation is initially defined by a two-cycle, Gaussian-windowed toneburst centered on 45 kHz. Its frequency spectrum extends from 6 to 85 kHz, down to −15 dB, and is used as the input to the FE model. The response of the plate is then calculated for each of the 42 frequency components of this spectrum.

Absorbing regions are simulated at the right-hand side of the aluminium pipe, and also of the bitumen. No absorbers are simulated at the left-hand side of the structure since the excitation is applied to it, and since no scatterer in the structure is present to produce backward-traveling waves. The implemented absorbers are of type 1 (AR1 in Fig. 1), i.e., those gradually increasing the imaginary part of the copper and bitumen Lamé moduli as the z position increases. Their length is, chosen to be equal to three times the maximum wavelength expected in this region of the structure, and in the whole frequency spectrum, i.e., $3^*167=500$ mm, if one considers that the wave velocity should be somewhat comprised between that in the copper ($\approx$2.2 mm/μs) and that in the bitumen ($\approx$0.4 mm/μs), i.e., around 1.3 mm/μs, at the lowest frequency of interest (8 kHz).

The mesh used for this problem is made of rectangular, Lagrange, linear elements of 1 mm length in the z direction, and of 0.68 mm through the bitumen and of 0.175 mm through the pipe wall, in the radial direction. In total, the structure is 1 m long, including the 500 mm long absorbing regions. It is meshed by 14 000 elements and 15 015 degrees of freedom. The time of computation lasted around 1 minute using a 1.8 GHz, biprocessor MACINTOSH G5 machine.

2. Results

Real $\tilde{u}_\theta$ displacements are monitored all along the outer pipe surface. Figure 3 shows the huge effect of the bitumen viscoelasticity on the torsional mode amplitude, at 50 kHz, which is a quite low frequency. Indeed, the mode amplitude is already almost zero when entering the absorbing region, if the bitumen viscoelasticity is taken into account. The efficiency of the absorbing regions placed at the right-hand side of the bitumen and the copper is clearly visible in Fig. 3(a). A selection of 200 points located from positions $z=20$ mm to $z=500$ mm is used for post-processing, which consists of Fourier transforming the complex $\tilde{u}_\theta(z)$ data, for each frequency. The wave-number-frequency diagram thus obtained is used to quantify the phase velocity and attenuation of the torsional mode, as explained in Ref. 15. Figure 4 shows the very good agreement obtained between the predicted and measured phase velocities and attenuations. Indeed, the numerical predictions are practically always inside the range of errors of the experimental data. The little differences between the two sets of results are probably due to the fact that the values of the Lamé modulus μ_{bit} used as input data to the

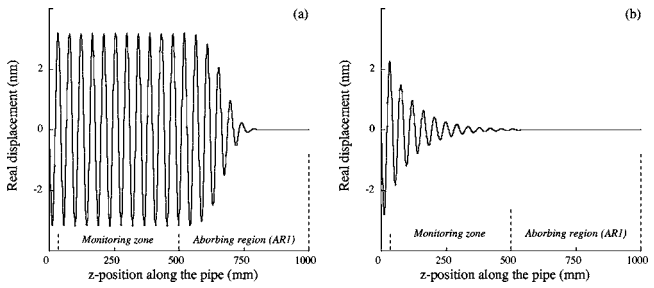


FIG. 3. Real $\tilde{u}_\theta$ displacement monitored along the outer copper pipe surface for $f=50$ kHz when the bitumen is (a) elastic and (b) viscoelastic.

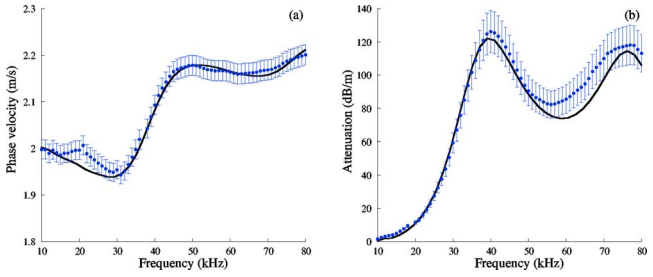


FIG. 4. (Color online) Phase velocity (a) and attenuation (b) of torsional mode along copper pipe filled with bitumen; measurements from [17] (dots with range of errors) and FE predictions (lines).

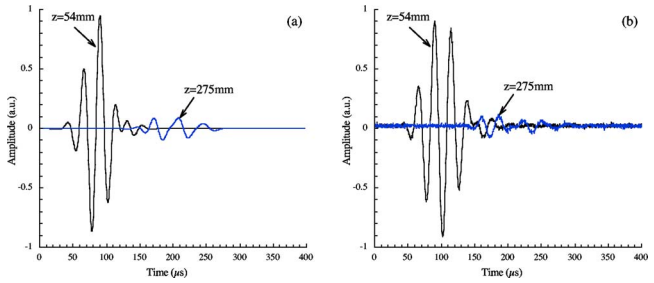


FIG. 5. (Color online) Wave forms for torsional mode detected at two positions (z=54 mm and z=275 mm) along the surface of the copper pipe filled with bitumen, (a) FE numerical predictions and (b) experimental data from [17].

FE model come from the interpolation functions given in Eqs. (18), which have been established from a set of eight estimated values of μ_{bit}, although the FE solution is produced for 42 frequencies. Nevertheless, the good correlations between the FE-predicted and the measured phase velocities and attenuations validate the frequency-dependent complex modulus of bitumen obtained from the experiments described in Ref. 17, and also the model used and presented in this paper.

Figure 5(a) shows temporal traces predicted at two positions, z=54 mm and z=275 mm at the pipe surface, when the bitumen is considered to be viscoelastic in the model, and for a five-cycle, Hanning-windowed, toneburst excitation centered on 25 kHz. These were obtained by running the FE model over a 27-frequency-component excitation, and by inverse Fourier transforming the set of 27 complex $\tilde{u}_\theta$ displacements obtained for each of the two positions. These signals can successfully be compared to those measured at the same positions [Fig. 5(b)], in the experiments described in Ref. 17. It is clear that the bitumen viscoelasticity, which strongly attenuates the mode over the 220 mm path of propagation, is correctly taken into account in the model.

B. Reflection of a torsion mode by an epoxy end in an aluminium pipe

In this second example, the purpose is to study the reflection of the torsional mode $T(0,1)$ by the front edge of an epoxy blockage cast inside an aluminium pipe (Fig. 6), i.e., to quantify the reflection coefficient defined by the ratio between the amplitude of the reflected wave and that of the incident wave. The inner radius of the pipe is 8.2 mm and its wall thickness is 1.4 mm. Both pipe and epoxy must be modeled to represent infinitely long regions since the purpose is to avoid reflections from the ends of the pipe.

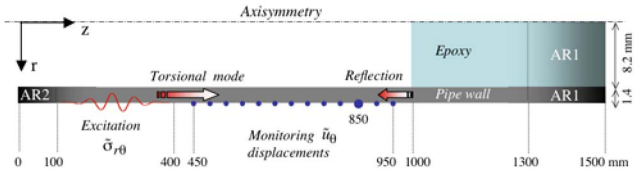


FIG. 6. (Color online) Schematic of aluminium pipe with epoxy end, plus excitation, monitoring, and absorbing zones defined for FE simulations.

1. Input data

Both aluminium and epoxy are considered to be isotropic materials, the properties of which are given in Table II. In the simulations, the epoxy is first considered to be an elastic material (real Lamé modulus is used as input data) and then a viscoelastic material (using complex Lamé modulus) considering two levels of viscoelasticity in order to emphasize the effect of the viscoelasticity on the reflected torsional mode.

The excitation is a complex shear stress applied on the outer surface of the pipe. Its amplitude varies with the position z along the pipe, so that it locally produces z-dependent deformations that correspond to the wavelength of the desired mode. Its phase is such that the mode is launched in the direction of the blockage only (in the positive z direction). This excitation is defined by the following function:

$$\tilde{\sigma}_{r\theta} = e^{-ikz} \cdot e^{-20(z-z_0)^2/L^2}, \tag{19}$$

where z_0=250 mm is the center position of the excitation zone and L=300 mm is its extent along z (L is chosen to be equal to three times the maximum wavelength). Note that $\tilde{\sigma}_{r\theta}$=1 at z=z_0, and tends toward 0 for z<100 mm and for z>400 mm, so that the outer surface of the pipe is free of stress except between 100 and 400 mm where the excitation is applied. $k=2\pi f/V_{\text{torsion}}$ is the wave number of the desired torsional mode $T(0,1)$, with V_{torsion} being its phase velocity and f the frequency.

The signal of the excitation is initially defined by a three-cycle, Gaussian-windowed toneburst centered on 150 kHz. Its frequency spectrum extends from 30 to 250 kHz, down to −25 dB, and is used as the input to the FE model. The response of the plate is then calculated for each of the 253 frequency components of this spectrum. The frequency step is very small because it is intended to predict temporal signals at a distance of 600 mm away from the center of the excitation zone and 150 mm before the epoxy blockage (z=850 mm), so the incident and reflected modes would have to travel about 1000 mm all together; this requires a long time domain to be defined (up to 800 μs), especially because echoes produced inside the epoxy have to

TABLE II. Aluminium and epoxy material properties.

	ρ (g/cm)^3	C_s (m/s)	A_s (dB/m)	μ (GPa)
Aluminium	2.7	3130	0	26.45
Elastic epoxy	1.17	1010	0	1.19
Low viscoelastic epoxy	-	-	205 at 150 kHz	1.19+i0.06
High viscoelastic epoxy	-	-	410 at 150 kHz	1.19+i0.12

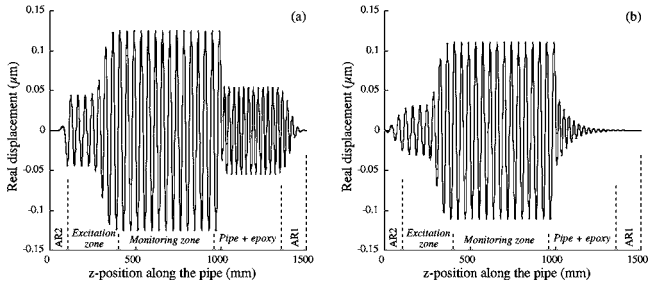


FIG. 7. Real $\bar{u}_\theta$ displacement monitored along the outer aluminium pipe surface for $f=75$ kHz when epoxy blockage is (a) elastic with $\mu=1.19$ GPa and (b) viscoelastic with $\mu=1.19+i0.06$ GPa.

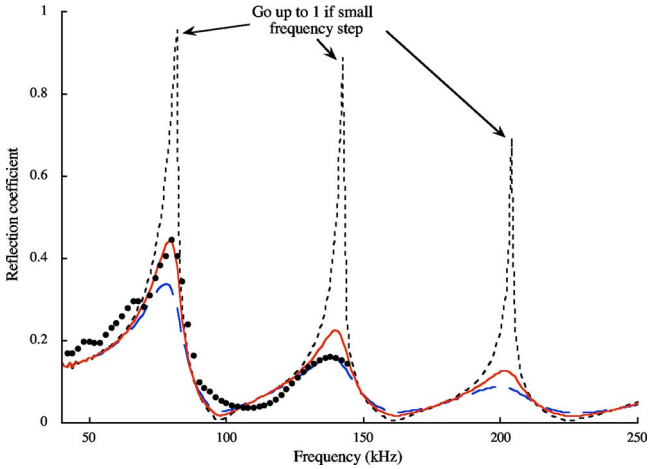


FIG. 8. (Color online) Reflection coefficient versus frequency for torsion mode reflected from epoxy blockage in aluminium pipe; FE numerical predictions for elastic epoxy with $\mu=1.19$ GPa (---), viscoelastic epoxy with $\mu=1.19+i0.06$ GPa(—) and strongly viscoelastic epoxy with $\mu=1.19+i0.12$ GPa (––), plus experiments (•••) from [18].

be considered in these simulations. This small frequency step will be convenient anyway to emphasize resonances in the predicted spectrum.

Absorbing regions are simulated at each end of the aluminium pipe, and also at the right-hand side of the epoxy blockage, which must be represented as infinitely long to suppress undesired reflections. The two types of absorbers mentioned in Sec. II B have been used. Absorbers of type 1 (AR1 in Fig. 6), which increase the imaginary parts of the Lamé moduli, have been used at the right-hand side of the aluminium and the epoxy. The length of this absorbing region is chosen to be equal to three times the maximum wavelength expected in the right-hand side of the pipe, and in the whole frequency spectrum, i.e., $3^*66=200$ mm, if one considers that the wave velocity should be somewhat comprised between that in the aluminium (≈ 3 mm/μs) and that in the epoxy (≈ 1 mm/μs), and that the lowest frequency of interest is 30 kHz. The second type of absorber (AR2 in Fig. 6) is used for suppressing reflections from the left-hand side of the pipe. It imposes the imaginary part, of the Lamé modulus to be 80% of its real part, so that the material becomes highly viscoelastic, its mass density being adjusted so that the acoustic impedance is kept the same as that in the pipe, as explained in Sec. II B, i.e., $\mu_a=26.45+i21.16$ GPa and $\rho_a=1.64-i1.32$ g/cm^3. The length of this absorbing region is chosen to be equal to the maximum wavelength expected in the free pipe, and in the whole frequency spectrum, i.e., 100 mm.

The mesh used for this problem is made of rectangular, Lagrange, linear elements of 0.5 mm length in the z direction, and of 0.5125 mm through the epoxy and of 0.35 mm through the pipe wall, in the radial direction. In total, the structure is 1.5 m long, including the absorbing regions. It is meshed by 28 000 elements and 31 021 degrees of freedom. The time of computation for the 253-frequency loop lasted around 10 min using a 1.8 GHz, biprocessor MACINTOSH G5 machine. Note that if the problem was solved in the time domain, the classical criterion for the steps of the time marching routines[25] would require around 28 times more solutions than that in the frequency domain.

2. Results

Real $\bar{u}_\theta$ displacements are monitored all along the outer pipe surface. Figure 7 shows the effect of the viscoelasticity of the epoxy blockage on the torsional mode amplitude, at 75 kHz. The efficiency of both types of absorbing regions (AR1 and AR2) is visible on this figure. A selection of 200 points located from positions $z=450$ mm to $z=950$ mm, i.e., between the excitation and the epoxy blockage, is used for post-processing, which consists of Fourier transforming the complex $\tilde{u}_\theta(z)$ data, for each frequency. The wave-number-frequency diagram thus obtained is used to identify the amplitudes of the incident and blockage-reflected torsional modes, and therefore to quantify the reflection coefficient.

Figure 8 presents the reflection coefficient versus frequency predicted when the epoxy is elastic and when it is defined with the two levels of viscoelasticity given in Table II. The effect of the epoxy viscoelasticity is tremendous, especially on the amplitudes of resonant peaks (80, 142, and 203 kHz), which are much lower than unity, this value being reached if the epoxy is modeled as an elastic material. Also, it can be seen that amplitudes of anti-resonances (98, 162, and 225 kHz) are zero if the epoxy is modeled as an elastic medium, which is no longer the case if its viscoelasticty is taken into account. Dots on the figure present experimental data measured according to the technique described in Ref. 18. Very good agreement is obtained with the reflection coefficient predicted using $\mu=1.19+i0.06$ GPa as input data, especially below 100 kHz. Indeed, the experimental procedure consisted of capturing series of signals along a circumferential line, and then post-processing them so that undesired flexural modes were eliminated. As explained in Ref. 18, this technique works better below than above 100 kHz, thus explaining the slight differences between the numerical and experimental data, above this frequency. However, the good correlation confirms that it is necessary to take the epoxy viscoelasticity into account to model properly the reflection phenomenon studied here.

Figure 9 presents temporal traces predicted considering the epoxy to be elastic [Fig. 9(a)] and then with the low level of viscoelasticity [Fig. 9(b)]. The effect of the epoxy viscoelasticity is slightly less visible than on the reflection coefficient plotted versus frequency (Fig. 8), since only the

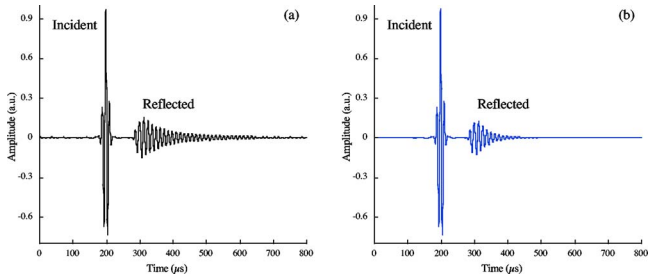


FIG. 9. (Color online) FE predicted wave forms showing incident and reflected torsion modes detected on the surface of the aluminium pipe, at $z=850$ mm, when epoxy is modeled as (a) an elastic material with $\mu=1.19$ GPa and (b) a viscoelastic material with $\mu=1.19+i0.06$ GPa.

small-amplitude tail of the reflected time trace disappears. No experimental signals are shown in this figure since these contain traces of flexural modes thus making the comparison with predictions not possible in the time domain.

C. Fluid-filled cast-iron pipe running through concrete wall

This last example concerns a NDT application to detect corrosion in fluid-transporting cast-iron pipes. Pipes sometimes run through concrete walls or blocks, both sides of which are not always accessible. It is therefore of great interest to launch ultrasonic waves from the accessible side, and to use them for inspecting hidden, critical zones of the pipe, for instance that immediately located behind the wall, which may be corroded if humidity oozes along the wall. Figure 10 illustrates an infinitely long cast-iron pipe of 100 mm in diameter and 5 mm wall thickness, running though a 200 mm thick concrete wall, which has an infinite extend in the direction normal to the pipe axis. The first purpose is to simulate a low-frequency (100 kHz) incident torsional wave mode $T(0,1)$ launched from the pipe surface located at left-hand side of the wall, and its reflection from the pipe-wall junction situated at the other side of the wall. The second goal is to check if this reflection can clearly indicate the presence of corrosion at this junction, and to quantify the effect of the attenuation of the waves on this indication.

1. Input data

As mentioned in Ref. 26, ultrasonic waves are much attenuated in cast-iron media as the nodularity percentage decreases. This paper presents velocity and attenuation measurements of the longitudinal bulk wave, and shows that they

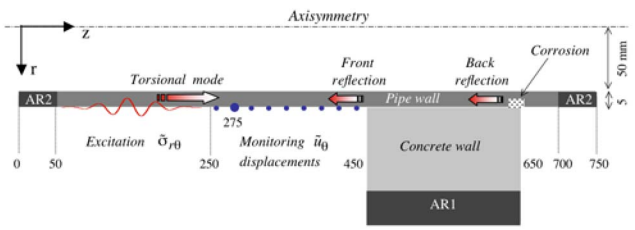


FIG. 10. (Color online) Schematic of cast-iron pipe running through a concrete wall, plus excitation, monitoring, and absorbing zones defined for FE simulations.

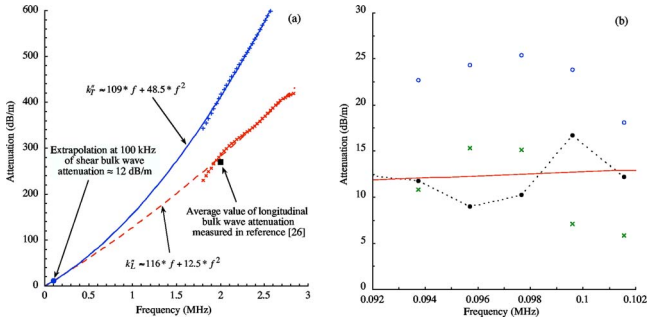


FIG. 11. (Color online) Attenuation in cast-iron plate: (a) of longitudinal and shear bulk waves, experimental data (crosses) and quadratic fitting functions (lines) and (b) of A0 Lamb mode with (○) attenuation in iron plate due to angular aperture and scattering, (×): attenuation in glass plate due to angular aperture and (•) attenuation in iron plate due to scattering only, plus (—) attenuation predicted with complex Lamé moduli.

can vary between 4000 and 5600 m/s, and between 67 and 473 dB/m at the frequency $f=2$ MHz, respectively, according to the quality of the iron. However, to our knowledge, no experimental values of the attenuation of shear waves in cast iron have been published. Therefore, a 12 mm thick cast-iron plate has been used to quantify both the velocity and the attenuation of shear waves, in the low frequency range of interest, and to estimate the value of the complex Lamé modulus that should be used as input data to the FE model. The classical immersion technique described in Ref. 9 is first used to collect data in the [1.8–2.8] MHz domain. The velocities of both longitudinal and shear bulk waves have been estimated to $V_L \approx 4100$ m/s and $V_T \approx 2400$ m/s, respectively. These values are consistent with standard velocities in cast iron. Then, the attenuations of both longitudinal and shear modes have been measured. These are presented in Fig. 11(a). At 2 MHz, the measured longitudinal attenuation $k_L''(\times\times\times)$ is very close to the average value (■) calculated from the range of values given in Ref. 26, thus giving good confidence in our results obtained with the immersion technique. Therefore, the shear wave attenuation k_T'' measured between 1.8 and 2.8 MHz (+++) is considered to be reliable, and used to estimate the attenuation at 100 kHz, which is the center frequency of the further FE modeling of the propagation of the torsional wave mode along a cast-iron pipe. A second-order polynomial function of frequency has been chosen to fit the measured shear wave attenuation k_T'', and to estimate its value at the center frequency of the further FE runs. The value 12 dB/m was then found for the attenuation k_T'' at 100 kHz. Consequently, the complex wave number of the shear bulk wave, at 100 kHz, was estimated: $k_T=(2\pi f/V_T)-ik_T''=0.2618-i0.0013$ mm^{-1}, and used in Eq. (17) to quantify the complex Lamé modulus for cast iron: $\mu_{\text{iron}}=45+i0.45$ GPa, the mass density being equal to 7.85 g/cm^3. In a similar way, using the measured velocity of the longitudinal bulk wave and its attenuation extrapolated at 100 kHz, the other complex Lamé modulus was estimated: $\lambda_{\text{iron}}=42+i0.51$ GPa.

To check whether scattering by the internal structure of cast iron has or not a significant effect on the attenuation of the waves, at 100 kHz, these two complex moduli have been used to predict the attenuation of the first-order, antisymmet-

ric Lamb mode A_0 in the 12 mm thick cast-iron plate. As shown in Fig. 11(b), this predicted attenuation (—) of A_0 is in good agreement with that measured (•) in the cast-iron plate sample. In the experiments, the A_0 mode was propagated along a 170 mm length path, and its decay in amplitude (ooo) was measured using air-coupled transducers and the technique described in Ref. 27. To remove the amplitude decay due to the angular aperture of the ultrasonic beam produced by the transmitter, this has first been measured on a perfectly nonabsorbing glass plate (×), and then substracted from the amplitude decay measured on the iron plate. The good correlation between the resulting experimental attenuation of A_0 and the predicted one confirms that the complex Lamé moduli were correctly estimated, and that the attenuation due to scattering by the internal structure of cast iron is not negligible, even at low frequencies around 100 kHz, especially if waves have to be propagated along a long distance, as is often the case in pipe testing problems.

Finally, the complex Lamé modulus $\mu_{\text{iron}}=45+\text{i}0.45$ GPa is going to be used as input data in the FE code, to model the propagation of the torsional wave mode along a cast-iron pipe. This modulus can be considered to be constant in the frequency range of interest [80–120] kHz, since the shear bulk wave attenuation varies practically linearly with frequency, in this low frequency domain [Fig. 11(a)].

The corroded zone is simulated by very large degradations of the mechanical properties of the cast-iron, in a 8 mm long by 3 mm thick zone located at the pipe surface, just at the junction between the back side of the wall and the pipe, as shown in Fig. 10. In fact, the software which was used[8] allows, without difficulty, the properties to be defined to be dependent on spatial coordinates. Therefore, both mass density and Lamé modulus are divided by a factor, corr, the value of which is 90 inside the corroded zone, and 1 elsewhere

$$\mu_{\text{iron}}=\frac{45+\text{i}0.45}{\text{corr}} \text{ and } \rho_{\text{iron}}=\frac{7.85}{\text{corr}} \text{ with corr}\begin{cases}=90 & \text{if } 650\leqslant z\leqslant 658 \text{ and } 52\leqslant r\leqslant 55 \text{ mm}\\ =1 & \text{elsewhere.}\end{cases} \tag{20}$$

The concrete properties have been measured using the ultrasonic technique described in Ref. 28. Several concrete blocks have been used to estimate the average mass density, as well as complex moduli for two directions running along the blocks sides. Contrary to results given in Ref. 29, a small anisotropy has been observed since moduli measured in the two directions are systematically slightly different for the various blocks. Indeed, C_{44} and C_{55}, which correspond to the Lamé moduli for the two considered directions, are shown to have different values 16+i0.3 GPa and 18+10.4 GPa, with less than 10% of error in the measurements. In these experiments, the excitation was a five-cycle toneburst the center frequency of which (150 kHz) was inside the frequency domain of the FE simulations presented here. Therefore, the input data used for the FE modeling of the concrete are $\rho_{\text{concrete}}=2.3$ g/cm^3, and $C_{44}=16+\text{i}0.3$ GPa and $C_{55}=18+\text{i}0.4$ GPa, the imaginary parts of C_{44} and C_{55} representing more attenuation due to scattering by heterogenities in concrete than attenuation due to viscoelasticty.

The pipe length is chosen to be equal to 750 mm, but it is in fact rendered infinitely long by two 50 mm long absorbing regions of type 2, defined as explained in Sec. II B. The concrete wall is 200 mm wide, and also infinitely high thanks to a 50 mm long absorbing region of type 1. The whole domain is meshed by 24 500 Lagrange, linear elements. These are 1 mm square elements in the concrete region, and rectangular elements in the pipe wall, of 1 mm length in the axial direction and 0.83 mm in the radial direction. The total number of degrees of freedom is 25 357.

The excitation is a complex shear stress applied to the outer surface of the pipe, as that defined in Eq. (19), except that $z_0=150$ mm, $L=200$ mm, and the wave-number $k=2\pi f/V_{\text{torsion}}$ takes different values versus frequency since the material and pipe geometry are different. The phase velocity of the desired mode is now $V_{\text{torsion}}=\sqrt{\mu_{\text{iron}}/\rho_{\text{iron}}}=\sqrt{45\cdot 10^9/7850}\approx 2400$ m/s and the frequency spectrum of the excitation is obtained by Fourier transforming ten-cycle, Gaussian-windowed toneburst of 100 kHz center frequency, defined over a 700 μs temporal window. This spectrum is composed of 43 components extending from 75 to 125 kHz, down to −30 dB. All other boundaries of the system are stress free.

2. Results

The computation lasted 2 min on the G5 MACINTOSH machine. Figure 12 shows temporal wave forms predicted without and with taking into account damping in the concrete wall and in the cast-iron pipe, i.e., by using either real or complex moduli as input data, respectively. These wave forms have been obtained by inverse Fourier transforming the complex $\tilde{u}_\theta$ predicted at the position $z=275$ mm on the pipe outer surface. Their amplitudes are normalized with respect to the maximum amplitude obtained for the incident wave mode in the nonabsorbing structure. The monitored position has been chosen so that wave packets corresponding to reflections from the front and back pipe-wall junctions are separated from each other, and also from the packet corresponding to the incident wave. The whole distance covered by the incident wave and a reflected wave is then 475 or 875 mm, if one considers the wave reflected by the front or back junction, respectively. In these conditions, the attenuation along the pipe has a tremendous effect since the wave

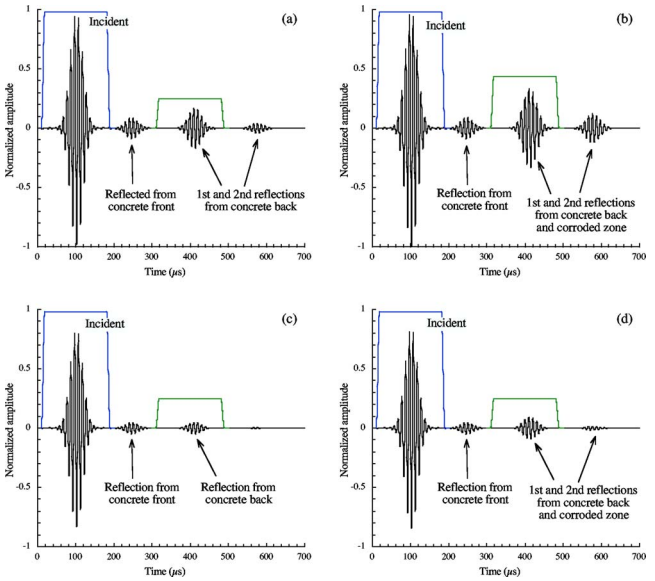


FIG. 12. (Color online) FE predicted wave forms showing incident and reflected torsion modes detected at position $z=275$ mm on the surface of the cast-iron pipe running through a concrete wall, with no corrosion (a) and (c) and with corrosion (b) and (d), and when damping in both concrete and iron is taken into account (c) and (d) or not (a) and (b).

packet reflected from the tested zone, i.e., from the back pipe-wall junction, has much smaller amplitude [Figs. 12(c) and 12(d)] than that predicted if the pipe is considered as nondamping material [Figs. 12(a) and 12(b)]. It is interesting to note that damping in the concrete has been checked to have a negligible effect on the attenuation of the torsional wave modes guided along the pipe. This comes from the fact that the amount of energy leaked into the concrete depends on its acoustic impedance, which is roughly the same whether damping is considered or not, because of the low damping level: MAX $(tg\delta)=C''_{55}/C'_{55}=0.4/18\approx 2\%$. Figure 12 shows that the presence of the corroded zone at the back pipe-wall junction similarly doubles the amplitude of the reflected signal whether damping in the iron is modeled or not. Therefore, one could argue that wave forms predicted using complex moduli in the FE model could be obtained if real moduli were used together with a correction factor to take into account the attenuation of waves reflected along the pipe. This is not so obvious since such a correction factor would have to take into account attenuation due to damping in the iron pipe, and also due to leakage in the concrete, both damping being differently frequency dependent, and also effective along different lengths of propagation. It is in fact much easier to use the complex moduli as input data in the FE model.

Figure 13 presents the modulus of the ratio (in decibels) versus frequency of $\tilde{u}_\theta$ displacements produced by the torsional wave mode reflected from the back pipe-wall junction, to those produced by the incident mode. These displacements have been picked up at the position $z=275$ mm on the outer surface of the pipe. In Fig. 13(a) damping in the cast-iron pipe is not taken into account, thus showing ratios much bigger that those presented in Fig. 13(b), Of course, these differences due to damping are more or less important according the position z chosen along the pipe. Indeed, the

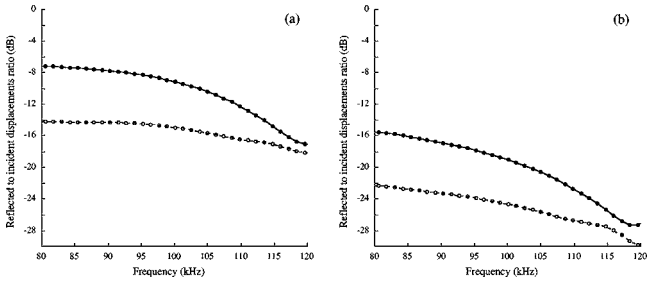


FIG. 13. FE predicted ratio of displacements produced by first reflection from wall back and by incident torsional modes at $z=275$ mm on the surface of the iron pipe running through a concrete wall, with no corrosion (◦) and with corrosion (•), and when damping in both concrete and iron is taken into account (b) or not (a).

ratio at a given position z between reflected and incident wave displacements is in fact that existing at the defect, commonly defined as the reflection coefficient, times a frequency-dependent attenuation factor which takes into account damping along the distance between position z and the defect.

The results shown in Figs. 12 and 13 indicate anyway that the use of torsional wave modes for experimentally detecting corroded pipe zones hidden behind concrete walls, may be less easy than what is predicted by FE models which do not properly consider material damping. This does not mean that torsional wave modes cannot be used for such purposes, but that careful interpretations of FE predicted results must be done before using simulations for setting or for optimizing long-range NDT techniques.

IV. CONCLUSIONS

The propagation and diffraction of torsional wave modes in axisymmetry problems have been simulated by solving equations of dynamic equilibrium in the frequency domain, using a finite element-based software. Previously measured frequency-dependent or -independent complex moduli have been used as input data. These allow simple modeling of realistic damping due either to materials viscoelasticity and/or to scattering on internal heterogeneities to be done. Spatially dependent material properties have also been used for simulating absorbing regions that suppress undesired reflections from some boundaries of the domain. A new, simple definition of very efficient absorbing regions has been proposed. This consists of defining a very high level of viscoelasticty for the whole absorbing domain, and in adjusting its mass density so that its acoustic impedance is perfectly continuous with that of the main: domain of propagation. However, the use of this definition is possible if the acoustic impedance of the guide is a simple function of the mass density and of the phase velocity, this latest being also a simple function of the material properties. This is the case for the fundamental, first-order torsional wave mode propagating along homogeneous guides. Although the use of such absorbing regions is quite restrictive, it is very simple to implement and leads to reducing the lengths of absorbing regions by a factor of 3, which may be very beneficial for further modeling in three dimensions.

Three cases have been modeled, all considering the propagation and/or scattering of the fundamental, first-order torsional wave mode. The first case consisted of simulating experiments conducted at Imperial College, London, on the propagation in a copper pipe filled with bitumen.[17] Good correlation between measured and predicted phase velocity and attenuation of the torsional mode validated the FE model. The second case consisted of simulating the reflection of the incident torsional mode from an infinitely long epoxy blockage cast inside an aluminium pipe. The new definition of absorbing regions was used here and its efficiency was successfully checked. The reflection coefficient (ratio between amplitudes of reflected and incident modes) was predicted for various levels of viscoelasticty of the epoxy, and comparison with measured reflection coefficients allowed the estimation of the order of magnitude of the imaginary part of the epoxy Lamé moduli. Finally, the last case consisted in simulating a torsional mode propagating along a cast-iron pipe running through a concrete wall, and its reflection from a hidden zone located at the pipe-wall junction, past the wall. Simulations were conducted by considering either no damping or damping both in the cast iron and in the concrete, and also when the pipe had, or not, corrosion damage all around its circumference at the pipe-wall junction. The corroded zone was simulated by locally dropping the cast-iron mass density and Lamé modulus. Results showed that not taking into account the damping of the waves can lead to overestimation of the potential of torsional wave modes for long-range NDT applications.

In the three cases, the interest of solving the equations of dynamic equilibrium in the frequency domain has been emphasized by the very convenient and easy way to use measured complex material properties to realistically model the damping of waves due to either viscoelasticity or scattering by the internal structure. Moreover, time traces have been systematically reconstructed by inverse transforming sets of complex displacements predicted at given positions and for series of frequencies, in order to demonstrate that working in the frequency domain also allows easy and quick temporal simulations to be done.

ACKNOWLEDGMENTS

The authors are very grateful to Mike Lowe, Jian Ma, Francesco Simonetti, and Peter Cawley for having kindly supplied the experimental data included in this paper. Also many thanks to Mike Lowe for his very helpful contribution in correcting this paper.

[1]J. McGregor, B. Nooteboom, and N. Ivory, "The application of long range guided ultrasonics for the inspection of riser pipes," *16th WCNDT 2004, World Conference on NDT*, http://www.wcndt2004.com

[2]H. Shin and J. Rose, "Guided wave tuning principles for defect detection in tubing," J. Nondestruct. Eval. **17**, 27–36 (1998).

[3]D. Alleyne and P. Cawley, "Long range propagation of Lamb waves in chemical plant pipework," Mater. Eval. **55**, 504–508 (1997).

[4]Yong-Moo Cheong *et al.*, "Detection of axial cracks in a bent pipe using EMAT torsional guided waves," *16th WCNDT 2004, World Conference on NDT*, http://www.wcndt2004.com

[5]A. Demma, P. Cawley, and M. J. S. Lowe, "The reflection of the fundamental torsional mode from cracks and notches in pipes," J. Acoust. Soc. Am. **114**, 611–625 (2003).

[6]J. Mackerle, "Finite-element modelling of non-destructive material evaluation, an addendum: A bibliography (1997–2003)," Modell. Simul. Mater. Sci. Eng. **12**, 799–834 (2004).

[7]E. Bécache, A. Ezziani, and P. Joly, "A mixed finite element approach for viscoelastic wave propagation," Comput. Geosci. **8**, 255–299 (2004).

[8]FEMLAB, User's Guide and Introduction. Version 3.0 by – COMSOL AB 2004 http://www.comsol.com/

[9]B. Hosten, "Heterogeneous structure of modes and Kramers-Kronig relationship in anisotropic viscoelastic materials," J. Acoust. Soc. Am. **104**(3), 1382–1388 (1998).

[10]L. Jiang and J. B. Haddow, "A finite element solution of plane wave propagation in inhomogeneous linear viscoelastic solids," J. Sound Vib. **194**(3), 429–438 (1995).

[11]I. A.Alshaikh, D. Turhan, and Y. Mengi, "Transient waves in viscoelastic cylindrical layered media," Eur. J. Mech. A/Solids **21**, 811–830, (2002).

[12]U. Basu and A. K. Chopra, "Perfectly matched layers for time-harmonic elastodynamics of unbounded domains: Theory and finite-element implementation," Comput. Methods Appl. Mech. Eng. **192**, 1337–1375 (2003).

[13]E. Bécache, S. Fauqueux, and P. Joly, "Stability of perfectly matched layers, group velocities and anisotropic waves," J. Comput. Phys. **188**, 399–433 (2003).

[14]G. R. Liu and S. S. Quek Jerry, "A non-reflecting boundary for analysing wave propagation using the finite element method," Finite Elem. Anal. Design **39**, 403–417 (2003).

[15]M. Castaings, C. Bacon, B. Hosten, and M. V. Predoi, "Finite element predictions for the dynamic response of thermo-viscoelastic material structures," J. Acoust. Soc. Am. **115**(3), 1125–1133 (2004).

[16]B. Hosten and M. Castaings, "Finite elements methods for modeling the guided waves propagation in structures with weak interfaces," J. Acoust. Soc. Am. **117**(3), 1108–1113 (2005).

[17]F. Simonetti and P. Cawley, "A guided wave technique for the characterization of highly attenuative viscoelastic materials," J. Acoust. Soc. Am. **114**(1), 158–165 (2003).

[18]J. Ma, F. Simonetti, and M. Lowe, "Scattering of the fundamental torsional mode by an axisymmetic layer inside a pipe," J. Acoust. Soc. Am. (submitted).

[19]J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1980).

[20]R. M. Christensen, *Theory of Viscoelasticity: An Introduction* (Academic Press, New York, 1971).

[21]M. Drozdz, M. J. S. Lowe, P. Cawley, L. Moreau, and M. Castaings, "Efficient numerical modeling of absorbing regions for boundaries of guided waves problems," in *Review of Progress in Quant. Non Destruct. Eval.*, edited by D. O. Thompson and D. E. Chimenti, American Institute of Physics, New York, Vol. 25, 2006 (in press).

[22]A. L. Shuvalov, "The frobenius power series solution for cylindrically anisotropic radially inhomogeneous elastic materials," Q. J. Mech. Appl. Math. **56**(3), 327–345 (2003).

[23]I. Mirsky, "Wave propagation in transversely isotropic circular cylinder—Part I: Theory," J. Acoust. Soc. Am. **37**(6), 1016–1021 (1965).

[24]B. Pavlakovic, M. Lowe, D. Alleyne, and P. Cawley, "DISPERSE: A general purpose program for creating dispersion curves," *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.

[25]K. J. Bathe, *Finite Element Procedures in Engineering Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1982).

[26]C. H. Gür and B. Aydinmakina, Non-destructive characterisation of Nodular Cast Irons by Ultrasonic method, WCNDT 2000 proceedings, http://www.ndt.net

[27]M. Castaings and B. Hosten, "The use of electrostatic, ultrasonic, air-coupled transducers to generate and receive Lamb waves in anisotropic, viscoelastic plates," Ultrasonics **36**, 361–365 (1998).

[28]M. Cinquin, M. Castaings, B. Hosten, P. Brassier, and P. Pérès, "Monitoring the moisture content in carbon-epoxy plates using Lamb waves," NDT & E Int. **38**(1), 37–44 (2005).

[29]M. Castaings, B. Hosten, and D. François, "The sensitivity of surface guided modes to the bond quality between a concrete block and a composite plate," Ultrasonics **42**, 1067–1071 (2004).

Bulk conical and surface helical acoustic waves in transversely isotropic cylinders; application to the stiffness tensor measurement

M. Perton and B. Audoin[a)]
Laboratoire de Mécanique Physique, UMR CNRS 8469, Université Bordeaux 1, 33405 Talence, France

Y. D. Pan
Institute of Acoustics, Tongji University, 200092, Shanghai, P. R. China

C. Rossignol
Laboratoire de Mécanique Physique, UMR CNRS 8469, Université Bordeaux 1, 33405 Talence, France

(Received 4 January 2006; revised 29 March 2006; accepted 29 March 2006)

A point-source-point-receiver technique, based on laser generation and laser detection of acoustic waves, allows determination of mechanical properties of anisotropic cylinders. The anisotropic nature of the material and the geometry of the samples make the acoustic signature difficult to interpret. In addition to multiple surface waves, quasi-longitudinal and quasi-shear bulk waves are diffracted and acoustic rays are reflected with or without mode conversion at the cylinder surface. Moreover both bulk and surface diffracted waves have a dispersive behavior. To bypass the intricacies, wave fronts are synthesized with signals provided by scanning a straight line on the cylinder with the laser point source. Conical waves propagating in the volume and helical waves propagating along the surface are then numerically produced. The recovery of the stiffness-tensor components is based on the inversion of the bulk waves, phase velocities. The method is presented and applied to signals simulated or experimentally recorded for a composite material. The five independent stiffness coefficients of the hexagonal symmetry are thus measured with waveforms provided by a single scan along the cylinder surface. The method provides a unique mean for the noncontact measurement of elastic properties of cylindrical parts. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2198182]

PACS number(s): 43.35.Cg, 43.20.El [YHB] Pages: 3752–3759

I. INTRODUCTION

Cylindrical parts made of composite materials are in particular favor for the design of rotating axes or cables when enhanced mechanical properties and lightweights are desired. High performances are achieved with manufactured rods made of fibers embedded in a matrix and aligned along the cylinder axis. There is thus an increasing demand on measuring nondestructively the elastic constants of cylinders showing the hexagonal symmetry.

There have been extensive studies of the acoustic scattering characteristics of a transversely isotropic cylinder to determine its elastic properties. As a typical example, resonance acoustic spectroscopy (RAS), e.g., the study of resonance effects present in acoustic echoes of an elastic target, was proposed to this purpose.[1] However, the immersion in a fluid of samples that are not allowed to do so limits the capability of RAS. Another technique, based on the measurement of longitudinal and torsional resonance frequencies of samples of known geometry, has permitted one to determine nondestructively the elastic constants of various crystals at high temperatures.[2,3] Free vibrations of a transversely isotropic finite cylindrical rod have also been analyzed for the purpose of elastic constants determination.[4] But these techniques require proper machining of samples and fine identification of the resonance peaks. Moreover, although laser interferometry was used for resonance measurements,[5] noncontact excitation has not been achieved for these techniques up to now.

The above limitations are circumvented when the ultrasonic waves are generated and detected with lasers. Recently, several studies have been focused on applications of surface acoustic waves generated on cylinders by this laser ultrasonic technique. Surface residual stress was estimated by measuring the velocities of the Rayleigh wave for steel rods receiving different heat treatments.[6] Surface breaking defects were also detected with surface waves propagating on a cylinder surface.[7] Due to the complexity of the cylinder Green function, very few publications are related to the analysis of bulk waves for the nondestructive evaluation purpose. Recently, the authors have shown that the two stiffness coefficients of an isotropic rod can be measured by analyzing the bulk acoustic waves generated with a laser line source.[8] To their knowledge, noncontact and nondestructive measurement of the five stiffness coefficients of a transverse isotropic cylinder has never been reported.

This paper follows the recent work[9] concerning the modeling of bulk and surface acoustic waves generated by a laser point source in a transverse isotropic cylinder. The main features of the intricate acoustic signature are briefly re-

[a)]Electronic mail: b.audoin@lmp.u-bordeaux1.fr

 0001-4966/2006/119(6)/3752/8/$22.50

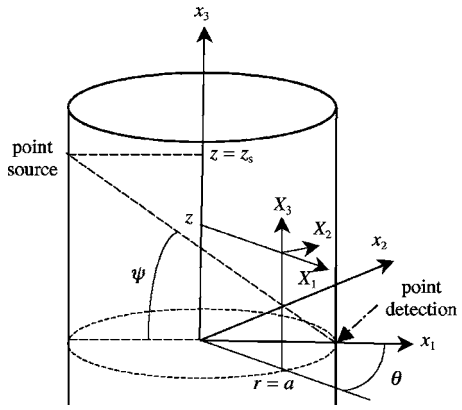


FIG. 1. Geometry and frames.

viewed in Secs. II and III of the present paper. Then Sec. IV depicts numerical synthesis of the bulk conical and surface helical waves that would have been actually generated if a broadband point source had moved along the cylinder surface at a super-ultrasonic velocity. Since the so-synthesized bulk conical waves propagate at phase velocity, the material stiffness coefficients are merely identified in Sec. V with a standard identification process based on the Christoffel equation.

II. POINT SOURCE AND ANISOTROPIC CYLINDER

A. Frames, material, and experimental setup

The geometry shown in Fig. 1 is that of a cylinder of infinite length. Cartesian coordinates (x_i), $i=1,\ldots,3$, are considered as well as cylindrical coordinates (r, θ, z) together with the local frame (X_i), $i=1,\ldots,3$. Directions X_1 and X_2 correspond to the radial and ortho radial axes of the cylindrical frame, respectively, and direction X_3 coincides with the axis z.

The cylinder is made of a composite material with carbon fibers embedded along the z direction in an epoxy matrix. The radius of the sample is $a=3.0$ mm and the mass density is $\rho=1.6$ kg/cm^3. The principal axes of the transverse isotropic symmetry coincide with axes X_i, $i=1,\ldots,3$, and the direction of symmetry is X_3. In the remainder, coefficients C_{ij}, $i, j=1,\ldots,6$, stand for the components of the stiffness tensor in contracted Voigt notation. Thus C_{11} and C_{66} denote the two independent elastic constants related to the isotropic plane perpendicular to the z axis. Imaginary parts of the stiffness coefficients account for a dispersion effect that may be due to the viscoelastic behavior of the matrix, or to the geometric dispersion caused by the heterogeneous nature of the material, or to a combination of both.[10] Stiffness coefficients of plates made of such an unidirectional carbon epoxy material were measured with ultrasonic methods.[11,12] Typical values are reported as reference data in Table I.

The beam of a Nd:YAG laser, with a pulse burst duration of about 10 ns, is focused on the cylinder surface by a set of lenses to form a point illumination. The spot diameter on the sample surface is about 0.1 mm. The electromagnetic-wave/material interaction causes the generation of bulk and surface acoustic waves.[13] By using a system composed of moveable mirrors, the setup allows displacement of the source along the z axis of the cylindrical frame. A CW laser interferometer, which measures the normal surface displacements, is used to detect the wave arrivals.[14] The detection laser beam is also focused down to a spot size of about 0.1 mm. A 10-μm-thick aluminum layer with a lateral size about 2 mm was stuck on the sample surface to improve signal-to-noise ratio. As shown in Fig. 1, the point detection is assumed to stand at a location $(r=a,\ \theta=0,\ z=0)$, whereas the coordinates of the source position are such that $(r=a,\ \theta=\pi,\ z=z_s)$. Thus the point-source to point-detection direction lies in the plane (x_1, x_3). In the remainder the angle between this direction and axis x_1 is denoted ψ.

B. Waves generated by a single point source

The acoustic waves generated on the surface and in the bulk of the material are now discussed. Since the material shows anisotropic behavior, three bulk modes can be generated. In the following, the quasi-longitudinal and quasi-shear modes whose polarizations lie in the plane (X_1, X_3) are indicated as (L) and (T_q), respectively, and the pure shear mode with a polarization along the X_2 direction is denoted (T_p).

When a single point source is considered, transient divergent bulk waves are generated in the solid, for which the propagation in an elastic medium is governed by the group-velocity surfaces. The directions of the acoustic-energy flux can be determined from the slowness surfaces by means of ray theory.[15] Owing to transverse isotropy, the tridimensional shape of the slowness surfaces is easily deduced from the section in plane (X_1, X_3) shown in Fig. 2, as long as invariancy for any rotation around the z axis is kept in mind. The direction of the energy flux at any point of the slowness

TABLE I. Stiffness tensor components (GPa) considered for signal calculations, stiffness coefficients calculated from signals simulated for the thermoelastic and the ablation regimes, and measured values from experimental signals recorded in both regimes.

	C_{11}	C_{13}	C_{33}	C_{44}	C_{66}
Used for signal calculations	12	6	130	6	3.5
Recovered from signals simulated for the thermoelastic regime	12.4	9	145	6.0	3.5
Recovered from signals simulated for the ablation regime	12.5	6.7	135	5.9	3.8
Measured in thermoelastic regime	12.7±0.1	4.8±0.5	118±3	7.4±0.1	3.3±0.1
Measured in ablation regime	12.6±0.1	2.4±0.4	120±3	7.6±0.2	3.4±0.1

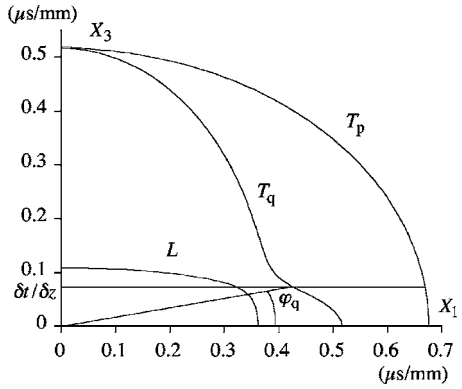


FIG. 2. Section of the slowness surfaces with plane (X_1, X_3) for the carbon-epoxy material.

surface is along the normal at that point. The anisotropy of the medium thus leads to the characteristic shape of the group velocity diagram for the quasi-shear mode, which contains cuspidal edges where the curve folds back on itself.[16] This caustic arises from the profile of the slowness curve, which is successively convex, concave, and convex. Thus for directions contained between the two cusp extremities, up to three contributions from this quasi-shear mode can be observed. Again, invariancy of this characteristic feature for any rotation around the axis z must be underlined.

In addition to the direct bulk longitudinal and shear wave arrivals, waves reflected off the cylinder free surface, with or without mode conversion, may be detected. Possible reflections of the acoustic rays are illustrated in Fig. 3(a) and 3(b). The relationship between incident and reflected rays is not straightforward when reflection of the transient divergent waves occurs with mode conversion. As discussed in Ref. 17 for a plate interface, the Huygens construction or the Fermat principle can be considered for modes without caustic only.

Surface waves and pseudo-surface waves propagating along the curved surface may also be detected. These waves propagate along the surface at group velocity and, owing to anisotropy, multiple arrivals may be observed for some ray orientations along the surface. Moreover, it propagates in either clockwise or counter-clockwise direction along odd

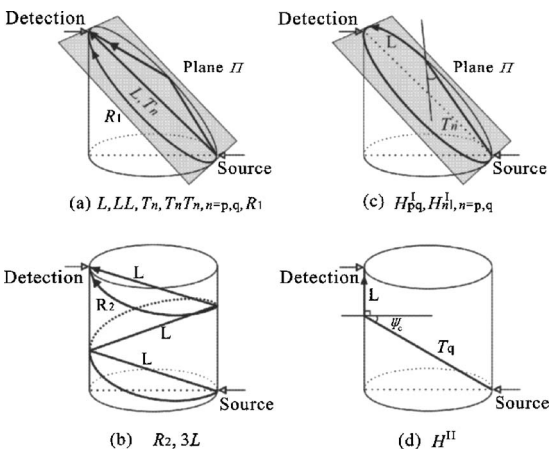


FIG. 3. Some ray trajectories of waves propagating in the bulk and on the surface of the cylinder.

numbers of half round-trip as illustrated, for instance, for the Rayleigh wave with R_1 and R_2 in Figs. 3(a) and 3(b), respectively.

Head waves resulting from the conversion of a bulk wave into a grazing wave propagating on the surface may additionally be detected. Let Π denote the plane that contains the point-source to point-detection direction and whose normal lies in the plane containing this direction and the axis z. For the purpose of clarity, two kinds of head waves are considered separately. As illustrated in Fig. 3(c), head waves of the first kind have their ray trajectory included in plane Π, with the grazing path along the circumference of the elliptic intersection of the cylinder with plane Π. The observed head waves are resulting from conversion at the critical angle of (i) a pure shear wave into a grazing quasi shear wave or (ii) shear wave, either T_q or T_p, into a grazing longitudinal wave. These head waves, shown in Fig. 3(c), are denoted H^I_{pq} for mode conversion case (i) or H^{I}_{nl} for mode conversion case (ii). Index $n=p, q$ indicates that the head wave arises from conversion of a pure (T_p) or of a quasi (T_q) shear wave, respectively. Head waves of the second kind result from mode conversion at a critical angle in plane (x_1, x_3), as illustrated in Fig. 3(d). Owing to the equality of the two shear wave slownesses along the direction X_3, as shown in Fig. 2, conversion of a shear wave into another grazing shear wave is not possible. Conversion of shear waves into a grazing longitudinal wave along the z direction is now considered. Since polarization of the pure shear mode lies along x_2, it cannot convert into a longitudinal wave at the free interface. Therefore only one head wave of this second type, arising from conversion of the quasi-shear into the longitudinal mode, may exist. It is denoted H^{II} in Fig. 3(d). This head wave H^{II} can only be observed for z_s positions such that angle ψ is greater than the critical angle ψ_c for the conversion of the quasi-shear mode into the quasi-longitudinal mode.

III. SIGNAL SIMULATIONS

A calculation scheme of the anisotropic cylinder Green function was already presented in a recent paper.[9] The assumptions and principles are now briefly mentioned since calculated signals will be considered in the following sections.

Modeling of the acoustic source is first presented. Assuming small optical penetration with respect to the acoustic wavelengths, volume source effects can be neglected. The source is thus considered to lie on the cylinder surface and it only interferes in the boundary conditions. Since a spherical lens is used to focus the laser beam, the spatial acoustic source can be modeled as a point on the surface. Apart from this point, the cylinder surface is supposed to be free. At low incident power densities, thermal expansion generates acoustic waves. Neglecting thermal diffusion, the thermo-elastic process is represented by a force dipole during a Heaviside step function in time. For high power densities, vaporization of a small amount of surface material occurs. Thereby, the source is modeled as a normal loading during a delta function of time in this ablation regime.

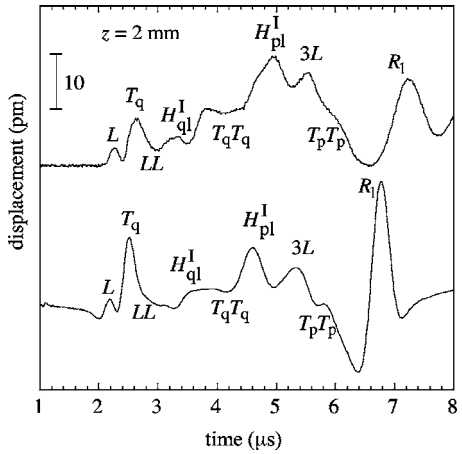


FIG. 4. Experimental (upper curve) and calculated (lower curve) displacements in the ablation regime for a point detection position ($r=a$, $\theta=0$, $z=0$) and a point source position ($r=a$, $\theta=\pi$, $z_s=2$ mm).

To calculate the acoustic field, wave motion and boundary equations are linearized with a three-dimensional Fourier transform. The Fourier dual variables of time t and space variables (z, θ) are ω and (k, ν), respectively. On noting $\nu=k_\theta a$, k and k_θ stand for the components along the X_3 and X_2 directions, respectively, of the wave vector **k**. An imaginary part of the stiffness coefficients accounts for the viscoelastic material behavior. Moreover, the profile of the source along the surface as well as the laser pulse width are accounted for by multiplying the displacement in the Fourier domain with convenient Gauss functions.[7] The main equations are derived in Ref. 9, where the inversion of the transform is also described.

Experimental and calculated signals are shown in Fig. 4 for a source position such as $z_s=2$ mm. In this figure, the component of the displacement normal to the surface is presented for the ablation regime. The stiffness coefficients considered for calculations will be specified in Sec. IV. The simulated waveform has been scaled vertically by a constant corresponding to the source strength to bring the amplitudes of the two signals into agreement. The shapes of the calculated and experimental displacements are similar and the spreading of the signals due to viscosity appears clearly. Direct quasi-longitudinal (L) and quasi-shear waves (T_q) are observed in these waveforms. Owing to phonon focusing the quasi-shear wave shows high amplitude for this point-source to point-detection direction oriented nearby the cuspidal edges.[16,18] The pure shear mode is not observed in these waveforms. This would hold for any z_s position since the measured component of the surface displacement lies perpendicular to the wave polarization. This direct pure shear wave could only be observed by diffraction effect in the near field, with small amplitude.

Waves resulting from the reflection of each of the three modes at the cylinder free surface according to ray paths shown in Fig. 3(a) are also observed. They are reflected without mode conversion for this source position $z_s=2$ mm, and they are referred as LL, T_pT_p, and T_qT_q in Fig. 4 for the quasi-longitudinal, pure, and quasi-shear modes, respectively. Appearance of the wave arising from reflection of the pure mode is worth being underlined since this mode is not observed for a direct trajectory. Other reflections of the quasi-longitudinal wave ($3L$) with ray path in plane (x_1, x_3) arise in both calculated and experimental waveforms.

Two head waves of first type H^I_{nl}, $n=p$, q [Fig. 3(c)], are also observed with large amplitude, whereas that of second type, H^{II} [Fig. 3(d)], cannot be measured since $\psi<\psi_c$ for this source position. Note finally the large amplitude of the Rayleigh wave R_1 with ray trajectory shown in Fig. 3(a). The small delay between the calculated and the measured Rayleigh waves, as well as the smoothing of the measured Rayleigh wave, is due to the thin aluminum layer used for signal detection.

IV. SYNTHESIS OF CONICAL WAVES

The waveform in Fig. 4 is very rich as it shows the signature of numerous waves with different natures and several ray paths. However, waveform interpretation is not a trivial matter since acoustic energy travels at a group velocity that differs from phase velocity owing to anisotropy. For this reason, interest is now turned to the numerical synthesis of the signal $s(t)$ that would be experimentally obtained if an array of laser point sources was used. For this purpose a set of N source positions is considered equally spaced with a constant step δz. Each signal $s_i(t)$, either calculated or measured, is recorded for the fixed detection position. Then a constant time delay δt is applied between two successive signals. Finally the synthesized signal associated to this delay is built by summing the shifted signals according to the relation:

$$s(t)=\sum_{i=-N}^{N} s_i(t+i\delta t). \tag{1}$$

For convenient sampling conditions,[19] the so synthesized signal is the signal that would have been recorded if the source had traveled along the surface with the slowness $\delta t/\delta z$. Accordingly to Snell's law the incident and the refracted waves at a plane interface must have the same projected wave number along the interface. Therefore, the source slowness along the z direction $\delta t/\delta z$ fixes the projection of wave vectors in this direction.[19] Thus, whatever the polarization, all synthesized bulk waves have the same wave vector projection in direction z. It will be denoted as k_s in this paper. Owing to the symmetry of the phase slowness surfaces, e.g., isotropy in the plane (x_1, x_2), projection of the wave vector in this plane holds a unique value for each mode. Thus the wave vectors lie along three cones whose symmetry axes are z and whose solid angles are $\pi/2-\varphi_m$, with φ_m such that

$$\frac{\sin\varphi_m}{V_m}=\frac{\delta t}{\delta x}. \tag{2}$$

Index $m=l$, p, q refers to the quasi-longitudinal, shear, and quasi-shear mode, respectively, and V_m stands for the phase velocity of the concerned mode. Angle φ_q is shown in Fig. 2 to provide a geometric presentation of Eq. (2). The corresponding bulk wave fronts are cones with symmetry axis ($r=a$, $\theta=\pi$, z) lying along the cylinder surface and with solid angles φ_m. The so synthesized wave fronts will be named as

bulk conical waves in the remainder. The angle of each conical refraction can merely be tuned continuously by conveniently choosing the delay δt, according to Eq. (2). The wave front propagates in the plane (x_1, x_3) at a phase velocity also fixed by the delay δt. It still propagates at group velocity in the plane (x_1, x_2) but owing to transverse isotropy, phase and group velocities are equal in this plane. Phonon focusing is now avoided as well as multiple direct arrivals of the quasi-shear mode.

These bulk conical waves are reflected at the cylinder free surface[20] with or without mode conversion, giving rise to several conical wave fronts with symmetry axes lying along the cylinder surface. According to Snell's law, the solid angles of the reflected conical wave fronts are angles φ_m satisfying Eq. (2).

The question arises as to what are the synthesized head waves. Let us firstly consider the projections in the plane (x_1, x_2) of the wave vectors related to a synthesized shear conical wave. There always exists one tangent to the cylindrical free surface that intersects the direction of the projected wave vector with an angle equal to the critical angle of the mode conversion: (*i*) T_p into T_q or (*ii*) either T_q or T_p into L. It gives rise to a grazing quasi-shear wave or to a grazing longitudinal wave, respectively. The component along x_3 of the grazing wave vector is still that, named k_S, fixed by the shear wave cone, e.g., by the source slowness. The corresponding wave fronts are thus helicoidal. Three head waves of this first type may exist and they are denoted as H^{I}_{pq} for mode conversion case (*i*) and H^{I}_{nl}, $n=p$, q, for mode conversion case (*ii*). Projection in the plane (x_1, x_3) is secondly considered. For each shear mode only one source slowness $\delta t/\delta z$ fixes angle φ_n equal to the critical value that gives rise to a grazing longitudinal wave propagating in the X_3 direction. Hence a head wave of the second type H^{II} can be synthesized for only one value of the source slowness along the surface.

Surface and pseudo-surface waves are also synthesized. They result from the coherent combination of surface waves generated by each point source. They propagate at phase velocity in a direction perpendicular to the curved line defined by the intersection of the cone with the cylinder. Surface phonon focusing disappears and the consequential multiple arrivals of the Rayleigh wave cannot occur anymore, so simplifying the acoustic response.

The signal synthesized for source slowness along the surface $\delta t/\delta z$ of 75 μs/m is shown in Fig. 5. A set of $N=41$ signals is calculated with a distance δz between two successive sources of 0.2 mm. As shown in Fig. 2, the source is super-ultrasonic for the three acoustic modes. Therefore three cones can be refracted from the interface. Waves observed in Fig. 4 are also visible in the synthesized waveform, Fig. 5. However, they may arrive in a different order since they have traveled at phase velocity instead of group velocity. Mapping of the phase velocities of three wave modes will be considered in the following section for stiffness coefficients measurement purpose.

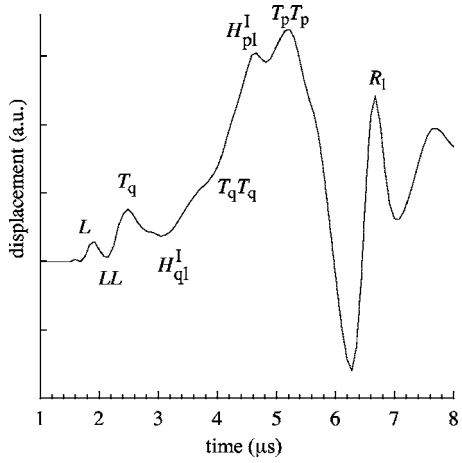


FIG. 5. Signal $s(t)$ built for source slowness $\delta t/\delta z$ along the cylinder surface equal to 75 μs/m. Signals calculated for the ablation regime were used for the numerical synthesis.

V. MATERIAL CHARACTERIZATION

Interest is now turned to the inverse problem of measuring the five independent stiffness tensor components, namely C_{11}, C_{33}, C_{44}, C_{66}, and C_{13}. The method is first applied to the set of simulated signals $s_i(t)$ considered in the previous section. By tuning the value of the parameter δt, conical waves are synthesized with controlled solid angles φ_m, $m=l$, p, q [see Eq. (2)]. A set of synthesized signals $s(t)$ is thus calculated, providing conical waves for chosen directions distributed in a wide angular sector.

Let us now consider the question of wave recognition. The following considerations allow wave identification. First, for a source position differing from $\theta=\pi$, the acoustic waves propagating along asymmetric ray paths on the surface attain the detection point at different times, and the corresponding peak is twofold. Therefore, waves propagating along the cylinder surface, such as Rayleigh waves and head waves of the first type, can obviously be distinguished merely by calculations or experimental measurements performed for a source position slightly shifted from $\theta=\pi$. This does not apply for head waves of the second type, but they can be recognized with the unique features that they are only detected for angles ψ greater than a critical value ψ_c. Second, waves having undergone reflection once at the free surface are easily distinguishable since their phase is shifted by π in comparison with that of the corresponding direct bulk wave (see LL and T_qT_q in Figs. 4 and 5).

For each synthesized waveform, signal processing[21,22] allows the accurate measurement of arrival times t_m of the direct quasi-longitudinal and quasi-shear bulk conical wave fronts. Let us now relate the phase velocity $V_m(\varphi_m)$ to the measured arrival time t_m. Figures 6(a) and 6(b) show the geometry needed for the derivation of this relationship. The dashed line in Fig. 6(b) is the intersection of a conical wave front with the plane (x_1, x_3). At that stage, it is important to note firstly that because of the antisymmetrical construction of the wave fronts with respect to time, Eq. (1), the position of the dashed line in Fig. 6(b) is consistent with time $t=0$; and secondly that the transit time t_m equals the time at which the wave front symbolized by this dashed line reaches point B_1 in Fig. 6.[19] Then the phase velocity is simply obtained with

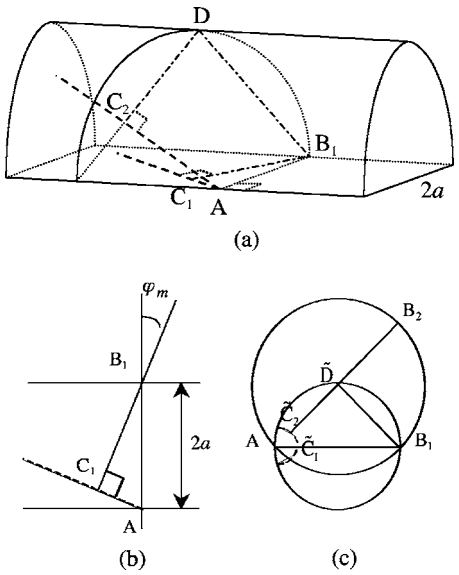


FIG. 6. Path of the direct and reflected conical waves; B_1: point detection, A: top of the conical front at time $t=0$. (a) AC_1 and AC_2 are two primitive lines of the same conical front at time $t=0$. C_1B_1: direct path of this conical front. C_2DB_1: path of the conical front with reflection at the free surface, either LL, T_qT_q, or T_pT_p. (b) Path C_1B_1 of a direct conical front (dashed line) in plane (X_1, X_3). (c) $\tilde{C}_2\tilde{D}B_1$ is the projection in plane (X_1, X_2) of the path of the reflected conical front (C_2DB_1) associated to the cylinder of radius a. $\tilde{C}_2B_2$ is the projection of the equivalent direct path in a cylinder of radius $\sqrt{2}a$.

$$\frac{\cos\varphi_m}{V_m} = \frac{t_m}{2a}. \tag{3}$$

Let us now additionally describe how the phase velocity $V_m(\varphi_m)$ of any mode can also be calculated with the transit time t_{mm} of the waves LL, T_pT_p, or T_qT_q reflected without mode conversion at the free surface. These waves propagate along the path C_2DB_1 drawn in Fig. 6(a). Figure 6(c) shows its projection $\tilde{C}_2\tilde{D}B_1$ in the plane (x_1, x_2). The length of the projected path equals that, $\tilde{C}_2B_2$ in Fig. 6(c), along which a direct wave front would have propagated through a cylinder with radius $\sqrt{2}a$. Therefore, the phase velocity $V_m(\varphi_m)$ can be calculated with the data corresponding to the reflected wave simply by using Eq. (3) in which t_m/a is changed into $t_{mm}/\sqrt{2}a$. Equations (2) and (3) then provide the projections in the plane (x_1, x_3) of the phase slownesses for these waves undergoing reflection at the free surface with wave vectors out of the plane (x_1, x_3). This point is of particular interest for the pure shear mode whose polarization makes the direct contribution T_p difficult to measure.

Waveforms synthesized with signals simulated for either the thermoelastic or the ablation regime have been processed to calculate the arrival times of both direct and reflected bulk conical waves. Intricacy of the waveforms, including waves overlapping, has made the arrival times processing difficult. However, corresponding slowness data are shown with crosses and circles in Fig. 7 for the thermoelastic regime only. The solid lines in Fig. 7 represent slowness curves obtained by solving the Christoffel equation using stiffness coefficients reported in Table I as reference values. Very good agreement between the recovered slownesses and reference counterparts is obtained for the quasi-longitudinal, pure, and quasi-shear modes.

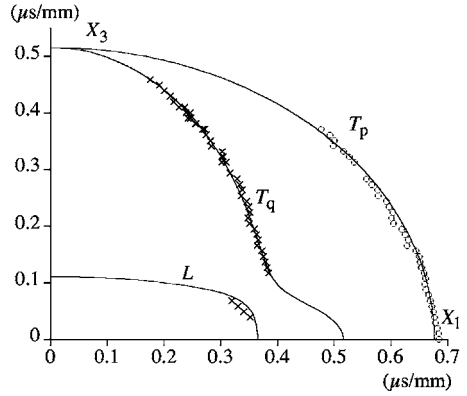


FIG. 7. Slowness data deduced from both direct (L, T_q: crosses) and reflected (T_p: circles) bulk conical waves. Processed signals were simulated for the thermoelastic regime. Solid lines are for slownesses calculated with reference stiffness coefficients used for signal simulations.

Let us now describe the stiffness coefficients identification from phase velocity data. For wave vectors included in a principal plane of a material showing hexagonal symmetry, say (X_1, X_3), the Christoffel equation factorizes into

$$(\lambda_m - \Gamma_p)(\lambda_m - \Gamma_q)(\lambda_m - \Gamma_l) = 0, \tag{4}$$

with

$$\Gamma_p = \Gamma_{22},$$
$$2\Gamma_q = \Gamma_{11} + \Gamma_{33} - \sqrt{(\Gamma_{11} - \Gamma_{33})^2 + 4\Gamma_{13}^2}, \tag{5}$$
$$2\Gamma_l = \Gamma_{11} + \Gamma_{33} + \sqrt{(\Gamma_{11} - \Gamma_{33})^2 + 4\Gamma_{13}^2}.$$

In Eq. (4), $\lambda_m = \rho V_m^2$, $m=l, p, q$ denote eigenvalues of Christoffel tensor; in Eqs. (5), Γ_{ij} stand for the components of this tensor. Nullity of the first, second, and third terms in Eq. (4) occurs for velocities V_m, $m=p, l, q$, respectively. Coupling of the quasi-longitudinal and quasi-shear velocities is obvious in view of the second and third equations in Eqs. (5). Moreover, assuming transverse isotropy, the pure shear velocity V_p does not uncouple with V_l or V_q since Γ_{22} depends on the stiffness coefficient C_{44}, and so do the other nonzero tensor components Γ_{11}, Γ_{13}, and Γ_{33}. Annulling each term in Eq. (4) for phase velocities V_m processed for several cone angles φ_m, e.g., several source slownesses $\delta t/\delta z$, provides an over determined system of nonlinear equations[23] whose unknowns are the five stiffness coefficients. At this point it is important to underline that the coefficient C_{66} enters in the Christoffel coefficient Γ_{22} only, thus measurement of this stiffness coefficient requires that phase velocities of the pure shear mode be measured.

The values of the stiffness coefficients recovered with the slowness data shown in Fig. 7, e.g., for the thermoelastic regime, are reported in Table I. Recovered values of the stiffness coefficients are close to their reference counterparts. Note that stiffness coefficients C_{44} and C_{66} are obtained with good accuracy. Since sensitivity of phase velocities to stiffness coefficients is the least for off-diagonal coefficients,[24]

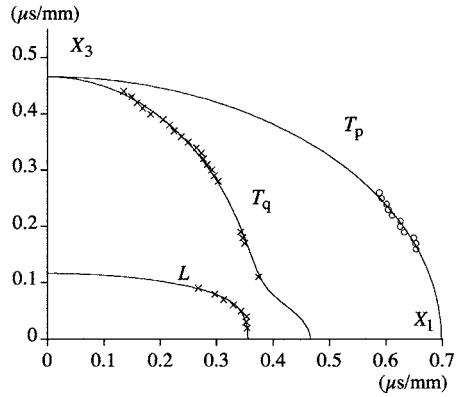


FIG. 8. Slowness data deduced from both direct (L, T_q: crosses) and reflected (T_p: circles) bulk conical waves. Processed signals were measured in the thermoelastic regime. Solid lines are for slownesses calculated with optimum stiffness coefficients shown in Table I.

maximum relative error is for the off-diagonal coefficient C_{13}. Slowness data calculated with signals simulated for the ablation regime have been processed also. The identified stiffness coefficients are shown in Table I. Still, recovered values of the stiffness coefficients are close to their reference counterparts, and again the highest relative error holds for the off-diagonal coefficient C_{13}. However, since the quasi-longitudinal wave generation is favored in the ablation regime, identification of the stiffness coefficients sensitive mainly to slownesses of quasi-longitudinal mode, namely C_{33} and C_{13}, is improved. For the reversed reason, e.g., favored generation of the shear and quasi-shear modes in the thermoelastic regime, measurement accuracy of shear coefficients, C_{44} and C_{66}, is better with signals simulated for the nondestructive regime.

Good agreement of the above results with the reference values led us to perform stiffness measurements with experimentally recorded signals. The process is now applied to a set of $N=51$ waveforms actually recorded by moving the laser point source along the cylinder surface with increments of δz equal to 0.2 mm. Since the method requires an array of identical laser point sources, several signals are averaged for each position of the laser beam to ensure reproducibility of the acoustic source and to prevent any random variation of the laser intensity. The thermoelastic regime is preferred for acoustic generation to avoid any burning of the cylinder surface. Despite low signal-to-noise ratio in this regime when a Nd:Yag laser is used,[25] slowness data were obtained for the three bulk propagation modes by processing experimental signals. Figure 8 shows superposition of these slowness data with the slowness curves calculated with Christoffel equation in which optimum stiffness coefficients are considered. These measured coefficients provided by the optimization scheme are reported in Table I together with their 90% confidence interval.[24] The confidence intervals are not representative of the exact errors because possible systematic errors associated with the slowness measurements are not taken into account. However, they are indicative of the uncertaincy on the stiffness coefficients estimated from the set of measured phase slownesses. Note that the measured values must not be compared to reference values given in Table I since the latter are approximate values chosen for signal simulations. Signals measured in the ablation regime have been processed also and relevant stiffness coefficients are listed in Table I. Since amplitude of the quasi-longitudinal bulk wave is higher is this regime, corresponding phase data could be measured with good accuracy and for angles φ_l included in a larger sector. The stiffness coefficients measured for the two regimes are almost identical except for the off-diagonal coefficient C_{13}. The low accuracy of this coefficient identification was shown with the simulation results and, again with experimentally recorded signals, this coefficient shows maximum relative uncertainty. Better recovering of shear coefficients C_{44} and C_{66} in the thermoelastic regime cannot be ascertained with experimental data owing to low signal-to-noise ratio.

VI. CONCLUSION

Laser ultrasonics with their unique feature of pointlike generation and detection has allowed noncontact and complete stiffness characterization of a cylinder rod made of anisotropic composite material. The generated transient divergent waves are difficult to analyze in anisotropic media. However, the intricacies associated with divergent waves could be circumvented. First, the point source was scanned along the cylinder surface to provide acoustic signatures that would have been recorded if an array had been used. Second, the signals deriving from point sources were shifted in time and summed to synthesize bulk conical and line surface waves that would have been generated if the point source had traveled at super-ultrasonic velocities along the cylinder surface. Bulk conical direct and reflected waves allow one to measure the quasi-longitudinal, quasi-shear, and also pure shear mode phase velocities for a large panel of wave vectors orientations. Despite bulk and surface waves overlapping, the coefficients of the material stiffness tensor were measured from the phase velocities with good accuracy starting with either calculated or experimentally recorded signals. These results will pave the way for new laser ultrasonic applications such as on-line monitoring of cylinders extrusion or for noncontact in-site control of stiffness decreasing induced by damage in cylinder rods.

ACKNOWLEDGMENTS

This work was partially supported by a grant from Conseil Régional d'Aquitaine and the European Community. The authors thank Soficar Company for its support and for providing the samples. Y. D. Pan was supported by KC-Wong and CNRS fellowship and by the Natural Science Foundation of China under Grant No. 10234020.

[1]F. Honarvar and A. N. Sinclair, "Nondestructive evaluation of cylindrical components by resonance acoustic spectroscopy," Ultrasonics **36**, 845–854 (1998).
[2]Y. A. Burenkov and S. P. Nikanorov, "Elastic properties and binding forces in crystals with diamond and sphalerite lattices, "Sov. Phys. Solid State **26**, 1940–1944 (1984).
[3]W. Hermann and H.-G. Sockel, "Torsional and flexural vibrations of rod," in *Handbook of Elastic Properties of Solids Liquids and Gases* (Academic, London, 2001), Vol. **1**, pp. 299–327.
[4]C. P. Lusher and W. N. Hardy, "Axisymmetric free vibrations of a trans-

versely isotropic finite cylindrical rod," J. Appl. Mech. **55**, 855–862 (1988).
[5]H. Ogi, N. Nakamura, M. Hirao, and H. Ledbetter, "Determination of elastic, anelastic, and piezoelectric coefficients of piezoelectric materials from a single specimen by acoustic resonance spectroscopy," Ultrasonics **42**, 183–187 (2004).
[6]M. Qian, M. Duquennoy, M. Ouatouh, F. Jenot, and M. Ourak, "Laser ultrasonic characterization of surface residual stresses in steel rods," Rev. Prog. Quant. Nondestr. Eval. **20**, 1487-1493 (2001).
[7]D. Clorennec, D. Royer, and H. Walaszek, "Nondestructive evaluation of cylindrical parts using laser ultrasonics," Ultrasonics **40**, 783–789 (2002).
[8]Y. Pan, C. Rossignol, and B. Audoin, "Acoustic waves generated by a laser line pulse in cylinders; application to the elastic constants measurement," J. Acoust. Soc. Am. **115**(4), 1537–1545 (2005).
[9]Y. Pan, M. Perton, B. Audoin, and C. Rossignol, "Acoustic waves generated by a laser point pulse in a transversely isotropic cylinder," J. Acoust. Soc. Am. **119**(1), 243–250 (2006).
[10]D. H. Hurley and J. B. Spicer, "An investigation of the effects of material anisotropy and heterogeneity on pulsed laser-generated acoustic signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**(6), 1387–1395 (1999).
[11]B. Hosten, "Elastic characterization of orthotropic composite-materials from ultrasonic inspection through non principal planes," Rev. Prog. Quant. Nondestr. Eval. **10B**, 1437–1444 (1991).
[12]S. I. Rokhlin and W. Wang, "Double through-transmission bulk wave method for ultrasonic phase velocity measurement and determination of elastic constants of composite materials," J. Acoust. Soc. Am. **91**(6), 3303–3312 (1992).
[13]R. M. White, "Generation of elastic waves by transient surface heating," J. Appl. Phys. **34**(12), 3559–3567 (1963).
[14]D. Royer and E. Dieulesaint, "Optical detection of sub-Angstrom transient mechanical displacement," in *Proceedings of the 1986 IEEE Ultrasonics Symposium* (IEEE, New York, 1986), pp. 527–530.
[15]K. Aki and P. Richards, *Quantitative Seismology* (Freeman, San Francisco, 1980).
[16]R. L. Weaver, M. R. Hausser, and J. P. Wolfe, "Acoustic flux imaging in anisotropic media," Z. Phys. B: Condens. Matter **90**, 27–46 (1993).
[17]A. Mourad and B. Castagnède, "Détermination des temps d'arrivée de fronts d'onde acoustiques divergents dans une lame anisotrope," J. Phys. III **3**, 1911–1927 (1993).
[18]H. J. Maris, "Enhancement of heat pulse in crystal due to elastic anisotropy," J. Acoust. Soc. Am. **50**, 812–818 (1971).
[19]F. Reverdy and B. Audoin, "Elastic constants determination of anisotropic materials from phase velocities of acoustic waves generated and detected by lasers," J. Acoust. Soc. Am. **109**(5), 1965–1972 (2001).
[20]Y. Pan, C. Rossignol, and B Audoin, "Identification of laser generated acoustic waves in the 2D transient response of cylinders," J. Acoust. Soc. Am. **117**(6), 3600–3608 (2005).
[21]S. Guilbaud and B. Audoin, "Measurement of the stiffness coefficients of a viscoelastic composite material with Laser generated and detected ultrasound," J. Acoust. Soc. Am. **105**(4), 2226–223 (1999).
[22]A. Grossmann, M. Holschneider, R. Kronland-Martinet, and J. Morlet, "Detection of abrupt changes in sound signals with the help of wavelet transforms," in *Inverse Problems*, edited by P. C. Sabatier (Academic, New York, 1987), pp. 289–306.
[23]J. Roux, B. Hosten, B. Castagnède, and M. Deschamps, "Caractérisation mécanique des solides par spectro-interférométrie ultrasonore," Rev. Phys. Appl. **20**(6), 351–358 (1985).
[24]B. Audoin, S. Baste, and B. Castagnède, "Estimation de l'intervalle de confiance des contantes d'élasticité identifiées à partir des vitesses de propagation ultrasonores," C. R. Acad. Sci., Ser. II: Mec., Phys., Chim., Sci. Terre Univers **312**, 679–686 (1991).
[25]M. Dubois, P. W. Lorraine, R. J. Filkins, T. E. Drake, K. R. Yawn, and S.-Y. Chuang, "Experimental verification of the effects of optical wavelength on the amplitude of laser generated ultrasound in polymer-matrix composites," Ultrasonics **40** (1-8), 809–812 (2002).

A method for modeling polymer viscoelastic data and the temperature shift function

Walter M. Madigosky
Department of Physics, Catholic University of America, Washington, DC 20064–0001

Gilbert F. Lee and Jan M. Niemiec
Carderock Division, Naval Surface Warfare Center 9500 McArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 5 December 2005; revised 21 March 2006; accepted 22 March 2006)

Dynamic viscoelastic polymer data is traditionally time-temperature shifted to obtain a temperature shift function (TSF) and then fitted to various analytical models. The process of obtaining the TSF can introduce considerable procedural or operator bias. Nevertheless the Havriliak and Negami (HN) model using the TSF methodology can generally describe polymers that are rheologically simple. In this paper the "wicket" plot is utilized as an important tool in analyzing data, as it is completely independent of time-temperature shifting (TTS). Using the wicket plot the data is fit to the HN equation to determine the four material HN constants independent of TTS. Having obtained the complete spectra of dynamic properties the specific relaxation time (frequency) at each temperature is obtained by matching the HN curve to the experimental data at that temperature thus determining the TSF. The procedure is illustrated by analyzing computer-generated data with random error in modulus and loss and finally real data on a standard material. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195292]

PACS number(s): 43.35.Mr, 43.35.Fj [YHB] Pages: 3760–3765

I. INTRODUCTION

Viscoelastic materials are used extensively in managing the propagation of sound between structures. Such systems require specific frequency dependent mechanical properties in order to function in an optimal manner. The Havriliak-Negami (HN) model has been shown to be an excellent model for modeling the alpha relaxation that occurs during the rubber to glass transition in polymers and which controls the acoustic properties.[1] Polymers with widely varying properties have shown good agreement with this model.[2] These polymers are said to be rheologically simple and only one relaxation is observed within the measurable limits. In order to obtain a complete spectrum of modulus and loss values, from the rubbery to glassy region, it is found necessary to take data over a wide temperature and frequency range. Frequency alone cannot provide the range of values required due to instrument constraints. To obtain the complete spectrum from the multiple sets of frequency and temperature it is necessary to invoke the principle of time-temperature superposition. Thus a temperature-shift function (TSF) which shifts one set of frequency data onto a corresponding set of data taken at a different temperature is established. Traditionally this is the first step in analyzing the data. The shift is often performed by simply matching adjacent frequency curves and noting the shift in frequency required[3–5] or by computational curve fitting.[6] The difficulty with this is that no standard method is used and each operator introduces a unique set of errors or bias. Hence, different experimenters may arrive at different results for the same set of data.[7]

In this approach the starting point is the HN equation and the wicket plot. The wicket plot, well know in electrical engineering and other disciplines, provides a method of analysis independent of the frequency and temperature under which the data was obtained. By fitting the HN equation to the wicket data plot the four basic material parameters are determined without recourse to time-temperature shifting. Subsequently, the fifth material parameter can be determined by choosing a reference temperature from the actual data set and comparing that data with the HN frequency curve by selecting the characteristic frequency which best matches the data. Finally, other reference temperatures can be obtained by determining the frequency shift required for each temperature to bring the data into agreement with the HN frequency curve. This determines a TSF and a "master" curve or nomogram.[8] Note that the TSF is determined last and not first. As such it does not distort or introduce errors or bias into the master curve.

The details of the method are illustrated in two ways. First by considering idealized test data that was computer generated with a given amount of error. Next resonance apparatus data for the American National Standard Institute (ANSI) polymer is processed and presented. In both cases excellent results are obtained.

II. HAVRILIAK-NEGAMI MODEL

The single relaxation time model predicts a very narrow relaxation spectrum. The Cole-Cole model broadens the spectrum and the Davidson-Cole model introduces an asymmetry. Havriliak and Negami combined the latter two models to produce a model that was both broad and asymmetric. The Havriliak and Negami equation is

$$E^* = E_\infty + (E_0 - E_\infty)/[1 + (i\omega\tau)^\alpha]^\beta, \quad (1)$$

where E^* is the complex Young's modulus

 0001-4966/2006/119(6)/3760/6/$22.50

$$E^* = E' + iE'', \tag{2}$$

and E_0 is the low frequency or rubbery modulus, E_∞ is the high frequency or glassy modulus, α is a parameter governing the width of the relaxation, β is a parameter governing the asymmetry of the relaxation, τ is a relaxation time governing the position of the relaxation in the frequency domain, and $\omega = 2\pi f$ is the angular frequency. At a specific temperature

$$\tau_0 = 1/(2\pi f_0), \tag{3}$$

where f_0 is a characteristic frequency at a reference temperature T_0.

Solving for the real E' and imaginary E'', parts of Young's modulus, respectively,

$$E' = E_\infty + (E_0 - E_\infty)\cos(\beta\theta)/[1 + 2(\omega\tau)^\alpha\cos(\alpha\pi/2) + (\omega\tau)^{2\alpha}]^\beta, \tag{4}$$

and

$$E'' = (E_\infty - E_0)\sin(\beta\theta)/[1 + 2(\omega\tau)^\alpha\cos(\alpha\pi/2) + (\omega\tau)^{2\alpha}]^\beta, \tag{5}$$

where

$$\theta = \tan^{-1}[(\omega\tau)^\alpha\sin(\alpha\pi/2)]/[1 + (\omega\tau)^\alpha\cos(\alpha\pi/2)]. \tag{6}$$

The measure of damping effectiveness known as the loss factor is

$$\text{loss factor} = \tan\delta = E''/E', \tag{7}$$

Thus

$$\tan\delta = (E_\infty - E_0)\sin(\beta\theta)/\{E_\infty[1 + 2(\omega\tau)^\alpha\cos(\alpha\pi/2) + (\omega\tau)^{2\alpha}]^\beta + (E_0 - E_\infty)\cos(\beta\theta)\}$$

or

$$\tan\delta = (E_\infty/E' - 1)\tan(\beta\theta). \tag{8}$$

In this model it is assumed that the only temperature dependent parameter is the relaxation time and that only one polymer relaxation is involved. If there are multiple relaxation processes then a more complicated analysis or the use of more than one HN equation is required.

Finally, it is important to note that in order to obtain data over a wide frequency range the principle of time-temperature superposition is universally used to generate master curves. In the HN equation the variables appear only as a product of frequency and the relaxation time, which is a function of temperature. Hence a change in frequency is equivalent to a change in temperature. In a rheologically simple material a modulus value measured at frequency f and temperature T is equivalent to a value at reduced frequency f_R and temperature T_0. The reduced frequency can be much greater than the experimental frequency available (by a factor a_T) since measurements over a wide range of temperatures are more experimentally accessible. The shift factor can be expressed as the ratio of the relaxation time at temperature T to the relaxation time at a reference temperature T_0

$$a_T(T) = \tau(T)/\tau(T_0). \tag{9}$$

Since the complex modulus is a function of the product of $\omega\tau$ only, it can be written as

$$E^* = E\{f a_T(T)\tau(T_0)\}, \tag{10}$$

and the reduced frequency is defined as

$$f_R = f a_T(T). \tag{11}$$

Traditionally the shift factor has been obtained by shifting each isotherm along the log frequency axis until it partially overlaps the reference isotherm. This process is continued with all the isotherms, in sequence, both higher and lower than the reference temperature. Because shifting data generates this function, it is known as the temperature shift function. The result is a plot of log E' over a wide range of log reduced frequency values known as a master plot. The amount of shift required to produce overlap can be plotted as a function of temperature. The Williams-Landel-Ferry[9] (WLF) or Arrhenius equations are often used although others may be appropriate.

This paper departs from the traditional shifting procedure, whether it is by visual inspection or computational curve matching of the isotherms. Instead the particular properties of the wicket plot and the HN equation will be used to uniquely determine the relevant parameters. This process will be illustrated first by a set of computer generated data and then by real data on a standard material.

III. TRIAL DYNAMIC VISCOELASTIC DATA

Polymer data can cover a wide range of properties. Different instruments can be used to measure the properties with unknown error and bias. Thus to properly test the present approach a data set was computer generated with a maximum 10% random error in the storage modulus and a maximum 10% random error in the loss factor. The data set represents data that could have typically been obtained using a Dynamic Mechanical Thermal Analyzer (DMTA), resonance apparatus, or other instrument using specific frequencies over a wide range of temperatures to obtain a complete modulus profile from the rubbery modulus to the glassy modulus. The HN parameters used in generating this data are $\alpha = 0.5$, $\beta = 0.1$, $E_0 = 10^6$ Pa, $E_\infty = 10^9$ Pa, $f_0 = 100$ Hz, and WLF constants C1 = 12.2 and C2 = 119 [see Eq. (13)]. It is common practice to apply a vertical correction $T_0\rho_0/T\rho$ to the moduli (real and imaginary) in order to account for density and temperature changes as suggested by polymer theory for amorphous materials.[9] The vertical correction may lead to smoother data but generally amounts to a small correction and should only be applied from the rubbery region through approximately 80% of the transition region. For the purposes of this paper any required temperature and/or density correction is already factored in and the HN parameters represent an ideal set of data with 10% random error.

IV. THE WICKET PLOT

The wicket plot is an extremely useful tool in analyzing dynamic viscoelastic data,[6,8] especially data that has been

gathered at a variety of frequencies and temperatures. The wicket plot is a log-log plot of the loss factor versus the modulus, usually the real part, and has its origins in circuit analysis, known there as a Nichols plot. In this plot all of the data, independent of temperature and frequency, appears as a single inverted "U" or wicket. That this is so can be seen from Eqs. (8) and (10) where the only dependence on frequency and temperature is through the reduced variable $\omega\tau$. The inverted "U" shape is also common to all the previous models: Zehner, Cole-Cole and Cole-Davidson.[2,6]

The approximate Kramer-Kronig[10] relation is also of interest

$$\tan\delta \cong (\pi/2)[d\log E'/d\log\omega]. \tag{12}$$

It is clear from the Kramer-Kronig relation that one dynamic property can be calculated from the knowledge of the other and that the shape of the rubber to glass transition curve point to the inverted "U" wicket plot.

Thus the wicket plot is a representation of all of the data independent of any temperature shift function (TSF) and the operator process of obtaining the TSF. The significance of this is paramount in our consideration of the HN equation. Finally, if there were no error in the data all of the data would appear as a single curve of vanishing width. Deviations from this single curve can represent errors in the measurement and extreme deviations from the set of data can represent bad data points that can be selectively eliminated from the data set. However low scatter in the data does not necessarily mean accurate data or lack of a systematic error in the data.[6]

Generally vertical shifts are applied to both the real and imaginary parts of the moduli. Recent measurements on a foam rubber[11] did not appear to follow this rule and bear further comment. If the morphology of the foam rubber changes with temperature (e.g., by way of cavity dimensions, geometry, humidity, or degree of crystallization) then the properties can change. Thus foam or highly compliant materials may require special considerations and corrections applied to the data with justification. Analyzing the data,[11] an excellent fit to the HN equation is achieved with a small increase in the static modulus E_0 of the foam with decreasing temperature. The significance of this change is not entirely clear but may be due to the strengthening of the foam as the sample volume changes.

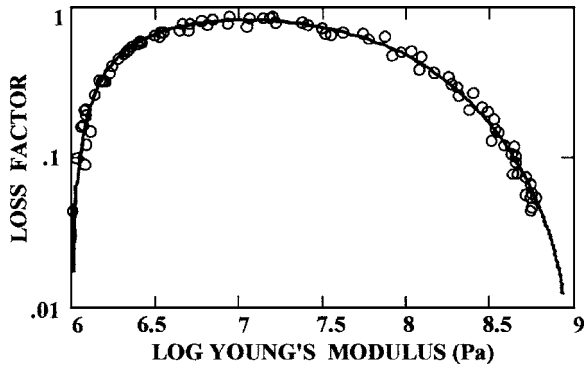


FIG. 1. Wicket plot of computer-generated data for a high loss material with a maximum 10% error in modulus and maximum 10% error in loss factor. o— data points, solid curve—best fit using the HN equation with $\alpha=0.501$, $\beta=0.099$, $E_0=9.92\times10^5$, and $E_\infty=1.008\times10^9$.

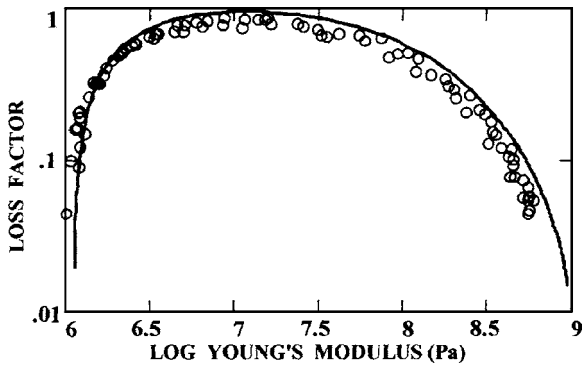


FIG. 2. Wicket plot as in Fig. 1, with a 10% change in the HN fitting parameters to show the sensitivity of fit to the HN parameters. HN parameters: $\alpha=0.55$, $\beta=0.11$, $E_0=1.1\times10^6$, and $E_\infty=1.1\times10^9$.

Finally, materials may not exhibit limiting moduli E_0 and E_∞ that are completely independent of temperature, frequency, and creep which are not contemplated in the HN equation. Such considerations are beyond the scope of the present paper. Nevertheless the HN equation can accurately model the transition region and is a useful tool in analyzing and optimizing acoustic materials.

The HN equation may be displayed as a wicket plot. To do this one chooses a set of parameters, E_∞, E_0, α, β, and an arbitrary characteristic frequency f_0. A wicket plot of this HN equation over a wide frequency range will then appear as a smooth curve with the inverted "U" shape. It is now possible to adjust the position and shape of the HN plot to coincide with the actual data points in the wicket plot of the experimental data if constants are properly assigned. Thus the four of the characteristic material constants E_∞, E_0, α, β used in the HN equation are determined without recourse to the time-temperature shifting (TTS) process. This is the inverse of the conventional method in which the TSF was first determined by TTS, either by sight or computational curve matching methods, and then the data best fit to an analytical equation such as the HN equation.

The minimum error fit to the computer generated test data set is shown in Fig. 1. The fit is excellent. Data with

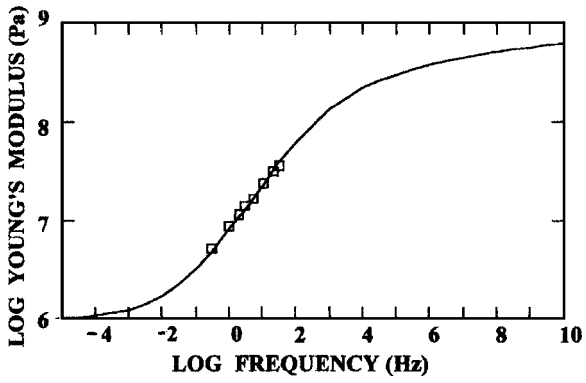


FIG. 3. Plot of the HN equation fit to the computer-generated data at the reference temperature $T_0=0\,^\circ$C with a characteristic frequency $f_0=93.5$ Hz. Solid line HN equation, □—data at $T_0=0\,^\circ$C.

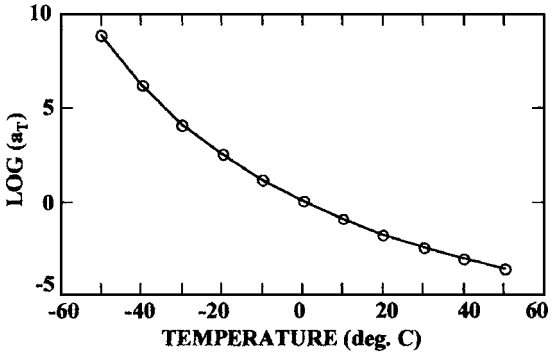


FIG. 4. Plot of the temperature shift function for the computer-generated data. o—data from fitting the HN equation to the actual frequency data at individual temperatures. Solid line—minimum error fit of Eq. (13) to the points with C1=12.13 and C2=119.2 and T_0 is 0 °C.

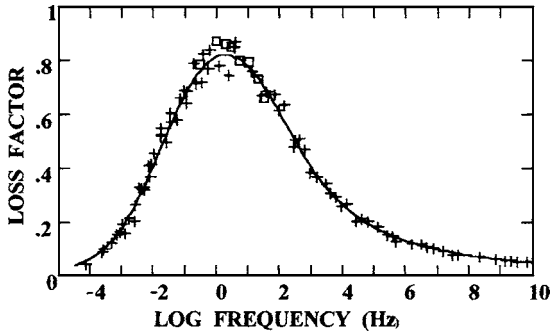


FIG. 6. Loss factor master curve for the computer-generated data at the reference temperature T_0=0 °C for all of the temperature sets shifted onto the HN curve. Solid line—HN equation, +—data points, □—actual unshifted data at T=0 °C.

error larger than the 10% used here would have a larger scatter around the solid HN equation curve and data with smaller error would coincide with the HN curve. Individual points that significantly deviate from the smooth curve could presumably be rejected as bad data.

A fit similar to Fig. 1 fit but now with a 10% deviation in the HN fitting parameters is shown in Fig. 2. This fit is clearly not as good as the one in Fig. 1. This demonstrates the sensitivity of the HN parameters in fitting the data. The deviation from a good fit is especially noticeable in the values of the two limiting moduli that are quite sensitive to the values chosen. The maximum in the loss factor is also considerably off from the optimum value, mainly through the choice of α. Finally, although not explicitly evident, the least sensitive parameter is β, as it requires almost a 20% variation in β to make a noticeable change.

V. THE HAVRILIAK-NEGAMI FREQUENCY PLOT

The next and final step is to determine the actual characteristic frequency f_0 appropriate for some temperature T_0 for which data exists. This is simply done by choosing the reference temperature T_0, plotting the actual data for that temperature on a modulus versus frequency plot, and performing a minimum error fit of the HN equation to the actual experimental data at that temperature. This uniquely determines the characteristic relaxation time or frequency. The results of this process for the computer-generated data are shown in Fig. 3.

Figure 3 can now be repeated for each temperature for which a data set exists. Each temperature will have a reduced frequency f_R relative to T_0 and the set of temperatures and reduced frequencies uniquely determines the TSF.

The data and a smooth curve that fit the data using the WLF equation is shown in Fig. 4. Equation (13) gives the form of the WLF equation where T_0 is the reference temperature.

$$\log(a_T) = -C1(T - T_0)/C2 + (T - T_0) \tag{13}$$

As expected the $\log(a_T)$ is zero at the reference temperature T_0=0 °C. With the TSF determined, the modulus and loss curves for any other reference temperature can easily be determined by simply shifting the frequency axes by an amount a_T. This is the method of reduced frequencies that can produce a two-dimensional temperature-frequency master curve in the form of a nomogram.[8]

The master plot with all of the temperature data at T =0 °C is shown in Fig. 5. All of the frequencies shift nicely onto the single smooth curve given by the HN equation.

Likewise the loss factor as a function of frequency using the same shifting values used in Fig. 5 is shown in Fig. 6. The loss factor also nicely shifts onto the HN equation prediction.

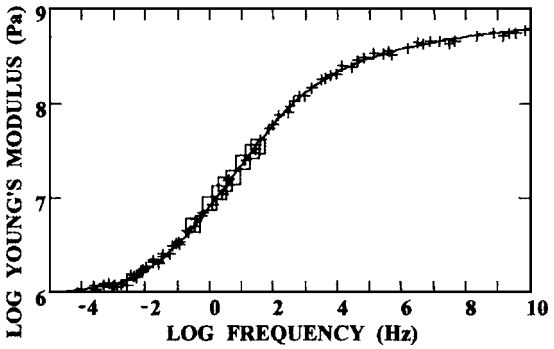


FIG. 5. Modulus master curve for the computer-generated data at the reference temperature T_0=0 °C for all of the temperature sets shifted onto the HN curve. Solid line—HN equation, +—data points, □—actual unshifted data at T=0 °C.

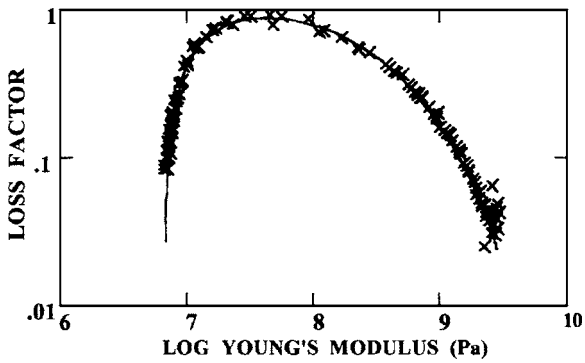


FIG. 7. Wicket plot of the standard ANSI polymer. x—resonance apparatus data points, solid curve—minimum error fit using the HN equation. α =0.568, β=0.059, E_0=6.6×10^6, and E_∞=4.04×10^9.

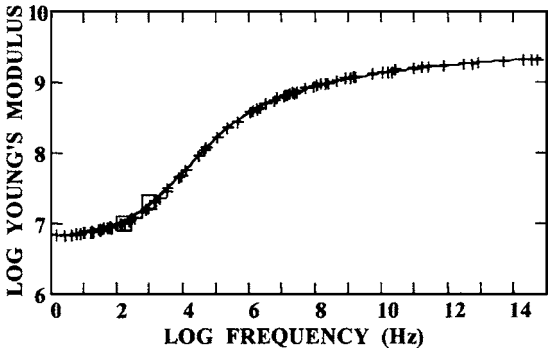


FIG. 8. Modulus master curve for the standard ANSI polymer at the reference temperature $T_0=25\ ^\circ\mathrm{C}$. The characteristic frequency at the reference temperature is $f_0=7.88\times10^4$ Hz. The minimum error WLF constants are C1=8.031 and C2=107.7. Solid line—HN equation, +—resonance apparatus data points, and □—actual unshifted data at $T=24.3\ ^\circ\mathrm{C}$.

VI. STANDARD ANSI VISCOELASTIC POLYMER DATA

Resonance data[3,4] obtained on the standard ANSI material[12] is taken from ANSI S2.24-2001.[8] The methodology is repeated for this real data. Using the wicket plot, the HN equation is first fit to the data to obtain the four HN material constants. Following this a reference temperature is chosen and the HN equation as a function of frequency is fit to the actual experimental data for the reference temperature, thus determining the characteristic frequency f_0 and displaying the entire HN frequency plot. The remaining temperature data is then shifted onto the smooth HN versus frequency curve to determine the TSF. In each step a minimum error technique can be and is used here to determine the relevant constants.

The result of a wicket plot fit is shown in Fig. 7. The scatter in the data is small and the fit is excellent. The fact that the HN equation fits the data indicates that the material is most likely a rheologically simple polymer.

The last step is to choose a reference temperature and fit the HN curve to the frequency data at that temperature and then determine the TSF by shifting the remaining data onto the HN curve. The outcome of this simple process for the standard polymer using the resonance apparatus data set is shown in Fig. 8. The result is a smooth curve with a tight data set.

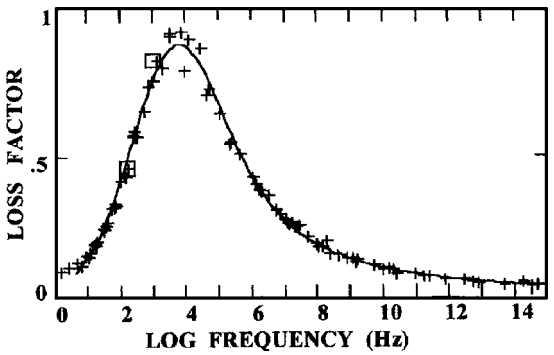


FIG. 9. Loss factor master curve for the standard ANSI polymer at the reference temperature $T_0=25\ ^\circ\mathrm{C}$. Solid line—HN equation, +—resonance apparatus data points, □—actual unshifted data at $T=24.3\ ^\circ\mathrm{C}$.

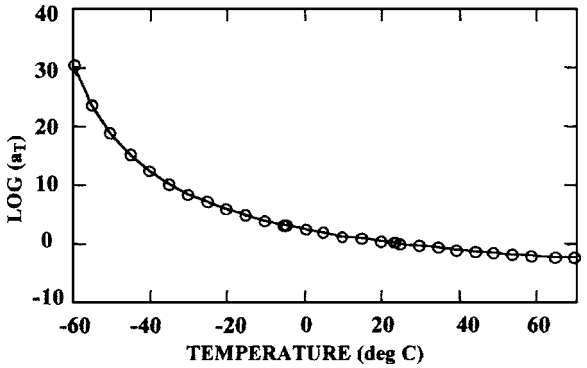


FIG. 10. TSF for the resonance apparatus data on the standard ANSI polymer at 25 °C. The minimum error WLF constants are C1=8.031 and C2=107.7. o—resonance apparatus data, solid line—WLF equation.

A similar plot of the loss factor for the standard ANSI material is shown in Fig. 9. Again the fit is excellent.

The TSF for the standard ANSI polymer is shown in Fig. 10. Note that unlike the previous method the WLF equation is found to work well over the entire temperature range. Thus the current method produces a considerably better fit, with constants that are slightly different than the previous method of graphically shifting the data to obtain the TSF and the master curve. Finally the required vertical temperature/density shift was applied to the data over the complete transition range and the analysis repeated. An equally good fit is obtained (not shown) but with slightly different constants: $\alpha=0.56$, $\beta=0.029$, $E_0=5.95\times10^6$ Pa, $E_\infty=8.24\times10^9$ Pa, $f_0=9.87\times10^4$ Hz, and WLF constants C1=7.9 and C2=102. Note that the values of β and E_∞ are materially changed due to the incorrect application of the temperature shift in the glassy region.

VII. CONCLUSION

The relaxation of many polymers having widely varying properties shows good agreement with the Havriliak and Negami equation. The wicket plot of dynamic material properties is unique as all the data appears as an inverted "U" independent of the temperature and frequency. Traditionally data was first time-temperature shifted by eye or computational curve fitting and then fitted to the Havriliak and Negami equation. In this paper the process is reversed and the wicket plot is used to determine the material constants in the Havriliak and Negami equation independent of time-temperature shifting. This method produces a considerably better fit than the previous method of graphically shifting the data to obtain the TSF. The procedure can also avoid many of the operator error and bias effects inherent in time-temperature shifting as well as shifting uncertainties where the modulus is essentially flat.

[1]S. Havriliak and S. Negami, "A complex plane analysis of α-dispersion in some polymer systems," *in Transitions and Relaxations in Polymers*, J. Polym. Sci. Part C, No. 14, edited by R. F. Boyer (Interscience, New York, 1966), pp. 99–117.
[2]B. Hartmann, G. F. Lee, and J. D. Lee, "Loss factor height and width limits for polymer relaxations," J. Acoust. Soc. Am. **95**, 226–233 (1994).
[3]W. M. Madigosky and G. F. Lee, "Improved resonance technique for material characterization," J. Acoust. Soc. Am. **73**, 1374–1377 (1983).

[4]American National Standard, "Resonance method for measuring the dynamic mechanical properties of viscoelastic materials," ANSI S2.22-1998.
[5]American National Standard, "Single cantilever beam method for measuring the dynamic mechanical properties of viscoelastic materials," ANSI S2.23-1998.
[6]L. Rogers and B. Fowler, "Smoothing, interpolating, and modeling complex modulus data for viscoelastic damping materials, including a new approach to temperature shift functions," US Air Force Technical Report: AFRL-VA-WP-TR-2004-3073 (July 2004).
[7]W. Madigosky and B. Hartmann, "Inter-laboratory comparison of dynamic mechanical measurements," *Ultrasonics International 1995*, Edinburgh, Scotland, 5–7 July (Elsevier Science B.V., Amsterdam, 1995).
[8]American National Standard, "Graphical presentation of the complex modulus of viscoelastic materials," ANSI S2.24-2001.
[9]J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1980), 3rd ed.
[10]T. Pritz, "Verification of local Kramers-Kronig relations for complex modulus by means of fractional derivative model," J. Sound Vib. **228**, 1145–1165 (1999).
[11]F. M. Guillot and D. H. Trivett, "A dynamic Young's modulus measurement system for highly compliant layers," J. Acoust. Soc. Am. **114**, 1334–1345 (2003).
[12]American National Standard, "Method for preparation of a standard material for dynamic mechanical measurements," ANSI S2.21-1998.

Acousto-mechanical and thermal properties of clotted blood[a)]

Volodymyr M. Nahirnyak
Department of Physics, University of Cincinnati, Cincinnati Ohio 45221-0011

Suk Wang Yoon[b)] and Christy K. Holland[c)]
Department of Biomedical Engineering, University of Cincinnati, Medical Science Building, Room 6167, 231 Albert Sabin Way, Cincinnati, Ohio 45267-0586

(Received 28 November 2005; revised 7 April 2006; accepted 10 April 2006)

The efficacy of ultrasound-assisted thrombolysis as an adjunct treatment of ischemic stroke is being widely investigated. To determine the role of ultrasound hyperthermia in the process of blood clot disruption, the acousto-mechanical and thermal properties of clotted blood were measured *in vitro*, namely, density, speed of sound, frequency-dependent attenuation, specific heat, and thermal conductivity. The amplitude coefficient of attenuation of the clots was determined for 120 kHz, 1.0 MHz, and 3.5 MHz ultrasound at room temperature ($20\pm2\ ^\circ$C). The attenuation coefficient ranged from 0.10 to 0.30 Np/cm in porcine clots and from 0.09 to 0.23 Np/cm in human clots. The experimentally determined values of specific heat and thermal conductivity for porcine clotted blood are $(3.2\pm0.5)\times10^3$ J/kg·K and 0.55 ± 0.13 W/m·K, respectively, and for human clotted blood are $(3.5\pm0.8)\times10^3$ J/kg·K and 0.59 ± 0.11 W/m·K, respectively. Measurements of the acousto-mechanical and thermal properties of clotted blood can be helpful in theoretical modeling of ultrasound hyperthermia in ultrasound-assisted thrombolysis and other high-intensity focused ultrasound applications. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2201251]

PACS number(s): 43.35.Wa, 43.80.Ev [CCC] Pages: 3766–3772

I. INTRODUCTION

An estimated 700 000 people experience a stroke in the U.S. every year, resulting in an aggregate lifetime cost of $56.8 billion (AHA, 2005). The only therapy for ischemic stroke that is approved by the FDA is the thrombolytic agent, recombinant tissue plasminogen activator (rt-PA). Recombinant t-PA is moderately effective in lysing thrombi in ischemic stroke patients and it improves neurologic deficits if given within three hours after the onset of stroke symptoms (Wolpert *et al.*, 1993). Unfortunately, thrombolytics also can cause intracerebral hemorrhage. Thus adjuvant therapies that lower the rt-PA dose or increase its efficacy would represent a significant breakthrough (Polak, 2004). Experimental evidence has demonstrated that ultrasound used as an adjuvant to rt-PA can increase thrombus dissolution in an *in vitro* model (Shaw *et al.*, 2001a, b, 2006; Holland *et al.*, 2002). Mild heating of only a few degrees and mechanical effects, such as acoustic streaming, radiation force, cavitation, and accompanying microstreaming all could contribute to the enhanced thrombolysis.

Siddiqi *et al.* (1995) reported increased permeation of fibrinolytic enzymes into a fibrin gel due to pulsed ultrasound exposure at 1 MHz with an intensity of 2 W/cm^2 and 50% duty cycle. The authors attributed the enhanced permeability to ultrasound-induced cavitation. Degassing the fluid in which the fibrin gel sample was immersed significantly reduced the effect of fibrinolysis.

Increased penetration of radiolabeled rt-PA into clots during 1-MHz insonation at 4 W/cm^2 was observed by Francis *et al.* (1995; Francis, 2001). Ultrasound exposure increased the transport of rt-PA into the plasma clot resulting in deeper penetration. Cavitation has also been shown to be an important physical mechanism in the ultrasound-assisted thrombolysis (Everbach and Francis, 2000). Cavitation is thought to loosen the fibrin network, increase the penetration rate of the fibrinolytic agent, and create additional binding sites causing the fragmentation of the thrombus. However, other mechanisms were hypothesized to participate in the enhancement because the application of overpressure, which prevents the nucleation of inertial cavitation, did not completely remove the acceleration of thrombolysis.

Sakharov *et al.* (2000) provided evidence for streaming and heating during continuous wave ultrasound-enhanced lysis of plasma clots at 1 MHz. A temperature increase in the clot of a few degrees produced a twofold increase in lytic rate. This group attributed 30% of the effect of ultrasound-enhanced fibrinolysis to heating and the remaining 70% to microstreaming. The peak-to-peak pressure employed was 0.26 MPa and the corresponding time-averaged intensity was 2.3 W/cm^2. Enhanced thrombolysis was also achieved by heating and mild stirring (without ultrasound exposure). There is general agreement that a temperature increase is only minimally responsible for accelerated thrombolysis (Dick *et al.*, 1998; Francis *et al.*, 1992; Lauer *et al.*, 1992; Blinc *et al.*, 1993; Harpaz *et al.*, 1993; Olsson *et al.*, 1994; Sakharov *et al.*, 2000).

[a)]A portion of this work was presented in "Acousto-mechanical and thermal properties of clotted blood," at the 149th meeting of the Acoustical Society of America, Vancouver, Canada, May 2005.
[b)]Dr. Suk Wang Yoon is a visiting professor from SungKyunKwan University, Suwon 440-746, Republic of Korea.
[c)]Author to whom correspondence should be addressed. Electronic mail: christy.holland@uc.edu

 0001-4966/2006/119(6)/3766/7/$22.50

Suchkova *et al.* (2002) explored ultrasound-enhanced fibrinolysis at low kilohertz frequencies (27–100 kHz) in an attempt to minimize ultrasonic heating and concomitant adverse bioeffects. Significant enhancement of fibrinolysis was achieved, with the greatest effect observed at 27 kHz. The largest effect was observed with continuous-wave ultrasound. A significant acceleration of ultrasound fybrinolytic effect was also observed with peak intensity of 1 W/cm^2 and duty cycles of 10% and 1%.

According to Cintas *et al.* (2004), the use of galactose-based microbubbles increased the ultrasound-assisted fibrinolytic effect during low- intensity, 2-MHz ultrasound exposure of whole human blood clots with rt-PA. After a bolus of a 0.2 mL 400 mg/ml of D-Galactose Palmitic Acid (Levovist®) was added to the closed-loop flow system containing a clot, the authors reported a 31% clot mass loss after 30 -min insonification at an intensity of 0.46 W/cm^2 in the presence of rt-PA. Clots exposed to rt-PA alone exhibit only a 13.2% mass loss. These authors concluded that cavitation and microstreaming were the main mechanisms responsible for clot disruption and fragmentation. These effects may allow a deeper penetration of rt-PA into the clot which in turn exposes a larger percentage of the fibrin network to rt-PA.

An analytical model is being developed to determine the contribution of mild heating as a mechanism for the acceleration of the enzymatic activity of rt-PA during lysis using pulsed ultrasound over the frequency range of 120 kHz to 3.5 MHz. For such a computational model to be developed, the acousto-mechanical and thermal properties of clotted blood must be well characterized. Duck (1990) has collated an extensive literature on the physical properties of tissue, but the specific heat and thermal conductivity of clotted blood have not yet been explored. Therefore, experiments were conducted *in vitro* with porcine and human clots at room temperature to measure the density, specific heat, thermal conductivity, speed of sound, and attenuation over the frequency of 120 kHz to 3.5 MHz.

II. METHODS

A. *In vitro* clot model

The synergistic thrombolytic effect of rt-PA and 120 -kHz ultrasound was assessed previously *in vitro* in a porcine clot model (Holland *et al.*, 2002). To determine the thermal and acousto-mechanical properties of these clots, the same blood clot production protocol was followed, which was approved by the local institutional animal care and use committee. Whole blood clots were prepared by aliquoting 1.5 ml arterial porcine or venous human blood into the 8 -mm inner diameter glass Vacutainer tubes, immersing the tubes in a 37 °C water bath for 3 h and storing the clots at 5 °C prior to use in comparative ultrasound and rt-PA studies, which ensured complete clot retraction. This type of clot is fairly similar to physiologic venous clots and avoids the alteration of the clot structure by employing anticoagulated blood and the addition of thrombin. Additional aliquots of blood from each pig or human were used to obtain a complete coagulation panel from Antech Diagnostics (Chicago, IL), including D-Dimer, A-PTT, fibrinogen and prothrombin time testing, as well as a complete blood count. Most pigs used as part of this study were found to be slightly anemic, with hematocrits in the range of 25% to 35%. Only donors with values in the range 10–900 ng/ml for the D-Dimer test, 10–25 s for A-PTT, 250–700 mg/dl for the fibrinogen concentration, and 9–13 s for prothrombin time were considered to be acceptable. The resulting clots were normally dark red in color, roughly cylindrical in shape with an average diameter of 7–10 mm, and with the typical mass about 0.5 g for each of them, shown in Fig. 1.

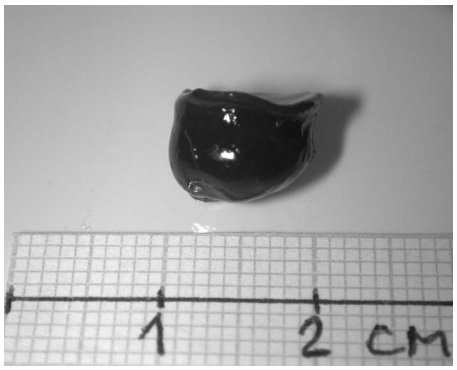


FIG. 1. Photo of a blood clot sample. Whole blood clots were prepared from either fresh porcine or human blood by aliquoting 1.5 or 2.0 ml into 10 -ml glass tubes (BD Vacutainer™, Franklin Lakes, NJ), immersing the tubes in a 37 °C water bath for 3 h and storing the clots at 5 °C for at least 3 days prior to assessment of the properties, which ensured complete clot retraction. The typical size of the clot samples was about 7–10 mm in diameter and the typical mass of each clot was about 0.5 g.

B. Density, calorimetric, and thermal conductivity measurements

For the measurements of clot density, the volume of the clots was determined using a fluid displacement method with an uncertainty of 0.08 cm^3. A standard laboratory scale (E200, Mettler-Toledo, Inc., Columbus, OH), was used to assess the mass of each clot in physiologic saline at room temperature (22±2 °C) with an uncertainty of 0.01 g. The overall uncertainty in the density measurements was 0.02×10^3 kg/m^3. A total of 202 clots from 28 human subjects and a total of 336 porcine clots from 38 pigs were employed for the density and calorimetric measurements. Thus the influence of physiologic variability and experimental uncertainty were captured in the standard deviation of the density and calorimetric measurements.

Direct calorimetric measurements using calibrated E-type thermocouples (Omega Engineering, Inc., Stanford, CT) were performed to determine the specific heat of human and porcine clots relative to the specific heat of a standard fluid, physiological saline (0.9% sodium chloride solution). The experimental setup included two 236 ml Styrofoam® containers with insulating lids, which contained E-type thermocouples. Two digital thermometers (Model HH 506-R, Omega Engineering, Inc., Stanford, CT) with thermocouples attached to them were used to record temperatures within the containers throughout the experiment. The method of mixtures (Semat, 1953) was used to determine the heat capacity of the sample clots in saline held at an inital temperature of

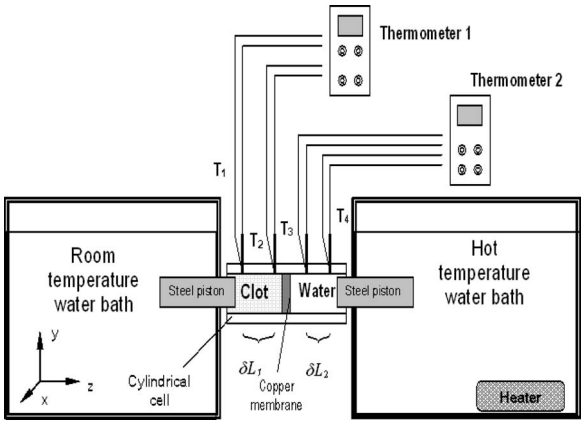


FIG. 2. Experimental apparatus for thermal conductivity measurements in blood clots.

6±1 °C and combined with a second aliquot of saline at room temperature (21±1 °C). The uncertainty in the temperature measurements was 0.06 °C and the uncertainty in the measurements of the clot mass was 0.01 g. The propagated error for the specific heat measurements in this case was 0.3×10^3 J/kg·K.

The comparative steady-state method (Kreith and Black, 1980) was used for determination of the thermal conductivity of whole blood clots in the experimental apparatus shown in Fig. 2. The sample chamber is inserted between two $25\times25\times25$ cm^3 Plexiglas® water tanks held at two different temperatures, T_1 and T_2. A thermal gradient was thus created horizontally across the cylindrical experimental cell made of polyethylene-terephthalate glycol (PETG) containing both the clot sample and a fluid of known thermal conductivity, namely deionized, degassed water. PETG has a significantly higher coefficient of thermal conductivity (0.33 W/m·K) than air (0.024 W/m·K), thus allowing us to simplify the experimental setup to one-dimensional heat flow along the axis of the cylindrical cell. Six Luerlock® connectors were attached to the experimental cell as ports for positioning thermocouples and expunging air (four along the chamber containing clot and two along the chamber containing water). The diameter of the cylindrical chambers containing the standard fluid and clot sample was 11.2 mm. The length of the chamber containing the clot sample was 55 mm and the length of the chamber containing water was 35 mm. A 1.27 -mm-thick copper membrane with high thermal conductivity separated the clot chamber from the chamber filled with water. The membrane was necessary to prevent contamination of the water from the sample chamber. Hypodermic, 0.29 -mm-diam T-type thermocouples (HYP1-30, Omega Engineering, Inc., Stanford, CT) were used to monitor the temperature at four different points along the chambers. Two stainless steel pistons of diameter 11.2 mm and length 80 mm provided thermal contact between the two water tanks and the sample chambers. This experimental design allowed the position of the thermocouple tips to be adjusted within the clot and the temperature drop at different locations along the clot to be measured. The magnitude of the thermal gradient between the tanks was approximately 5 °C which, given losses in the stainless steel pistons, corresponded to an approximate 1 to 1.5 °C thermal gradient across the clot sample.

To validate our method of measuring thermal conductivity, the sample chamber was filled with a fluid of known thermal conductivity, ethanol, as a test. A time period of an hour was required to allow the temperature gradients to stabilize in the sample chambers. The coefficient of thermal conductivity of the investigated substance (either clot or ethanol) was calculated from the following formula:

$$\lambda_{\text{unknown}} = \lambda_w \frac{\delta T_2}{L_2}\frac{L_1}{\delta T_1}, \tag{1}$$

where λ_{unknown} is the thermal conductivity of the investigated substance, λ_w is the thermal conductivity of water, $\delta T_1 = T_2 - T_1$ is the magnitude of the thermal gradient across the sample under investigation, L_1 is the distance between the two thermocouples in the sample, $\delta T_2 = T_4 - T_3$ is the thermal gradient across the water sample, and L_2 is the distance between the two thermocouples in the water-filled chamber. The uncertainties in the measurements of temperature and distance were 0.06 °C and 0.2 mm, respectively. The propagated error in determination of the coefficient of thermal conductivity was 0.02 W/m·K. The thermal conductivity measured using ethanol as the investigational substance was 0.20±0.02 W/m·K, which compares favorably with the published value for the thermal conductivity of ethanol, 0.19 W/m·K (Kreith and Black, 1980).

The coefficient of thermal conductivity was determined in three 10-ml specimens of clots prepared according to the protocol described above. We recorded the thermal gradients in the clot samples at three different positions within the clots. The thermal conductivity of porcine and human clots was determined by employing Eq. (1), averaging the results of all nine experiments, and computing the standard deviation of the measurements.

C. Sound speed in clotted blood

We used an insertion pulse-echo method to measure the speed of sound in clotted blood. A polystyrene cuvette of 4.5 ml capacity and dimensions $12.5\times12.5\times46$ mm^3 (Curtin Matheson Scientific, Inc. Baltimore, MD) was immersed in a water tank and positioned within the focus of a 3.5 -MHz transducer. Ultrasonic pulses were generated with a function generator (Model 33250, Agilent Technologies, Inc., Palo Alto, CA), amplified (Model AR 150LA, Ampifier Research, Souderton, PA), and monitored with an oscilloscope (LT372, LeCroy Corp., Chestnut Ridge, NY). The water tank was connected to the water thermostat (EX111, Neslab Instruments, Inc., Newington, NH). Experiments were conducted at two temperatures: 20 °C and 37 °C. The choice of the working frequency was determined by two factors. First, the axial resolution of the ultrasound measurement is dictated by the acoustic wavelength, the shorter the better. On the other hand, the choice of too high a frequency is limited by the attenuation in clot and cuvette walls, which were 6 mm apart.

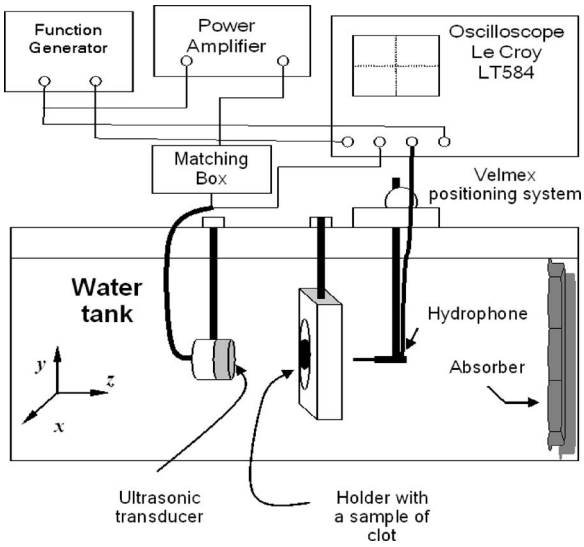


FIG. 3. Experimental apparatus for the measurements of the amplitude attenuation coefficient in blood clots.

The time delay between echoes reflected from the front face and far wall of the cuvette was used to determine the speed of sound at two temperatures, 20 °C and 37 °C. The pulse repetition period for all experiments was 10 ms which corresponds to a PRF of 100 Hz. The three-cycle pulse duration was 0.9 μs. First, we measured the time delays between echoes from the front and back walls of the cuvette when it was filled with water. By using the sound speed in water at 20 °C and 37 °C, 1482 and 1524 m/s, respectively, we determined the distance between two walls of the cuvette. After the water was drained, a blood clot was carefully put into the same cuvette and the time between echoes from the cuvette walls was measured. The speed of sound in the clot was calculated by dividing twice the distance between the cuvette walls by the measured time delay. Three human clot samples and three porcine clot samples were used and the mean and standard deviation were calculated. For validation of the technique, the cuvette was filled with ethanol instead of clotted blood and the measurements of the sound speed were carried out. The measured speed of sound in ethanol, 1200 m/s at 20 °C and 1150 m/s at 37 °C, agree well with the tabulated values of 1191 m/s at 20 °C and 1159 m/s at 37 °C (CRC, 1979).

D. Frequency-dependent attenuation

We employed an insertion loss method to measure the ultrasound attenuation in clotted blood (Hill *et al.*, 2004). The measurements of the pressure amplitude attenuation coefficient were conducted in a Lucite™ water tank with dimensions of $68 \times 55 \times 55$ cm^3 filled with degassed and deionized water at 22±2 °C for three center frequencies: 120 kHz, 1 MHz, and 3.5 MHz. The sample holder, shown in Fig. 3, consisted of a Lucite™ frame which supported two acoustic windows (Tegaderm™, 3M Health Care, St. Paul, MN) sandwiching the clot sample in saline solution. Each clot was placed in the focal plane of one of the calibrated ultrasonic transducers. Two holders were employed in our experiments. The first had an opening diameter of 70 mm and was used for the 120-kHz experiments. The second had an opening diameter of 40 mm and was used for experiments at 1 and 3.5 MHz. In both cases the diameter of the openings was much bigger than the acoustic wavelength and the transducer beamwidth at the corresponding frequencies. No distortion of acoustic field due to diffraction of ultrasound was evident in the transverse beam profiles obtained when the clot sample was absent.

TABLE I. Characteristics of transducers and hydrophones used in the measurements of ultrasound attenuation in blood clots.

Transducer frequency	Hydrophone	Transducer aperture (cm)	−3 dB beamwidth (cm)	Focal distance (cm)
120 kHz	TC4038, RESON, Inc., Goleta, CA	6.14	2.2	7.4
1 MHz	Precision Acoustics 0.2 mm needle probe	2.5	0.7	9.4
3.5 MHz	Precision Acoustics 0.2 mm needle probe	1.9	0.25	9.3

All transducers were preliminarily calibrated in a water tank at room temperature (20±1 °C). Their experimentally measured focal distance, focal depth, and beam width are presented in Table I. The focal distance and beamwidth for the 120-kHz transducer were 74 and 22 mm, respectively. The 1-MHz transducer focal distance was 94 mm and the beamwidth was 7 mm. The focal length for the 3.5-MHz transducer was 93 mm and the beamwidth was 2.5 mm. The transducers were driven by signals from a function generator (Agilent 33250) as shown in Fig. 3. For the experiments at 120 kHz, an Ultra Series 2021LF/HF linear amplifier (T&C Power Conversion Inc., Rochester, NY) with an accompanying 50-Ω impedance matching unit (Sonic Concepts, Inc., Woodinville, WA) amplified the driving signal. For the experiments at 1 and 3.5 MHz, an AR 150LA (Ampifier Research, Souderton, PA) with no additional impedance matching was utilized. At each frequency the pulse repetition period was 1 ms corresponding to a 1-kHz PRF and each tone burst contained 10 cycles. A TC4038 hydrophone (RESON, Inc., Goleta, CA) and a 0.2-mm needle hydrophone (HP 0.2-mm interchangeable probe, Precision Acoustics Ltd., Dorset, UK) were used to measure the peak-to-peak pressure amplitude, initially in the free field (without the clot) and subsequently with the clot in the transducer's focal plane. The position of the hydrophones in the water tank was computer controlled by a Velmex system with a stepping motor controller (NF 90 series, Velmex, Inc., Bloomfield, NY). A sheet of acoustic absorbing material (Aptflex F28, Precision Acoustics Ltd., Dorset, UK) was fixed on the far wall of the Plexiglas tank to prevent the reflection of acoustic waves.

By comparing the acoustic pressure amplitude at the same location in the free field with that from the intervening clot, the magnitude of the coefficient of attenuation was mea-

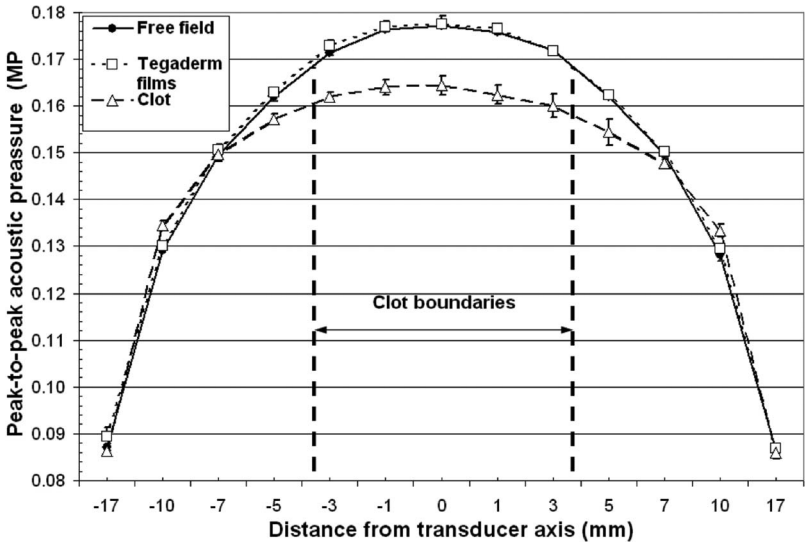


FIG. 4. Beam profiles in the focal plane of the 120 -kHz transducer. The decrease of the pressure amplitude (lower curve) was caused by the introduction of a clot sample into the acoustic field. The relative reduction of the acoustic pressure at the center of the beam was used to determine the amplitude coefficient of attenuation. Error bars represent the standard deviation of the measurements in three clots. The Tegaderm™ films used in the experiments do not cause a significant change of the pressure amplitude due to attenuation in the material of the film.

sured. The pressure amplitude coefficient of attenuation, μ, was determined using the following formula (Leighton, 1997):

$$\mu = \frac{1}{l} \ln\left(\frac{P_{\text{free}}(x,y)}{P_{\text{clot}}(x,y)}\right), \tag{2}$$

where l is the length of clot along the ultrasound beam axis, $P_{\text{free}}(x,y)$ is the pressure amplitude in the focal plane of ultrasound transducer, $P_{\text{clot}}(x,y)$ is the pressure amplitude at the same location after the clot was introduced into the acoustic field, and x and y are the transverse coordinates of the position of the hydrophone in the focal plane of transducer (Fig. 3). In our experiments, the thickness of clots ranged from 0.69 to 0.89 mm. The thickness was assessed with a calibrated caliper before placing the clot into the holder. The magnitude of the pressure amplitude was measured over the transverse beam profile using a temporal peak, spatial peak pressure output of 0.18 MPa for all three transducers. The transverse beam profile of the 120-kHz transducer with the presence and absence of the clot is shown in Fig. 4. The pressure amplitude attenuation coefficient for each frequency was determined by averaging the pressure reduction data within three blood clot samples by scanning the transverse beam profile along two perpendicular directions. The results of these measurements were averaged and the mean and the standard deviation were computed to assess the variability.

III. RESULTS

The clot densities were $(1.08\pm0.02)\times10^3$ kg/m^3 for human clots and $(1.06\pm0.01)\times10^3$ kg/m^3 for porcine clots. The speed of sound in porcine clots was 1547 ± 1 m/s at 20 °C and 1577 ± 2 m/s at 37 °C. The speed of sound in human clots was 1597 ± 9 m/s at 20 °C and 1633 ± 4 m/s at 37 °C.

The results for the measurements of amplitude attenuation coefficients in human and porcine clots with the corresponding standard deviations are presented in Table II. Their magnitudes vary between 0.10 and 0.3 Np/cm in porcine clots and between 0.09 and 0.23 Np/cm in human clots over the frequency range between 120 kHz and 3.5 MHz. A least squares fit to the acoustic attenuation data in both porcine and human clots shown in Fig. 5 demonstrates that the frequency dependence of the amplitude attenuation coefficient obeys approximately a one-third power law.

The average magnitudes of the coefficients of specific heat of the human and porcine clots were $(3.5\pm0.8)\times10^3$ J/kg·K and $(3.2\pm0.5)\times10^3$ J/kg·K, respectively. The coefficient of thermal conductivity for porcine clots was 0.55 ± 0.13 W/m·K and for human clots was 0.59 ± 0.11 W/m·K.

IV. DISCUSSION

An analysis of the results obtained should start with a look at the structure of clotted blood presented in Fig. 6. A formed blood clot has the same main constituents as whole blood: red cells (or erythrocytes), platelets, and fibrin fibers. Density, specific heat, and thermal conductivity for whole human blood can be found in the literature. The density reported to be 1.05×10^3 to 1.6×10^3 kg/m^3, the specific heat is 3.61×10^3 to 3.89×10^3 J/kg·K, and the thermal conductivity 0.49 to 0.55 W/m·K (NCRP, 1992; Duck, 1990; Lehmann and DeLateur, 1982). Human red blood cells have a density of 1.093×10^3 kg/m^3 (Duck, 1990). Their specific heat is 3.21×10^3 J/kg·K and thermal conductivity is 0.45–0.482 W/m·K (Toubal *et al.*, 1999; Duck, 1990). Isolated fibrin fibers have density of 1.3×10^3 kg/m^3 (Mourad and Kargl, 2003). The normal value for the human blood hematocrit lies in the range of 39% to 52% (Alanen *et al.*,

TABLE II. Pressure amplitude attenuation coefficients in whole porcine and human blood clots.

Frequency	Attenuation in clotted porcine blood (Np/cm)	Attenuation in clotted human blood (Np/cm)
120 kHz	0.10±0.01	0.09±0.02
1 MHz	0.17±0.02	0.17±0.02
3.5 MHz	0.3±0.1	0.23±0.04

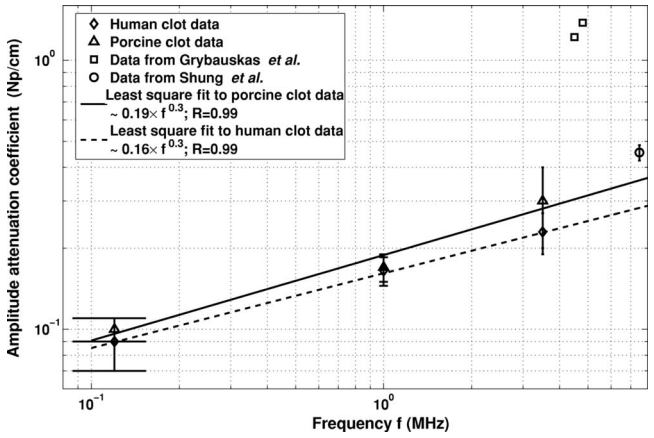


FIG. 5. Frequency dependence of the pressure amplitude attenuation coefficient in blood clots. The solid and dashed lines refers to a least square fit of the data to the form $\mu = a \cdot f^b$, where μ is the pressure amplitude attenuation coefficient, f is the center frequency, and the R values listed in the legend are the correlation coefficients and represent a goodness of fit.

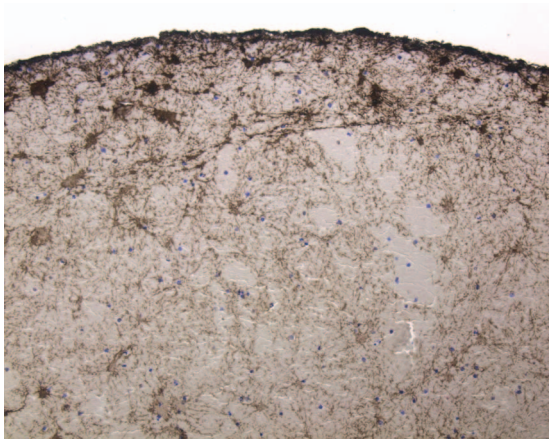


FIG. 6. Photo of the cross section of a porcine clot, with 20× magnification stained with an antifibrinogen antibody (Dako Corp., Carpinteria, CA) which stains fibrin fibers brown (Datta, 2005).

1985). For weanling (age 6 to 10 weeks) pig blood, the value of hematocrit ranges between 29% and 33% (Osborne and Meredith 1971) and is slightly anemic.

It is helpful to compare the measured acoustomechanical and thermal properties of clotted blood with other types of soft tissues. Typical values for the density of the human organs such as kidney, liver, and brain range between 1.04×10^3 and 1.06×10^3 kg/m^3 (Lehmann and De Lateur 1982). Thus, the density of blood clots, $(1.08 \pm 0.02) \times 10^3$ kg/m^3 for human clots and $(1.06 \pm 0.01) \times 10^3$ kg/cm^3 for porcine clots, is consistent with the upper range of values for soft tissue densities. The coefficient of specific heat for brain is 3.68×10^3 J/kg·K, for liver is 3.60×10^3 J/kg·K, and for kidney is 3.89×10^3 J/kg·K (Lehmann and De Lateur, 1982). The thermal conductivity of brain is 0.528 W/m·K, that of liver is 0.565 W/m·K, and that of kidney is 0.544 W/m·K (Lehmann and De Lateur 1982). Note that our data for the thermal properties of clotted blood are also within the range of values for soft tissues in the literature.

Our results of the measurements of sound speed in clotted blood compare favorably with previously published results (Grybauskas *et al.*, 1978; Shung *et al.*, 1984). Grybauskas *et al.* (1978) obtained an experimental value for sound speed in human blood clots of 1602 m/s that were 90 min old at 37 °C. Shung *et al.* (1984) measured a sound speed of 1590 m/s for human blood clots that were 1 h old at 23 °C. These investigators also noted an increase of sound speed with the age of the clot. Our measured value for the speed sound at 20 °C, 1597±9 m/s at 20 °C, agrees with these data.

Previously published data for the attenuation of ultrasound at 1 MHz in whole human blood are 0.017 Np/cm (Hoskins *et al.*, 2003). Grybauskas *et al.* (1978) reports 1.22–1.38 Np/cm for the magnitude of absorption of ultrasound in the frequency range between 4.5 and 4.8 MHz. Shung *et al.* (1984) obtained 0.45 Np/cm for ultrasound attenuation in human clotted blood at 7.5 MHz. In our experiments, the attenuation of ultrasound in porcine and human clotted blood at 1 MHz was 0.17 Np/cm. If we compare the obtained results with the published ones, we see that the values obtained by Grybauskas *et al.* (1978) are significantly higher than the typical attenuation of ultrasound in soft tissues for this frequency range. These data are closer to the ultrasound attenuation in bone. On the other hand, our experimental results are in good agreement with the data presented by Shung *et al.* (1984).

The measured attenuation coefficients agree well with the published values for other types of soft biological tissues such as liver, kidney, and brain, which contain the same types of proteins. For example, the value of ultrasonic attenuation in liver is 0.17 Np/cm, in muscles is 0.2–0.25 Np/cm, and in brain is 0.12 Np/cm (Damask, 1978). All of them have the same order of magnitude of the amplitude attenuation coefficient. The attenuation of the ultrasound in clotted blood is influenced by ultrasonic absorption by macromolecules of the various proteins which constitute clot (NCRP, 1992).

A two-parameter least squares fit was performed on the attenuation data of the form $\mu = af^b$, where μ is the pressure amplitude attenuation coefficient and f is the center frequency. The power dependence, b, was 0.32 for porcine clots and 0.28 for human clots. In contrast to the linear dependence of attenuation on frequency, which is typical in many soft tissues, our results demonstrated an approximate one-third power dependence between 120 kHz and 3.5 MHz. The frequency dependence of attenuation in clotted blood is similar to the frequency dependence of ultrasound attenuation in lung (~0.2 power), skeletal muscles (~0.8 power), and skin (~0.6 power) (Hill *et al.*, 2004). Theoretical calculations of the frequency dependence of the acoustic absorption in whole clotted blood were reported by Mourad and Kargl (2003) and exhibited a 0.9 power dependence.

The standard deviation in the thermal conductivity and specific heat data reflects the natural variability of the physical properties of blood clots. The presence of air pockets, cavities, and the inhomogeneous fibrin structure in clots might also contribute to the standard deviation of the thermal conductivity and specific heat experimental results. The simple techniques used in this work may easily be employed

to investigate other types of biological soft tissues.

V. CONCLUSIONS

The acousto-mechanical and thermal properties of the human and porcine clotted blood were determined experimentally and add to the knowledge base of the physical properties of clotted blood. The results obtained will allow us to make numerical estimates of a magnitude of the thermal elevation during ultrasound insonification of blood clots for enhanced thrombolysis and may be helpful in the development of high-intensity focused ultrasound and other therapeutic ultrasound applications.

Understanding the potential for thermal mechanisms involved in the interaction of ultrasound with blood thrombi during ultrasound-assisted thrombolysis is an important first step in improving thrombolytic efficacy while minimizing unwanted thermal bioeffects.

ACKNOWLEDGMENTS

The authors want to thank Sampada Vaidya, Jun Tan, Ashima Dhamija, and Nazli Bavani for the help in preparation of the blood clot samples. This research was supported by the National Institutes of Health, Grant No. NIH/NINDS R01-NS047603.

Alanen, A., and Kormano, M., (**1985**). "Correlation of the ecogenisity and structure of clotted blood," J. Ultrasound Med. **4**, 421–425.

American Heart Association (AHA) (2005). *Heart Disease and Stroke Statistics—2005 Update.*, (American Heart Association, Dallas, TX).

Blinc, A., Francis, C. W., Trudnovski, J. L., and Carstensen, E. L. (**1993**). "Characterization of ultrasound-potentiated fibrinolysis in vitro," Blood **81**, 2636–2643.

Cintas, P., Nguen, F., Boneu, B., and Larrue, V. (**2004**). "Enchancement of enzymatic fibrinolysis with 2-MHz ultrasound and microbubbles," J. Thrombosis Haemostasis **2**, 1163–1166.

CRC (**1979**). *CRC Handbook of Chemistry and Physics* edited by Robert C. West, (CRC, Baca Ratan, FL).

Damask, A. C. (**1978**). *Medical Physics. Volume I: Physiological Physics, External Probes* (Academic, New York), p. 215.

Datta, S., McAdory, L. E., Tan, and J., Holland, C. K. (**2005**). "Cavitation detection during ultrasound-assisted thrombolysis in porcine blood clots," J. Acoust. Soc. Am. **117**, 2558.

Dick, A., Neuerburg, J., Schmitz-Rode, T., Alliger, H., Schmid-Schonbein, H., and Gunter, R. W. (**1998**). "Thrombolysis of mural thrombus by ultrasound. An experimental in vitro study," Invest. Radiol. **33**, 85–90.

Duck, F. A. (**1990**). *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, London).

Everbach, E. C., and Francis, C. W. (**2000**). "Cavitational mechanisms in ultrasound-accelerated thrombolysis at 1 MHz," Ultrasound Med. Biol. **26**(7), 1153–1160.

Francis, C. W. (**2001**). "Ultrasound-enhanced thrombolysis," Echocardiogr. **18**(3), 239–246.

Francis, C. W., Blinc, A., Lee, S., and Cox, C. (**1995**). "Ultrasound accelerates transport of recombinant tissue plasminogen activator into clots," Ultrasound Med. Biol. **21**(3), 419–424.

Francis, C. W., Onundarson, P. T., Carstensen, E. L., Blinc, A., Meltzer, R. S., Schwarz, K., and Marder, V. (**1992**). "Enhancement of fibrinolysis in vitro by ultrasound," J. Clin. Invest. **90**, 2063–2068.

Grybauskas, P., Kundrotas, K., Sukackas, V., and Jaronis, E. (**1978**). "Ultrasonic digital interferometer for investigation of blood coagulation," Ultrasonics **16**(1), 33–36.

Harpaz, D., Chen, X., Francis, C. W., Marder, V. J., and Meltzer, R. S. (**1993**). "Ultrasound enhancement of thrombolysis and reperfusion in vitro," J. Am. Coll. Cardiol. **21**, 1507–1511.

Hill, C. R., Bamber, J. C., and ter Haar, G. R. (**2004**). *Physical Principles of Medical Ultrasonics* (Wiley, Chichester, England), p. 123.

Holland, C. K., Vaidya, S. S., Coussios, C.-C., and Shaw, G. J. (**2002**). "Thrombolytic effects of 120 MHz and 1 MHz ultrasound and tissue plasminogen activator on porcine whole blood clots," J. Acoust. Soc. Am. **112**, 2370.

Hoskins, P. R., Trush, A., Martin, K., and Whittingham, T. A. (**2003**). *Diagnostic Ultrasound. Physics and Equipment* (GMM, London).

Kreith, F., and Black, W. Z. (**1980**). *Basic Heat Transfer* (Harper and Row, New York).

Lauer, C. G., Burge, R., Tang, D. B., Bass, B. G., Gomez, E. R., and Alving, B. M. (**1992**). "Effect of ultrasound on tissue-type plasminogen activator-induced thrombolysis," Circulation **86**, 1257–1264.

Lehmann, J. F., and DeLateur, B. J. (**1982**). *Therapeutic Heat and Cold* (Williams & Wilkins, Baltimore).

Leighton, T. G. (**1997**). *The Acoustic Bubble* (Academic, London), p. 28.

Mourad, P. D., and Kargl, S. G. (2003). "Acoustic properties of fluid-saturated blood clots," Technical Report of the Applied Physics Laboratory, University of Washington. APL-UW TR 2003, p. 16.

National Council on Radiation Protection and Measurements (NCRP) (1992). "Exposure criteria for medical diagnostic ultrasound: I. Cretiria based on thermal mechanisms," NCRP Publications, Report no. 113, June 1, p. 35.

Olsson, S. B., Johansson, B., Nilsson, A. M., and Olsson, C., Roijer, A. (**1994**). "Enhancement of thrombolysis by ultrasound," Ultrasound Med. Biol. **20**, 375–382.

Osborne, J. C., and Meredith, J. H. (**1971**). "Hematological values of the normal weanling piglet," Cornell Vet. **61**(1), 13–23.

Polak, J. F. (**2004**). "Ultrasound energy and the dissolution of thrombus," N. Engl. J. Med. **351**(21), 2154–2155.

Sakharov, D. V., Hekkenberg, R. T., and Rijken, D. C. (**2000**). "Acceleration of fibrinolysis by high-frequency ultrasound," Thromb. Res. **100**(4), 333–340.

Semat, H. (**1953**) *Fundamentals of physics.* (*Rinehart*, New York), p. 275.

Shaw, G. J., Bavani, N., Dhamija, A., and Lindsell, C. J. (**2006**). "Effect of mild hypothermia on the thrombolytic efficacy of 120 kHz enchanced thrombolysis in an in-vitro human clot model," Thromb. Res. **117**, 603–608.

Shaw, G. J., Hahn, N. L., Wagner, K. R., Kanter, D. S., and Holland, C. K. (**2001a**). "Ultrasound assisted clot lysis for stroke therapy," J. Acoust. Soc. Am. **109**, 2456.

Shaw, G. J., Hahn, N. L., Wagner, K. R., Kanter, D. S., and Holland, C. K. (**2001b**). "Ultrasound assisted thrombolysis in an *in vitro* clot model," Acad. Emerg. Med. **8**, 542.

Shung, K. K., Fei, D. Y., Yuan, Y. W., and Reeves, W. C. (**1984**). "Ultrasonic characterization of blood during coagulation," J. Clin. Ultrasound **12**(3), 147–153.

Siddiqi, F., Blinc, A., Braaten, J., and Francis, C. W. (**1995**). "Ultrasound increases flow through fibrin gels," Thromb. Haemostasis **73**(3), 495–498.

Suchkova, V., Carstensen, E. L., and Francis, C. W. (**2002**). "Ultrasound enhancement of fibrinolysis at frequencies of 27 to 100 kHz," Ultrasound Med. Biol. **28**(3), 377–382.

Toubal, M., Asmani, M., Radziszewski, E., and Nongaillard, B. (**1999**). "Acoustic measurement of compressibility and thermal expansion coefficient of erythrocytes," Phys. Med. Biol. **44**(5), 1277–1287.

Wolpert, S. M., Bruckmann, H., Greenlee, R., Wechsler, L., Pessin, M. S., and del Zoppo, G. J. (**1993**). "Neuroradiologic evaluation of patients with acute stroke treated with recombinant tissue plasminogen activator," AJNR Am. J. Neuroradiol. **14**(1), 3–13.

Ultrasonic evaluation of residual stresses in flat glass tempering: Comparing experimental investigation and numerical modeling

Marc Duquennoy,[a] Dany Devos, and Mohammadi Ouaftouh
Institut d'Electronique de Microélectronique et de Nanotechnologie, Département d'Opto-Acousto-Electronique, UMR CNRS 8520, Université de Valenciennes, Le Mont Houy, 59313 Valenciennes, France

Dominique Lochegnies and Eric Roméro
Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines, UMR CNRS 8530, Université de Valenciennes, Le Mont Houy, 59313 Valenciennes, France

(Received 26 October 2005; revised 16 March 2006; accepted 28 March 2006)

In order to control residual stress distribution in glass, techniques based on the phenomenon of photoelasticity are efficient, though subject to the inherent limitations of all optical techniques. To mitigate these limitations, we exploit the phenomenon of acoustoelasticity to estimate residual stress distribution, using surface acoustic waves. Experimental results are obtained for a 6 mm thick soda-lime silicate flat glass plate that had been subjected to nonuniform thermal tempering and whose stress distribution is calculated using finite element modeling. The estimated stress distributions provided by our ultrasonic method compare quite well with the results from the modeling, from both the qualitative and quantitative points of view. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197806]

PACS number(s): 43.35.Zc, 43.35.Pt, 43.35.Cg [YHB] Pages: 3773–3781

I. INTRODUCTION

Knowledge of the levels and the homogeneity of a residual stress field is crucial to guaranteeing the quality of tempered glass. Today, most of the nondestructive techniques used to estimate glass stresses are photoelasticity techniques,[1] based on optical properties. When first proposed, such techniques were impossible to implement because the specialized equipment was difficult to use in a production environment. However, the recent development of integrated, portable equipment has made this type of control possible during production. In addition, progress has been made to allow the stresses near the edges and across the thickness of the tempered glass to be controlled. For example, in the domain of automotive glass, Redner[2] developed a portable contact system connected to a fully automated personal computer, which permits the accurate measurement of surface and edge compression and average near-edge tensile stresses. Anton and Aben[3] used a compact light polariscope and a high-performance data recording system in a scattered-light contact method that applies the photoelasticity theory to measure stress distribution across the thickness of the tempered glass of a cathode ray tube bulb. Other polariscopes have also been developed for analyzing hollow tempered glass, such rim-tempered drinking glasses.[4] In order to measure the stresses within tempered glass, Cannon *et al.*[5] developed a noncontact method based on Bragg scattering. Using a pair of thermal gratings, their device was able to directly access the stresses located midplane in a tempered glass plate. The authors compared their results for the Bragg scattering model to experimental investigations involving scattered beams and stress measurements. Cavaillé *et al.*,[6] on the other hand, used Brillouin light scattering to measure sound velocity, and therefore elastic moduli, in several samples of tempered glass. This scattering technique is also a noncontact technique, but its instrumentation requires specific expertise; meaning that it is often assembled in the laboratory. Nonetheless, given highly spatial resolution measurements, the technique can be used to analyze the stress profiles in tempered glass.

All of the studies mentioned above used photoelasticity techniques, which are subject to certain inherent limitations due to the optical properties on which they are based. In order to mitigate some of these limitations (e.g., spatial resolution, difficulty of inspecting curved surfaces or dark glass, need to control the surface of the glass that has not come in contact with the tin during the float process), we propose an ultrasonic technique based on acoustoelasticity theory. Stress fields can be analyzed using ultrasonic waves because the stresses within a material change the characteristics of the wave (velocity, attenuation).[7] Using surface ultrasonic waves makes it possible to narrow the field to estimate surface residual stresses.[8,9]

According to acoustoelasticity theory and experimental results, the variation of the ultrasonic wave velocity, when it encounters stresses, is small. Consequently, in order to correctly evaluate residual stresses, the accuracy of the relative variation of velocity measurement must be better than 10^{-3}. Ultrasound methods have primarily been used to test residual stress distributions on metals. However, we wanted to try this kind of method on glass, whose stress level is typically lower than the stress level observed in metals.[7] Our experiments

[a]Author to whom correspondence should be addressed. Electronic mail: marc.duquennoy@univ-valenciennes.fr

were conducted on a 6 mm thick flat glass plate that had been subjected to nonuniform thermal tempering prior to our experiments in order to allow us to demonstrate the potential of our ultrasonic method for estimating stress distribution in tempered flat glass and its sensitivity toward low residual stress levels. The specificity of this study is that it analyzes the distribution of surface residual stress, and not applied stress, in glass.

II. THEORETICAL EXPRESSIONS OF RAYLEIGH WAVES IN A STRESSED MEDIUM

Establishing expressions of the ultrasonic wave velocities in a stressed material requires taking the type of wave and the mechanical behavior of the material into account. To distinguish the phases of the deformation process, three deformation states were defined, although the different material loading and unloading stages are not completely known. The undeformed reference configuration (i.e., with no stress and no strain) is called the *natural state*. The finite deformation, either applied or residual, is brought about by homogeneous prestress in the elastic area, transforming the natural state of the reference configuration into the static equilibrium of the *initial state*. The position of this material point is represented by the $\overline{\mathbf{x}}$ position vector. Finally, the configuration of the initial state changes when a dynamic perturbation is superposed on it, shifting the material into a third state, called *the final state*, whose position is defined by the $\widetilde{\mathbf{x}}$ position vector. A common Cartesian coordinate system (ξ_1, ξ_2, ξ_3) is used to refer to the position of the material points in the three states. The physical quantities of the natural state are denoted by "*O*" superscript, those that refer to the initial state by "*I*" superscript, and those that refer to the final state by "*F*" superscript.[7] The initial deformation corresponding to the passage from the natural state to the initial state and the final deformation corresponding to the passage from the initial state to the final state are denoted by $\overline{\mathbf{u}}$ and $\widetilde{\mathbf{u}}$, respectively. The perturbation resulting from the acoustic wave is assumed to be small in relation to the initial deformation $(\|\widetilde{\mathbf{u}}\| \ll \|\overline{\mathbf{u}}\|)$.[9]

The motion equation of an acoustic wave propagating in a stressed material has already been published several times.[7–9] To establish such equations, it is necessary to assume the following series of hypotheses: The solid is hyperelastic; all deformations are in the elastic area; the predeformation is static; and the predeformation and rotations are small. With these initial coordinates, the motion equation is:

$$\frac{\partial}{\partial \overline{\mathbf{x}}_J}\left[(R_{IJKL} + \delta_{IK} t^{\mathrm{I}}_{JL})\frac{\partial \widetilde{\mathbf{u}}_K}{\partial \overline{\mathbf{x}}_L}\right] = \rho^{\mathrm{I}}\frac{\partial^2 \widetilde{\mathbf{u}}_{\mathrm{I}}}{\partial t^2},$$

$$I = 1,2,3 \text{ and } J = 1,2,3, \tag{1}$$

with

$$R_{IJKL} = C_{IJKL}\left(1 - \frac{\partial \overline{\mathbf{u}}_M}{\partial \overline{\mathbf{x}}_M}\right) + C_{IJKLMN}\frac{\partial \overline{\mathbf{u}}_M}{\partial \overline{\mathbf{x}}_N} + C_{MJKL}\frac{\partial \overline{\mathbf{u}}_I}{\partial \overline{\mathbf{x}}_M} + C_{IMKL}\frac{\partial \overline{\mathbf{u}}_J}{\partial \overline{\mathbf{x}}_M} + C_{IJML}\frac{\partial \overline{\mathbf{u}}_K}{\partial \overline{\mathbf{x}}_M} + C_{IJKM}\frac{\partial \overline{\mathbf{u}}_L}{\partial \overline{\mathbf{x}}_M}, \tag{2}$$

where C_{IJKL}, C_{IJKLMN} are the second- and third-order elastic constants, ρ is the density, e^I is the infinitesimal strain tensor, and t^I is the Cauchy stresses tensor, defined over an area of material elements in the deformed state. The stress tensors provide the stress for the initial state.

Let us consider a surface wave [i.e., a Rayleigh wave (RW)] propagating on the free surface of a homogeneous, elastic, and semi-infinite material subjected to uniform static deformation. The solid half-space occupies the region $\xi_2 \geqslant 0$, and the RW is assumed to propagate in the ξ_1-direction. Equations (1) and (2) are established by taking into account the stress development to the second order in terms of strain, and by assuming small static deformations and small static rotations. The (ξ_1, ξ_2) sagittal plane is perpendicular to the ξ_3-binary axis, thus the $\widetilde{\mathbf{u}}_3$ component, corresponding to the shear harmonic mode, is completely independent of the RW. The analysis becomes two dimensional as in the isotropic case. The displacement associated with the RW has the following form:

$$\widetilde{\mathbf{u}}_{\mathrm{I}} = f_I(\overline{\mathbf{x}}_2)\exp[ik(\overline{\mathbf{x}}_1 - V_R \cdot t)], \quad I = 1,2, \tag{3}$$

where $\widetilde{\mathbf{u}}_3$ is zero, f_I are the amplitudes according to the depth $\overline{\mathbf{x}}_2$, k is the wave number, V_R is the RW velocity, and t is the time.

At the boundary surface ($\overline{\mathbf{x}}_2 = 0$), the component t^I_{22} vanishes and is assumed to stay at zero at the depth of a Rayleigh wavelength. By substituting the displacement $\widetilde{\mathbf{u}}_{\mathrm{I}}$ into the motion equation and satisfying the boundary conditions imposed by the free surface, we obtain Eq. (4) with which the phase velocity of the RW in a uniformly strained material is determined:[8,10,11]

$$[R_{22}\{\alpha - (t^I_{11} + R_{11})\} + (R_{12})^2]^2\{\alpha - (t^I_{11} + R_{66})\} - R_{66}R_{22}(\alpha - t^I_{11})^2\{\alpha - (t^I_{11} + R_{11})\} = 0, \tag{4}$$

where $\alpha = \rho^I V_R^2$, ρ^I is the density of the prestress material, and the Voigt's notation is used to contract the indices.

By replacing the R_{IJKL} terms in this equation with their expressions in terms of the second- and third-order elastic constants [Eq. (2)], and by neglecting the higher-order terms for the strains, the following equation is obtained:

$$(a_0 + a_1 e^I_{11} + a_2 e^I_{22})\alpha^3 + (b_0 + b_1 e^I_{11} + b_2 e^I_{22})\alpha^2 + (c_0 + c_1 e^I_{11} + c_2 e^I_{22})\alpha + (d_0 + d_1 e^I_{11} + d_2 e^I_{22}) = 0. \tag{5}$$

The a, b, c, and d constants in Eq. (5) depend on the symmetry of the material (i.e., on the second- and third-order elastic constants). These expressions are provided in Appendix A.

Given the stress $t^I_{22} = 0$ when $\overline{\mathbf{x}}_2 = 0$, only the principal strains, e^I_{11} and e^I_{22}, are independent of each other. Thus, the strain e^I_{33} is a function of e^I_{11} and e^I_{22}, and the solution to Eq. (5) can be written as: $\alpha = \alpha_0 + \alpha_1 e^I_{11} + \alpha_2 e^I_{22}$, where α_0, α_1, and α_2 are independent of e^I_{11} and e^I_{22}.

TABLE I. Values of the a, b, c, and d constants.

Constant	Values
$a_0(\mathrm{Pa}^2)$	33×10^{20}
$a_1(\mathrm{Pa}^2)$	-29.1×10^{20}
$a_2(\mathrm{Pa}^2)$	433.9×10^{20}
$b_0(\mathrm{Pa}^3)$	-709.3×10^{30}
$b_1(\mathrm{Pa}^3)$	-3002.7×10^{30}
$b_2(\mathrm{Pa}^3)$	-7707.9×10^{30}
$c_0(\mathrm{Pa}^4)$	43×10^{42}
$c_1(\mathrm{Pa}^4)$	401.8×10^{42}
$c_2(\mathrm{Pa}^4)$	405.6×10^{42}
$d_0(\mathrm{Pa}^5)$	-64.3×10^{52}
$d_1(\mathrm{Pa}^5)$	-897.5×10^{52}
$d_2(\mathrm{Pa}^5)$	-535.2×10^{52}

Given the approximation: $(\Delta V/V^O)^2 \ll (\Delta V/V^O)$,[9] the relative variation of the RW velocity in isotropic media (glass), expressed in terms of the static strains, can be written as

$$\frac{\Delta V_R}{V_R^o} = \frac{V_R - V_R^o}{V_R^o} = \frac{1}{2}\frac{\alpha_1}{\alpha_0}e_{11}^I + \frac{1}{2}\left(\frac{\alpha_2}{\alpha_0} + 1 - \frac{C_{11}}{C_{12}}\right)e_{22}^I, \qquad (6)$$

where V_R and V_R^o are, respectively, the RW velocities for the stressed and unstressed medium.

The relative variation of the velocity can then be easily expressed in terms of static stresses, since the principal directions of the strains coincide with the symmetrical axes of the material. Moreover, the stress tensor can be connected to the strain tensor via the second-order elastic constant tensor, because the development of R_{IJKL} is taken up to the first order in terms of static deformations:

$$\frac{\Delta V_{R12}}{V_{R12}^o} = A_{R12}^{(1)} \cdot t_{11}^I + A_{R12}^{(3)} \cdot t_{33}^I, \qquad (7)$$

where $A_{RIJ}^{(K)}$ are the acoustoelastic coefficients, the superscript K denotes the direction of the applied loading, t_{KK}^I is the applied stress in direction K, V_{RIJ} and V_{RIJ}^O are, respectively, the RW velocities of waves propagating along the ξ_I-axis and polarized in the $(\xi_\mathrm{I}, \xi_\mathrm{J})$ plane, for the stressed and unstressed medium. The RW acoustoelastic coefficients are given in Appendix A.

A relation similar to Eq. (7), with the acoustoelastic coefficients $A_{R21}^{(2)}$ and $A_{R21}^{(3)}$, is obtained for the RW propagated along the ξ_3-axis and polarized in the (ξ_2, ξ_3) plane. Moreover, for the case of isotropic media, $A_{R12}^{(1)} = A_{R32}^{(3)}$ and $A_{R12}^{(3)} = A_{R32}^{(1)}$.

Based on this theoretical development, the RW acoustoelastic coefficients can be estimated. The second- and third-order constants—given by Landa et Plešek[12] for a glass quite similar to the one studied in this paper—were used, and the glass density was 2280 kg m^{-3}, $C_{11}=72.4$ GPa, $C_{12}=18.8$ GPa, $C_{112}=58$ GPa, $C_{155}=14.7$ GPa, and $C_{456}=-6.7$ GPa. From the relation between the third-order elastic constants, the following values can be deduced: $C_{111}=116.9$ GPa, $C_{123}=16.6$ GPa, and $C_{144}=28.2$ GPa.

The estimated acoustoelastic coefficients, in turn, allow us to estimate the order of magnitude of the acoustoelastic constants, and we obtain 38.7×10^{-6}/MPa and -0.7×10^{-6}/MPa for $A_{R12}^{(1)}$ and $A_{R12}^{(3)}$. The values of these numerical constants given in Appendix A are summarized in Table I. Thus, this theoretical development allows us to completely predict the dependence of the Rayleigh wave with regard to a stress field. For example, a variation of 100 MPa in residual stress of a piece of glass leads to a variation of 0.38% of the relative variation of the RW velocity.

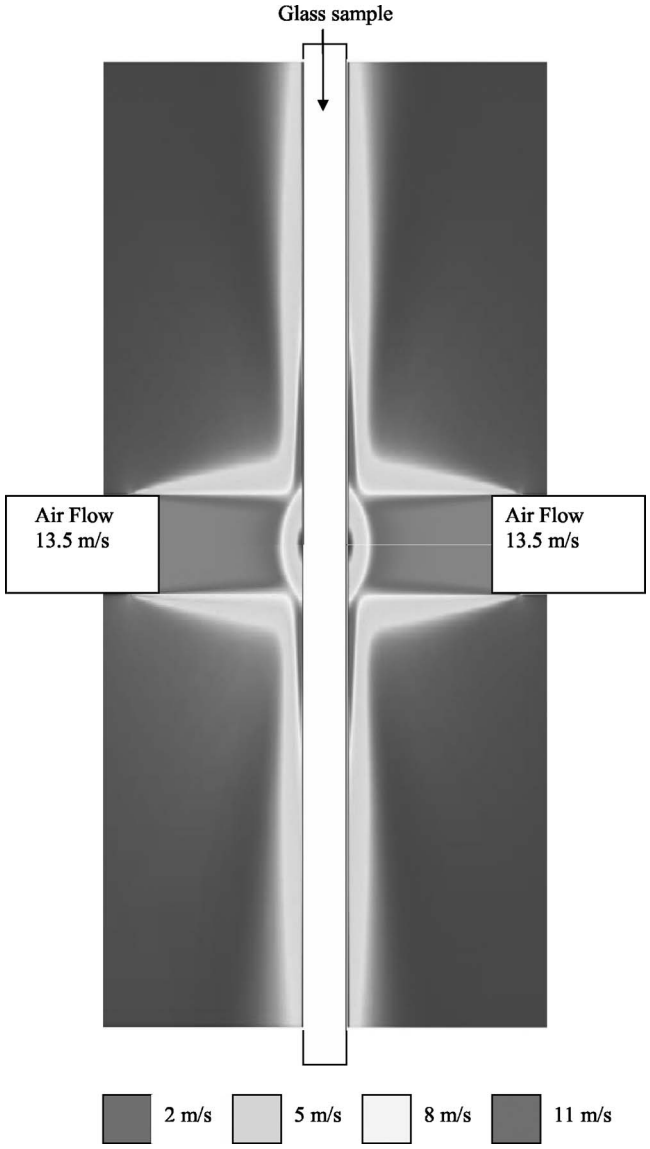


FIG. 1. Laboratory tempering installation.

III. THERMAL TEMPERING OF FLAT GLASS AND NUMERICAL MODELING

The experimentation began by tempering a flat soda-lime silicate glass sample (200 mm × 200 mm × 6 mm) in a vertical position on a laboratory tempering installation (Fig. 1). After its initial heating in a furnace, the sample was moved out the furnace by translation to the zone between two nozzles. The glass was cooled rapidly by the two air jets, which impinged perpendicularly on the center of the sample.

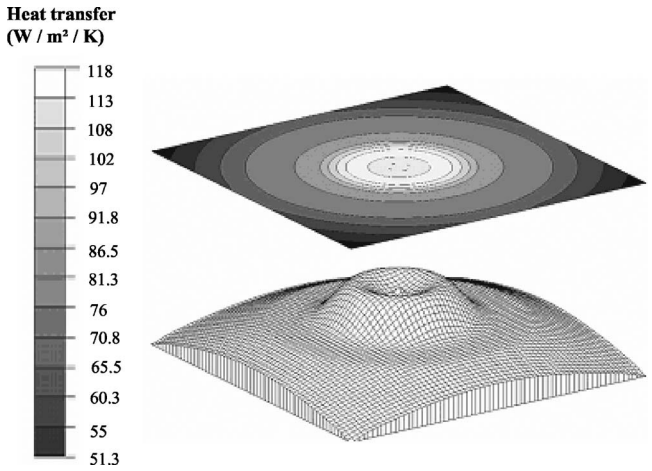


FIG. 2. Heat transfer on the glass sample for a nozzle with an outer velocity of 13.5 m/s.

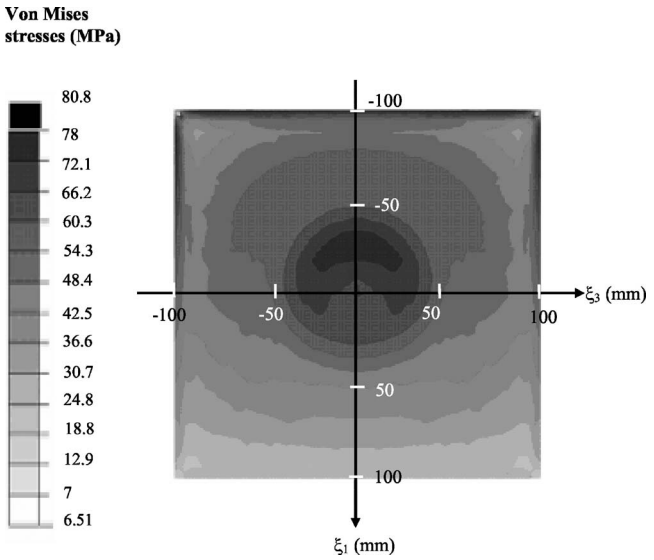


FIG. 3. Distribution of the Von Mises calculated stresses on the surface of the tempered glass sample.

For identical air jets (diameter of the nozzle: 42 mm, distance from the surface: 100 mm, air flow: 13.5 m/s at the nozzle exit) on each side of the flat sample, the first model was designed to estimate the air flow on the glass surface. A computational fluid mechanics (CFD) model was developed to estimate the air velocities from the nozzle exit to the glass surface for one impinging jet.[13,14] The wall heat flux computed for the glass sample, q, is expressed by the heat transfer coefficient h, defined as:

$$h = \frac{q}{(T_a - T_O)}, \tag{8}$$

where T_a and T_o are, respectively, the temperature of the air and the temperature of the glass.

The three-dimensional (3D) representation of the CFD-computed heat transfer on the surface of the glass sample is given in Fig. 2. The distribution is axisymmetrical, as is the CFD model itself. The variation of the heat transfer coefficient on the glass surface is nonlinear from 118 W/m^2/K in the center of the sample to 51 W/m^2/K on the edges.

Because the temperature distribution was graduated from the top to the bottom of the glass sample when it came out of the furnace, at which point its two surfaces were subjected to symmetrical cooling, two symmetries were present in our experiments. The first one passed through one-half of the thickness of the glass sample; and the second one, through the middle of the glass sample. Consequently, the finite element model is limited to one-quarter of the real sample (i.e., only one-half of the sample thickness is considered during the modeling). The 3D mesh is composed of 15,246 eight-node bricks, which are 3 mm in height and width. The half-thickness grid was discretized into seven sections, whose sizes ranged from 0.34 mm on the midplane in the ξ_2-direction to 0.085 mm on the surface. This distribution of element size allowed the typical parabolic stress distribution[15] to be calculated over the glass thickness with no significant error. In fact, the error for the compressive value of the stress on the glass surface was lower than 1%, as compared to a mesh composed of 40 elements along the thickness.

To estimate the tempering residual stresses, it was necessary to decouple the temperature and stress computations. For our thermal computations, the symmetry conditions were expressed as no heat flux on the symmetry planes. On the air-side surface of the glass, convection was modeled using the CFD estimation of the heat transfer (Fig. 2). On the outer edges, a mean heat transfer coefficient equal to 30 W/m^2/K was applied. Infrared measurements of the glass sample at the furnace exit highlighted the existence of a thermal gradient from the top of the glass sample to the bottom. This is mainly due to the positioning of the electric heaters inside the furnace and the lack of a ventilation system to homogenize the air temperatures inside the furnace. Consequently, a linear variation from 940 K at the top of the glass to 873 K at the bottom was considered for the initial temperatures of the mesh nodes prior to cooling. The ambient air temperature was 300 K. (All glass properties used for the temperatures computations are given in Appendix B).

Stress relaxation in glass was modeled with the viscoelastic generalized Maxwell law, using shear and bulk moduli functions.[16] Narayanaswamy's model[15] was used for structural relaxation with the so-called fictive temperature, computed with the algorithm proposed by Markowsky and Soules.[17] (The main equations used for the stress computations are summarized in Appendix C.) To take both the symmetries present in our experiment into account, displacement constraints were imposed on the nodes of the finite element mesh located in the symmetry planes. No displacement in the vertical direction was prescribed on the nodes at the bottom of the glass sample in order to represent the real boundary conditions related to attaching the glass sample to its insulating support. (All mechanical glass properties are given in Appendix B).

Figure 3 shows the distribution of the Von Mises residual surface stress:

$$t^{VM} = \sqrt{\frac{3}{2}[(t_{11}^I - t_{22}^I)^2 + (t_{22}^I - t_{33}^I)^2 + (t_{33}^I - t_{11}^I)^2 + 2t_{12}^{I\,2} + 2t_{13}^{I\,2} + 2t_{23}^{I\,2}]}, \quad (9)$$

in the tempered glass sample after mirroring the numerical results obtained on 3D mesh. The distribution of the computed Von Mises stresses on the glass surface is nonuniform, with a maximum 80 MPa zone in the center of the sample. Figure 3 shows that the stress distributions are different in the top, center, and bottom of the glass sample. This is due to the nonuniform heat transfer (Fig. 2) and the nonuniform initial temperatures of the glass sample prior to tempering. Moreover, except for certain edge effects, the computed stress results show that the difference between t_{11}^I and t_{33}^I is low, less than 10%. In Sec. IV, the evolution of the stress t_{11}^I in the ξ_1-direction is analyzed in detail for the three separate areas (top, center, and bottom) of the glass sample, and the ultrasonic measurements and numerical results are compared.

IV. EXPERIMENTAL RESULTS COMPARED TO THE NUMERICAL MODELING

The block diagram of the experimental device for determining stress profiles is illustrated in Fig. 4. Though is also possible to generate and detect ultrasonic waves using a laser,[18,19] which is useful when contact with the piece being examined is impossible for different reasons (e.g., temperature or geometry), we chose to use wedge probes to generate and receive the RWs in this experiment. Such sensors are appropriate and convenient, and they permit an interesting range of amplitudinal levels.

In order to increase the spatial resolution, small probes were selected (width of 6 mm). The emitter and receiver transducers were mounted in the same support in order to ensure the fixed spacing and proper alignment of the sensors. The temperature variation was limited during the measurements (<0.2 K), and any thermal expansion of the support was negligible.[11,20] The frequency used for this experiment was 5 MHz, producing a RW wavelength of only 0.6 mm, with a relatively weak attenuation. For the electrical excitation of the emitter, a 400 V pulse was used. Both emitter and receiver probes were moved in the ξ_3-direction, and the RW was propagated in the ξ_1-direction. Both transducers were moved via motorized mechanical translation, with a minimum step of 1 μm. A digital oscilloscope was employed to store the signal from the receiver probe. The signal was sampled at a rate of 2 G samples/s, which implies a good time resolution (500 ps). The time of flight of the RW was measured by a zero-crossing technique[21,22] and recorded for each position. All devices were connected by a general purpose interface bus to a computer in order to control the measurement system.

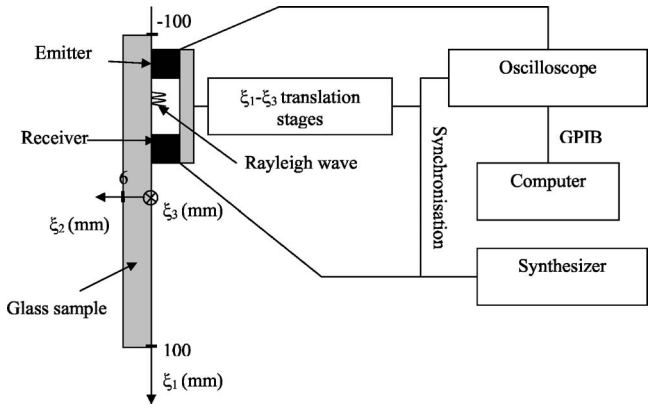


FIG. 4. Ultrasonic experimental device for determining stress profiles.

From the results of the modeling (Fig. 3), which estimated the nonhomogeneity of the stress field, we chose to analyze the distribution of the t_{11}^I stress along three specific lines in the ξ_3-direction. The first line at $\bar{\mathbf{x}}_1 = -15$ mm is located in the center of the glass sample. Along this line, the model indicates the most important stress variation in the glass sample. The second line at $\bar{\mathbf{x}}_1 = -60$ mm corresponds to the top of the sample. According to the modeling (Fig. 3), a lower stress variation, around 20 MPa, exists along this line. The third line, at $\bar{\mathbf{x}}_1 = +60$ mm at the bottom of the sample, shows the last modification of the stress; here, the numerical stress variation is low, around 10 MPa. This last line will be used to highlight the sensitivity of ultrasonic method.

In order to estimate the evolution of the stress along a line, wedge probes were symmetrically positioned on both sides of the line, at a regular distance of 20 mm. Thus, the relative variation of stress that is measured corresponds to the average variation of the stress over the distance that separates the probes. In order to compare our experimental measurements with the numerical model, two stresses in both the ξ_1- and ξ_3-directions were averaged over 20 mm in the ξ_1-direction and 6 mm in the ξ_3-direction. At each position along the line, the time of flight was measured, making it possible to determine the relative variation of the time of flight in relation to the relative variation of the velocity. We were then able to plot the evolution of the residual stresses using the acoustoelastic coefficients and the relative variation of velocity.

In order to compare the acoustoelastic coefficient obtained theoretically, we also measured these acoustoelastic coefficients experimentally from different tempered glass samples. The experimental results of this calibration phase confirmed that the linear evolution of the relative variation of the velocity is a function of the stress, and that the slope corresponds directly to the acoustoelastic coefficient. The value obtained for our glass sample was 29.9×10^{-6}/MPa for the sum of both coefficients. This value is quite close to the one proposed by Jassby *et al.*[23] (25×10^{-6}/MPa) and the ones calculated in theoretical part of Sec. II (38×10^{-6}/MPa). The influence of the stress perpendicular to

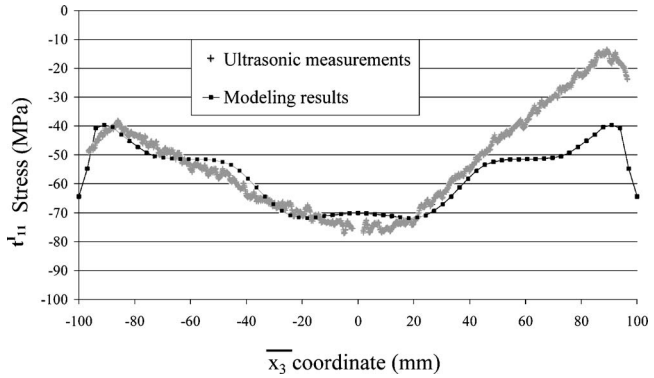


FIG. 5. Evolutions of the stress t^I_{11} along the line of coordinate $\overline{\mathbf{x}}_1 = -15$ mm for the ultrasonic measurement and the modeling.

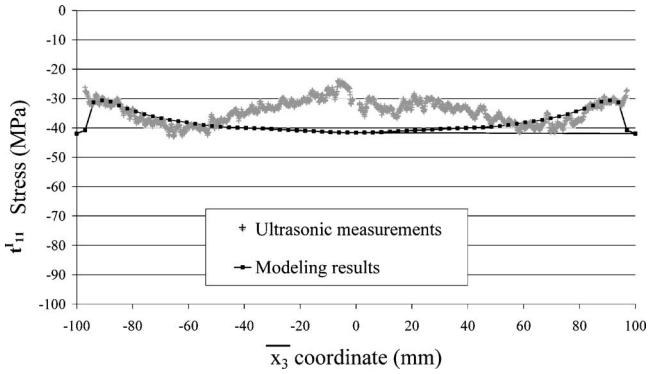


FIG. 7. Evolutions of the stress t^I_{11} along the line of coordinate $\overline{\mathbf{x}}_1 = +60$ mm for the ultrasonic measurement and the modeling.

the direction of propagation seemed to be quite weak (see Sec. II and Jassby *et al.*).[23] Since $A^{(3)}_{R12}$ was weak and since t^I_{11} and t^I_{33} were of the same order, it was possible to combine the influence of the acoustoelestic coefficients in Eq. (7). For this reason, one measurement of the relative variation of the velocity was sufficient for determining the stress profile. We compared the evolution of the three numerical graphs with the three experimental graphs taking the stress value of $\overline{\mathbf{x}}_3 = -90$ mm as reference. Both the experimental and numerical stress evolutions along the three lines are reported in Figs. 5–7.

In Fig. 5, measurements along the line of coordinate $\overline{\mathbf{x}}_1 = -15$ mm (according to the numerical results (Fig. 3), it corresponds to the most stressed zone where the heat exchange was maximum) indicate an estimated stress evolution of around 30 MPa between the center and the edge of the plate. However, there is a dissymmetry in the experimental stress evolution in the ξ_3-direction, which does not appear in the numerical results because only one-half of the glass sample was modeled, and the 3D modeling approach assumed perfect vertical symmetry in the tempering installation. Of course, it is possible that, during the experiment, the perpendicular direction of the airflow toward the glass sample was not perfectly respected, which would also explain the discrepancy between the numerical and experimental results. The dissymmetry of the measured stresses is particularly noticeable in Fig. 5 and since the line $\overline{\mathbf{x}}_1 = -15$ mm corresponds to the center of the glass sample where the heat transfer was the maximum (Fig. 2), a small dissymmetry will have a greater effect on the region in the center of the glass. Moreover, the results must be considered more carefully for the area near the edges ($|\overline{\mathbf{x}}_3| > 96$ mm), because edge effects must be taken into account. Both approaches have their limits: The modeling approach has trouble taking the tempering conditions entirely into account, and the divergence of the ultrasonic beams used in the experimental approach can cause some reflection off the flat edges, which can disturb the signal. This remark is to be taken into account for the three graphs.

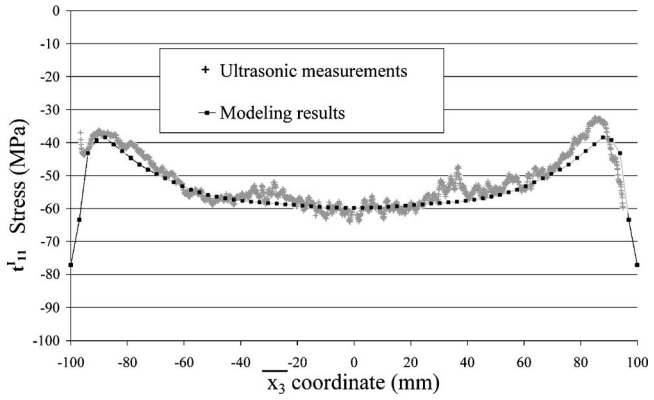


FIG. 6. Evolutions of the stress t^I_{11} along the line of coordinate $\overline{\mathbf{x}}_1 = -60$ mm for the ultrasonic measurement and the modeling.

In Fig. 6, measurements along the line of coordinate $\overline{\mathbf{x}}_1 = -60$ mm indicate an estimated stress evolution of around 20 MPa between the center and the edge of the plate. Though this stress level is low, the ultrasonic measurements are still compatible with the numerical results. In Fig. 7, measurements along the line of coordinate $\overline{\mathbf{x}}_1 = +60$ mm indicate an estimated stress evolution of around 10 MPa between the center and the edge of the plate, except for a local disturbance between $\overline{\mathbf{x}}_3 = -5$ mm to $\overline{\mathbf{x}}_3 = -1$ mm. This stress level is extremely low. The detection of such low-level stresses and the local disturbance demonstrates the great sensitivity of our ultrasonic method. Though it is true that some fluctuations of the measured stress evolution appear along this line, these fluctuations are not random, and they are reproducible for each measurement. For very low stress levels, any local modification of a characteristic can cause disturbances of the velocities of the same order of magnitude as the disturbances caused by the low-level stresses. Thus, they might be the consequence of a surface state.

The results of this comparison are quite encouraging. There is a good correlation between these experimental ultrasonic measurements and those indicated by the numerical modeling. For the three areas inspected, where the inhomogeneity stress fields are different, a similar evolution along the width of the sample was obtained. Moreover, whatever the stress variations between the center of the sample and the edge on these three areas (i.e., 30 MPa, 20 MPa, and 10 MPa) for both the experimental and modeling approaches, the stress variations are relatively similar. Thus, the important point in this analysis is that the results of ultrasonic measurement and numerical modeling of the evolution of the surface residual stresses are coherent, both quali-

tatively and quantitatively. On the other hand, the estimation of such low-level stresses and local disturbances demonstrate the great sensitivity of our ultrasonic method.

V. CONCLUSION

This paper compares stress distributions measured experimentally with an ultrasonic method with the residual stress distributions using CFD and finite element modeling. Both methods were tested on the same glass sample: A 6 mm-thick flat soda-lime silicate glass plate that was subjected to nonuniform thermal tempering, with stress distributions being estimated along three areas where the stresses are nonuniform. This comparison highlighted the qualitative and quantitative coherence of the two methods' results. These preliminary results demonstrate the potential of ultrasonic methods for characterizing stress distribution in glass. Other analyses of more complex samples must be done to complete this research.

We have shown that ultrasonic waves are a potentially advantageous method for estimating residual stresses in glass, offering an alternative to the optical techniques with their inherent limitations. Though this method requires highly accurate ultrasonic measurements, it is entirely possible that this method could be integrated into a production environment in the not-so-distant future.

ACKNOWLEDGMENTS

The present research was supported by the CNRS, the European Community, the French Regional Council of *Nord/Pas-de-Calais* (Interreg III, Stimat), the Belgian Ceramic Research Centre (Mons-Belgium), the Laboratory of Materials and Processes (Valenciennes-France), and the Laboratory of Mechanics and Energetics (Valenciennes-France). The authors gratefully acknowledge the support of these institutions.

APPENDIX A: EXPRESSIONS OF THE *a*, *b*, *c*, AND *d*, CONSTANTS FOR ISOTROPIC SYMMETRY AND ACOUSTOELESTIC COEFFICIENTS OF RAYLEIGH WAVE

$N_1 = C_{11}^2 - C_{12}^2$; $N_2 = C_{12}(C_{12} - C_{11})$; $N_3 = C_{11}N_1 + 2C_{12}N_2$ and $N_4 = (C_{11}/C_{12})$.

$$a_0 = C_{11}[C_{11} - C_{44}].$$

$$a_1 = C_{11}(-2C_{44} + C_{144} - C_{155}).$$

$$a_2 = -C_{44}(4C_{11} + C_{111}) + C_{11}(6C_{11} + 2C_{111} - C_{155}) - N_3[C_{11}(2C_{112} - C_{11} - C_{12} - C_{144}) - C_{44}C_{112}].$$

$$b_0 = 2N_1C_{11}.$$

$$b_1 = C_{11}[-8N_1 - 2C_{11}(C_{44} + C_{111} - C_{112}) + 4C_{12}(C_{112} - C_{123}) + C_{44}(6C_{44} + C_{111} - C_{112})].$$

$$b_2 = \left(\frac{2}{(2C_{44} - C_{11})}\right)\left[\begin{array}{l} C_{11}^3(8C_{456} + 31C_{44}) + C_{11}^2C_{44}(16C_{155} - 67C_{44} - 16C_{456}) \\ +2C_{11}C_{44}^2(17C_{44} - 8C_{112} - 26C_{155}) + 8C_{44}^3(C_{112} + 4C_{155}) \end{array}\right].$$

$$c_0 = N_1(N_1 + 2C_{11}C_{44}).$$

$$c_1 = N_1[8N_1 + C_{11}(20C_{44} + 2C_{111} - 2C_{112} - 2C_{144} + 2C_{155}) + 4C_{12}(-C_{112} + C_{123})] + 2C_{11}C_{44}[C_{11}(C_{111} - C_{112}) + 2C_{12}(-C_{112} + C_{123})].$$

$$c_2 = \left(\frac{16}{(2C_{44} - C_{11})}\right)\left[\begin{array}{l} C_{11}^3C_{44}(-6C_{456} + 11C_{44}) + C_{11}^2C_{44}^2(-C_{155} + 29C_{44} + 15C_{456}) \\ +2C_{11}C_{44}^3(-13C_{44} + 3C_{112} + 5C_{155} - 8C_{456}) + C_{44}^4(-5C_{112} - 4C_{155} + 8C_{44}) \end{array}\right].$$

$$d_0 = -N_1^2C_{44}.$$

$$d_1 = 6C_{12}^5 + C_{12}^4(5C_{144} - 5C_{11} - C_{155}) - C_{12}^3C_{11}(4C_{144} + C_{111} + 14C_{11}) + C_{12}^2C_{11}^2(12C_{11} + C_{111} - C_{112} + 2C_{155} - 6C_{144}) + C_{12}C_{11}^3(2C_{144} + C_{111} + 8C_{11} - C_{132}) + C_{11}^4(C_{144} - 7C_{11} + C_{112} - C_{111} - C_{155}).$$

$$d_2 = \frac{3}{2}C_{12}^5 + C_{12}^4(2C_{112} - C_{155}) + C_{12}^3C_{11}\left(C_{144} - 2C_{123} - C_{111} - \frac{7}{2}C_{11} - 3C_{112}\right) + C_{12}^2C_{11}^2(2C_{123} + C_{112} - 2C_{11} + C_{111} + 2C_{155}) + C_{12}C_{11}^3\left(2C_{123} + C_{112} + \frac{9}{2}C_{11} + C_{111} - 2C_{144}\right) + C_{11}^4(2C_{11} - C_{111} - C_{155} - 2C_{123} - 3C_{112}) + \left(\frac{C_{11}^5}{C_{12}}\right)\left(2C_{112} + C_{144} - \frac{5}{2}C_{11}\right).$$

$$A_{R12}^{(1)} = \frac{1}{2N_3}\left[N_1\frac{\alpha_1}{\alpha_0} + N_2\left(\frac{\alpha_2}{\alpha_0} + 1 - \frac{C_{11}}{C_{12}}\right)\right].$$

$$A_{R12}^{(3)} = \frac{1}{2N_3}\left[N_2\frac{\alpha_1}{\alpha_0} + N_1\left(\frac{\alpha_2}{\alpha_0} + 1 - \frac{C_{11}}{C_{12}}\right)\right].$$

$$\alpha_1 = -\left(\frac{a_1\alpha_0^3 + b_1\alpha_0^2 + c_1\alpha_0 + d_1}{3a_0\alpha_0^2 + 2b_0\alpha_0 + c_0}\right).$$

$$\alpha_2 = -\left(\frac{a_2\alpha_0^3 + b_2\alpha_0^2 + c_2\alpha_0 + d_2}{3a_0\alpha_0^2 + 2b_0\alpha_0 + c_0}\right).$$

APPENDIX B: MODELING DATA FOR THE STRESS COMPUTATIONS

The chemical composition of the glass is given in Table II. The thermal and mechanical properties are summarized in Table III. T_g is the glass transition temperature in K, $k_h(T)$ is the thermal conductivity in W/m/K, $c_p(T)$ is the specific heat in J/kg/K, ρ is the density of glass in kg/m^3, ν is Poisson ratio, and E is Young modulus in Pa. The viscoelastic properties of the glass used can be found in Table IV.

APPENDIX C: GENERAL EQUATIONS FOR THE STRESS COMPUTATIONS

Glass stress relaxation can be correctly modeled with the viscoelastic generalized Maxwell law, using shear $G(t)$ and bulk $K(t)$ moduli functions of time t, described with instantaneous values (G_g, K_g) and differed moduli (G_e, K_e) expanded into Prony's series of six terms:[17]

$$G(t) = 2G_g\sum_{i=1}^{6} \mathrm{W}_i \exp\left(\frac{-t}{\tau_i}\right), \tag{C1}$$

$$K(t) = 3K_e - (3K_e - 3K_g)\sum_{i=1}^{6} \mathrm{W}_i \exp\left(\frac{-t}{6\tau_i}\right), \tag{C2}$$

where W_i and τ_i are, respectively, the weights and relaxation times assumed to be constant with temperature. The shear differed value G_e is considered equal to zero.

These relaxation functions are known for the reference temperature T_r and thus can be determined for any temperature, since glass behaves in the temperature range like a thermo rheological simple material. Temperature variation is then replaced by time variation as reduced time $\xi(x,t)$ is introduced:

$$\xi(x,t) = \int_0^t \exp\left(-\frac{H}{R}\left[\frac{1}{T(x,t')} - \frac{1}{T_r}\right]\right)dt'. \tag{C3}$$

In Eq. (C3), R is the ideal gas constant, H is the activation energy, x is the position in the glass, and t' is the integration variable.

TABLE II. Chemical composition of soda-lime silicate glass (%).

SiO_2	CaO	Na_2O	MgO	Fe_2O_3	Other oxides
71.05	9.90	13.60	4.00	0.09	1.35

TABLE III. Thermal and mechanical properties.[a]

T_g(K)	824.4
k_h(W/m/K)	$0.975+5.58\times10^{-4}(T-273.15)$
c_p(J/kg/C)	$T<T_g$: $893+0.4\times T-1.8\times10^7/T^2$
	$T>T_g$: $1433+6.5\times10^{-3}T$
α_g(/K)	9×10^{-6}
α_1(/K)	25×10^{-6}
ρ(kg/m^3)	2550
ν	0.22
E(Pa)	$T<T_g$: $7.1\times10^{10}-4.916\times10^6T$
	$T>T_g$: $9.67\times10^{10}-3.611\times10^7T$

[a]See Ref. 24.

Narayanaswamy[15] established a structural relaxation model with the so-called fictive temperature $T_f(x,t)$ that takes both multiple relaxation mechanisms (memory) and nonlinearity into account. The deviation of the glass from its state of equilibrium due to structural relaxation can be estimated. Structural relaxation in glass can be monitored by measuring changes in certain macroscopic properties. The response $M_p(x,t)$ of any property $p(t)$ to a step change in temperature from T_1 to T_2 in the transition region is time dependent and may be written as

$$M_p(x,t) = \frac{p(t) - p_1(\infty)}{p_g(0) - p_1(\infty)} = \frac{T_f(x,t) - T_2}{T_1 - T_2}. \tag{C4}$$

The subscripts 0 and ∞ denote the instantaneous initial and long-time value of p. $p_g(0)$ is the property value immediately following the temperature change; afterward, the property relaxes in time toward the equilibrium liquid value $p_1(\infty)$.

The formulation of the structural model is completed with the following definition of fictive temperature. The fictive temperature $T_f(x,t)$ can be calculated from Eq. (C4), by presenting the relaxation function as a function of time and temperature, using the Boltzmann superposition principle and integration over the thermal history:

$$T_f(x,t) = T_f(x,t) - \int_0^t M_p[\xi(x,t) - \xi(x,t')]\frac{dT(x,t')}{dt'}dt'. \tag{C5}$$

Reduced time $\xi(x,t)$ depends on $T_f(x,t)$:

TABLE IV. Typical viscoelastic properties.[a]

W_i	τ_i(s)	Properties	Values
0.0427	19	H/R(K)	150000/1.9872
0.0596	291.9	X	0.5
0.0877	1843	T_r(K)	763.66
0.2454	11,800	K_e/K_g	0.3
0.2901	49,490		
0.2498	171,700		

[a]See Ref. 16.

$$\xi(x,t) = \int_0^t \exp\left(-\frac{H}{R}\left[\frac{X}{T(x,t')} + \frac{1-X}{T_f(x,t')} - \frac{1}{T_r} \right] \right) dt', \tag{C6}$$

where X is a constant parameter and exp( ) denotes the exponential function.

By analogy with the viscous relaxation [Eq. (C1)], the relaxation function is represented as series of six exponential relaxation functions:

$$M_p[\xi(x,t)] = \sum_{i=1}^{6} W_i \exp\left(\frac{-\xi(x,t)}{9\tau_i} \right). \tag{C7}$$

The Markowsky and Soules algorithm[17] [Eqs. (C8) and (C11)] is used to calculate the fictive temperature $T_f(x,t)$:

$$T_{fi}(x,0) = T_0, \tag{C8}$$

$$\tau_{i\,\mathrm{fic}} = 9\tau_i \exp\left(\frac{H}{R}\left[\frac{X}{T(x,t)} + \frac{1-X}{T_f(x,t-\Delta t)} - \frac{1}{T_r} \right] \right), \tag{C9}$$

$$T_{\mathrm{fi}}(x,t) = \frac{\left(T_{\mathrm{fi}}(x,t-\Delta t) + \dfrac{T(x,t)\Delta t}{\tau_{i\,\mathrm{fic}}} \right)}{\left(1 + \dfrac{\Delta t}{\tau_{i\,\mathrm{fic}}} \right)}, \tag{C10}$$

$$T_f(x,t) = \sum_{i=1}^{6} W_i T_{\mathrm{fi}}(x,t). \tag{C11}$$

The response function $M_q(x,t)$ [Eq. (C4)] is expressed as a sum of exponential relaxations using the differential form of Eq. (C5). Initial values of the partial fictive temperature $T_{\mathrm{fi}}(x,t)$ are assigned to T_0, and numerically integrated during the transient analysis using an implicit scheme. Δt is a time increment of a step change in temperature.

Finally, thermal strains $\epsilon_{\mathrm{th}}(x,t)$ induced by tempering of the glass are given by

$$\epsilon_{\mathrm{th}}(x,t) = \alpha_g[T(x,t) - T_0] + \alpha_s[T(x,t) - T_f(x,t)], \tag{C12}$$

where α_g is the thermal expansion coefficient of the glass in the solid state. α_s is the structural expansivity responding to $\alpha_s = \alpha_1 - \alpha_g$ with α_1 the thermal expansion coefficient of the liquid glass.

[1] H. Aben and C. Guillemet, *Photoelasticity of Glass* (Springer, Berlin, 1993).

[2] A. S. Redner, "Automated measurement of edge stress in automotive glass," Glass Processing Days 2003, edited by J. Vitkala, in *Proceedings of the 8th International Conference on Architectural and Automative Glass*, Tampere, Finland (2003), pp. 578–579.

[3] J. Anton and H. Aben., "A compact scattered light polariscope for residual stress measurement in glass plates," edited by J. Vitkala, in *Glass Processing Days 2003, Proceedings of the 8th International Conference on Architectural and Automative Glass*, Tampere, Finland (2003), pp. 86–88.

[4] H. Aben., J. Anton, and A. Errapart, "Measurement of tempering stresses in axisymmetric glass articles," edited by D. Lochegnies, in *M.G.F.T. 2002: Modeling of Glass Forming and Tempering, Proceedings of the 2nd International Colloquium*, Valenciennes, France (2002), pp. 276–282.

[5] B. Cannon, C. Shepard, and M. Khaleel, "Stress measurements in glass by use of double thermal gratings," J. Opt. A, Pure Appl. Opt. **40**, 5354–5369 (2001).

[6] D. Cavaillé, C. Levelut, R. Vialla, R. Vacher, and E. Le Bourhis, "Third-order elastic constants determination in silico-soda-lime glass investigated by Brillouin scattering," J. Non-Cryst. Solids **260**, 235–241 (1999).

[7] Y. H. Pao, W. Sachse, and H. Fukuoka, *Acoustoelasticity and Ultrasonic Measurements of Residual Stress*, edited by Mason and Thurston (Academic, New York, 1984), pp. 61–143.

[8] M. Hirao., H. Fukuoka, and K. Hori, "Acoustoelastic effect of Rayleigh surface wave in isotropic material," ASME J. Appl. Mech. **48**, 119–124 (1981).

[9] M. Duquennoy, M. Ouaftouh, M. Ourak, and F. Jenot, "Theoretical determination of Rayleigh wave acoustoelastic coefficients: comparison with experimental values," Ultrasonics **39**, 575–583 (2002).

[10] M. Duquennoy, M. Ouaftouh, M. Ourak, and W. J. Xu, "Influence of natural and initial acoustoelastic coefficients on residual stress evaluation: Theory and experiment," J. Appl. Phys. **86**, 2490–2498 (1999).

[11] M. Duquennoy, M. Ouaftouh, and M. Ourak, "Ultrasonic evaluation of stresses in orthotropic materials using Rayleigh wave," NDT & E Int. **32**, 189–199 (1999).

[12] M. Landa and J. Plešek, "Ultrasonic techniques for nondestructive evaluation of internal stresses," *Proceedings of the 15th World Conference on Nondestructive Testing*, Rome, Italy (2000).

[13] E. Roméro, D. Lochegnies, and M. Duquennoy, "Numerical modeling, experimental investigation, stress control for flat glass tempering," Glass Processing Days 2001, edited by J. Vitkala, in *Proceedings of the 8th International Conference on Architectural and Automative Glass*, Tampere, Finland, pp. 468–471.

[14] D. Lochegnies, E. Roméro, A. Leriche, M. Duquennoy, and F. Monnoyer, "Numerical modelling, experimental investigation, stress control for flat glass tempering," edited by H. De Waal, R. G. C. Beerkens, J. Peelen, and J. H. A. Van de Woude, I.C.G. Annual Meeting Glass 2000, *Proceedings of the New Millenium Challenges and Break-Through Technologies Session T1*, Amsterdam, Netherlands, pp. 1–8.

[15] O. S. Narayanaswamy, "Stress and structural relaxation in glass," J. Am. Ceram. Soc. **61**, 146–152 (1978).

[16] S. Rekhson, *Viscoelasticity of Glass* (Academic, Orlando, 1986) Vol. 3, pp. 1–117.

[17] A. Markovsky and T. F. Soules, "An efficient and stable algorithm for calculating fictive temperatures," J. Am. Ceram. Soc. **67**, 56–57 (1984).

[18] D. Devos, M. Duquennoy, E. Roméro, F. Jenot, D. Lochegnies, M. Ouaftouh, and M. Ourak, "Ultrasonic evaluation of residual stresses in flat glass tempering by an original double interferometric detection," *World Congress on Ultrasonics 2005*, Ultrasonics International, Beijing, China (2005).

[19] F. Jenot, M. Ouaftouh, M. Duquennoy, and M. Ourak, "Interferometric detection of acoustic waves at air-solid interface applications to nondestructive testing," J. Appl. Phys. **97**, 094905-1–8 (2005).

[20] K. Jassby and D. Saltoun, "Use of ultrasonic Rayleigh waves for the measurement of applied biaxial surface stresses in aluminium 2024-T351 Alloy," Mater. Eval. **40**, 198–205 (1982).

[21] R. B. Thompson, S. S. Lee, and J. F. Smith, "Angular dependence of ultrasonic wave propagation in a stressed, orthorhobic continuum: theory and application to the measurement of stress and texture," J. Acoust. Soc. Am. **80**, 921–931 (1986).

[22] A. V. Clark, D. V. Fukuoka, and J. C. Mitrakovic, "Characterization of residual stress and texture in cast steel railroad wheels," Ultrasonics **24**, 281–288 (1986).

[23] K. Jassby, A. Aharoni, and J. Bouvast, "Determining surface residual stress in isotropic solids by combined ultrasonic velocity measurements," Ultrasonics Int. 79–85 (1985).

[24] C. Guillemet, R. Gy, and M. Labrot, "Viscosity, configurational entropy, and structural relaxation of a silica-soda-lime glass," in *Proceedings of the 16th International Congress on Glass 1992*, Madrid, Spain (1992), pp. 183–188.

In-plane vibration of thin circular structures submitted to pulsed microwave

Anis Redha Hadj Henni,[a] C. Bacon, and B. Hosten
Laboratoire de Mécanique Physique, Bordeaux 1 University, UMR C.N.R.S. 5469, 351 cours de la Libération, 33405 Talence Cedex, France

(Received 11 January 2006; revised 6 March 2006; accepted 17 March 2006)

The absorption of microwaves by material induces a temperature rise that generates a mechanical deformation in the irradiated structure by thermoelastic effect. As a continuous development of this recent generation technique, the present study investigates the in-plane mechanical vibration of disks and rings submitted to pulsed microwaves in order to develop a contactless evaluation method. The vibration of a disk submitted to microwave irradiations depends on the distribution of the electromagnetic field in the sample. If the irradiation is axisymmetric, the purely radial eigenmode is generated. On the contrary, if the irradiation field is not axisymmetric, the vibration is angular dependent. Consequently, one can choose to use various eigenfrequencies for determining the geometrical properties of a circular plate and the mechanical properties of the material. A semi-analytic model is presented to predict the in-plane vibration eigenfrequencies of thin rings and disks. Wave equations are solved by taking into account the angular dependence in order to estimate the maximum number of disk or ring eigenfrequencies. The model is validated by comparison between the theoretical and the experimental results and applied for the evaluation of rings and disks geometries and their constitutive material properties using the microwave excitation technique. [DOI: 10.1121/1.2195190]

PACS number(s): 43.40.Dx, 43.35.Zc, 43.40.Le [YHB] Pages: 3782–3792

I. INTRODUCTION

Since the works of Pochhammer[1] and Chree[2] in the 19th century on the vibration of infinite circular cylinders, the circular vibrating structures have been investigated by many authors. The problems of transverse vibrations of thin circular plates, generally excited by impact, have been largely studied, for instance by Lamb,[3] Carrington,[4] and Rayleigh,[5] for different boundary conditions.[6] Other authors concentrate their research on the vibration of thick circular plates, using the associated plate theory,[7,8] to predict their acoustic generation.[9] On the other hand, compared to the out-plane vibration, the in-plane vibration of disks has been studied by fewer authors, who principally investigated analytically the in-plane vibration eigenmodes of thin disks without experimental validation[10,11] or with a comparison to numerical results obtained by a finite element approach.[12] The purely radial vibration has been experimentally tested in electrically excited[13] piezoelectric disks or by inserting a small piezoelectric transducer in the center of the disk[14] in order to generate a pure radial elastic wave.

In this work, the in-plane vibration of thin disks will be studied theoretically and experimentally by a recent technique of microwave acoustic generation.[15,16] This method has been applied for studying one-dimensional structure vibrations[17,18] and for characterizing the mechanical and electromagnetic properties of different materials (viscoelastic, composites, piezoelectric,…).[19–23] Acoustic generation by microwaves is only possible in materials that absorb electromagnetic waves, otherwise the latter are reflected or totally transmitted and no thermoelastic effect is observable. The absorption of microwave energy by a material causes the vibration of its microstructure, producing material heating that induces thermal expansion and finally the generation of acoustic waves. The microwave excitation method allows the irradiation of a structure, and consequently the acoustic generation, in a large volume without contact. In addition, it offers the possibility to excite structures by short pulses in order to obtain the vibration responses on a frequency range going from a few kHz to several hundred kHz. Compared to the impact technique, the microwave excitation presents the advantage to be a contactless technique, allows us to force the vibration of a structure at a given frequency in order to observe its resonances, and permits us, during the signal treatment, to average the experimental responses since they are perfectly reproducible from shot to shot. Finally, the microwave excitation applied to acoustic generation in disks or rings permits us to excite either the purely radial eigenmode or several other vibration modes by using an axisymmetric irradiation or a shifted one from the center of the plate.

As a first attempt to apply the microwave generation technique to the resolution of forward and inverse problems for circular structure vibrations, in order to develop a method of nondestructive testing or evaluation (NDT, NDE), the simplest shapes, like ring and disk, are used.

First, a theoretical model that predicts the vibrations of thin circular structures is presented. Because of the small thickness in comparison with the lateral dimensions, a plane stress state is assumed in the model, and the Helmholtz decomposition method is used for describing the particular displacement fields in the structures.[24,25] The wave equations

[a]Author to whom correspondence should be addressed. Electronic mail: ar.hadjhenni@lmp.u-bordeaux1.fr

are solved by the mean of infinite Bessel and trigonometric series,[11,12] in accordance with the zero stress boundary conditions. Compared to a finite element model, the semianalytic model presented here has the advantages of being fast and easily introduced in the resolution of inverse problems. This model will be used to evaluate ring and disk eigenfrequencies and the radial displacement fields in order to compare them to those experimentally obtained by microwave excitation. As an application of the classical method of evaluation of structure characteristics starting from its experimental eigenfrequencies,[7,26,27] it will be shown that the combination of two eigenfrequencies furnishes enough information to solve the geometrical and material inverse problems. The possibility to only excite the purely radial eigenmodes by the microwave technique is very important for identifying these modes. This fact represents the main advantage of this technique. It will be explained how to extend the direct resolution and the inverse procedures developed here to the case of circular cylinders or tubes.

II. THEORY

A. Problem formulation and wave equations

The vibration of a thin plate will be studied by determining its eigenmodes of vibration and by calculating, when it is possible, the displacement field. The plate will be considered as a ring or a disk, and the material is assumed to be viscoelastic. The expression of a viscoelastic material behavior law in the time domain uses convolution products that are transformed to classical products in the frequency domain. In the following developments, all quantities are expressed in the frequency domain by applying a Fourier transform to the variables expressed in the time domain. Any exception will be mentioned.

Taking the symmetry of the problem into account, the circular cylindrical coordinate system $(O,\mathbf{e}_r,\mathbf{e}_\theta,\mathbf{e}_z)$ is chosen. The displacement field has the following vector form:

$$(U_r(r,\theta)\ U_\theta(r,\theta)\ U_z(z))^T. \tag{1}$$

Since all the fields are expressed in the frequency domain and depend on the angular pulsation ω, this latter will not appear in the following developments.

The assumption of plane stress state allows us to consider only the displacement field in the plane of the plate $(\mathbf{e}_r,\mathbf{e}_\theta)$. Then, one can choose for the displacement in this plane the following expression:

$$\mathbf{U} = (U_r(r,\theta)\ U_\theta(r,\theta))^T. \tag{2}$$

The material is supposed linearly viscoelastic and has complex coefficients that take into account the attenuation. The plate is submitted to a sudden rise of temperature $\Delta T(r,\theta)$, which is assumed to be symmetric with respect to the axis $\theta=0$ of the disk plane, as it will be realized by the experimental setup (Sec. III). After calculation, under the hypothesis of plane stress state, the stresses in the plane $(\mathbf{e}_r,\mathbf{e}_\theta)$ are

$$\begin{aligned}
\sigma_{rr} &= Q_{11}\varepsilon_{rr} + Q_{12}\varepsilon_{\theta\theta} - B_1\Delta T(r,\theta),\\
\sigma_{\theta\theta} &= Q_{12}\varepsilon_{rr} + Q_{22}\varepsilon_{\theta\theta} - B_2\Delta T(r,\theta),\\
\sigma_{r\theta} &= 2Q_{66}\varepsilon_{r\theta}.
\end{aligned} \tag{3}$$

In addition, if the material is supposed to be isotropic, and designating by E, ν, and α, respectively, the complex Young's modulus, the Poisson's ratio, and the coefficient of thermal dilatation, then the coefficients Q_{ij} and B_i are defined by

$$Q_{11} = Q_{22} = \frac{E}{1-\nu^2},\quad Q_{12} = Q_{21} = \frac{E\nu}{1-\nu^2},$$

$$Q_{66} = \frac{E}{2(1+\nu)},\quad \text{and}\quad B_1 = B_2 = \frac{E}{1-\nu}\alpha \tag{4}$$

and it is possible to write in a contracted way

$$\boldsymbol{\sigma} = Q_{12}(\boldsymbol{\nabla}\cdot\mathbf{U})\mathbf{I} + (Q_{11}-Q_{12})\boldsymbol{\varepsilon} - B_1\Delta T(r,\theta)\mathbf{I}, \tag{5}$$

where $\boldsymbol{\sigma}$ is the stress tensor whose components are given in Eq. (3), and $\boldsymbol{\varepsilon}$ is the strain tensor given by the following expression:

$$2\boldsymbol{\varepsilon} = \boldsymbol{\nabla}\mathbf{U} + (\boldsymbol{\nabla}\mathbf{U})^T. \tag{6}$$

The field function $\Delta T(r,\theta)$ is the Fourier transform of the temperature increase in the material due to the absorption of microwave irradiations. Its temporal form is modeled as follows:

$$\Delta T(t,r,\theta) = \Delta T(r,\theta) f(t), \tag{7}$$

where $\Delta T(r,\theta)$ and $f(t)$ represent, respectively, the space distribution and the temporal evolution.

If the temperature rise occurs without conduction and dissipation during a time τ, which is short compared to the mechanical response duration, then the temporal evolution $f(t)$ and its Fourier transform $\tilde{f}(\omega)$ can be expressed as follows:[20,21]

$$f(t) = \begin{cases} 0, & t \leqslant 0,\\ t/\tau, & 0 < t < \tau,\\ 1, & t \geqslant \tau, \end{cases} \qquad \text{and}$$

$$\tilde{f}(\omega) = -\frac{1}{\omega^2\tau}(1-e^{-i\omega\tau}) \quad (\omega \neq 0), \tag{8}$$

One can note that in the case of an infinite cylindrical structure, the hypothesis of plane stress state is replaced by the hypothesis of plane strain state.[24,25] Equation (3) will keep the same form if the coefficients Q_{ij}, given in Eq. (4), are replaced by the coefficients C_{ij} of the complex stiffness matrix $\mathbf{C}$. Then, the method presented in the following for thin rings and disks can be used in the same way for studying the vibration of infinite cylinders and pipes.

By taking into account the local equilibrium equation,

$$\boldsymbol{\nabla}\cdot\boldsymbol{\sigma} = -\rho\omega^2\mathbf{U}, \tag{9}$$

the material behavior law (5), and the kinematic relation (6), it is possible to obtain the displacement differential Navier equation. In an undefined system of coordinates and under

the hypothesis of plane stress state, this equation is written

$$Q_{11}\,\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{U}) - \frac{Q_{11}-Q_{12}}{2}\,\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{U}) + \rho\omega^2\mathbf{U} = B_1\,\boldsymbol{\nabla}(\Delta T(r,\theta)). \tag{10}$$

The resolution of the circular plate vibration problem depends on the space distribution of temperature rise, which is itself related to the microwave excitation field. The knowledge of the excitation field is necessary to formulate the researched solution as the sum of a homogeneous solution (independent of temperature distribution) and a particular one (dependent on the excitation). Here, it is unknown as the second member of Eq. (10) is unknown.

Considering the difficulty to predict the temperature spatial distribution, the displacement field will only be determined for the case of a uniform temperature excitation. In this case, the temperature rise gradient is null and Eq. (10) becomes homogeneous and analytically solvable:

$$Q_{11}\,\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{U}) - \frac{Q_{11}-Q_{12}}{2}\,\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{U}) + \rho\omega^2\mathbf{U} = \mathbf{0}. \tag{11}$$

In contrast, the search of the structure vibration eigenmodes does not impose the preliminary knowledge of the excitation profile. In fact, it will be sufficient to solve the homogeneous equation (11) since the eigenfrequencies are contained in the homogeneous solutions. Details of this approach are given in Sec. II D.

In the following, the resolution of Eq. (11) is presented in the general case by taking the angular dependence into account in order to evaluate a maximum number of eigenfrequencies of circular plates submitted to microwave irradiations. The calculation of the displacement fields will be achieved only for the case of a uniform excitation in order to visualize and compare the experimental and computed spectra together.

A Helmholtz decomposition into potentials[24,25] is applied to the displacement field of Eq. (11),

$$\mathbf{U} = \boldsymbol{\nabla}\varphi + \boldsymbol{\nabla}\times\boldsymbol{\psi}, \tag{12}$$

where φ and ψ are, respectively, the scalar and the vectorial potentials of the displacement.

Since the displacement field $\mathbf{U}$ does not depend on the z coordinate, the vector $\boldsymbol{\psi}$ is given by

$$\boldsymbol{\psi} = (0 \quad 0 \quad \psi)^T. \tag{13}$$

Finally, the two potential functions verify the following wave equations,

$$\boldsymbol{\nabla}^2\varphi + k_L^2\varphi = 0,$$

$$\boldsymbol{\nabla}^2\boldsymbol{\psi} + k_T^2\boldsymbol{\psi} = 0, \tag{14}$$

where $k_L=\omega/c_L$, $k_T=\omega/c_T$, $c_L^2=E/\rho(1-\nu^2)=Q_{11}/\rho$ and $c_T^2=E/2\rho(1+\nu)=Q_{66}/\rho$ are, respectively, the longitudinal and transverse wave numbers, and the longitudinal and transverse waves velocities in the material.

The differential equations (14) are Helmholtz equations that are solvable in a circular cylindrical system of coordinates by the technique of separation of variables.[24,25]

Since the Helmholtz equation is separable in the cylindrical system of coordinates, one can apply the technique of variable separation to the functions φ and ψ.[24,25] By calling $\varphi_r(r)$ and $\psi_r(r)$ the radial parts of the potentials and $\varphi_\theta(\theta)$ and $\psi_\theta(\theta)$ their angular parts, one can find after calculation the following differential equations:

$$\frac{\partial^2\varphi_\theta}{\partial\theta^2} + n^2\varphi_\theta = 0,$$

$$r^2\frac{\partial^2\varphi_r}{\partial r^2} + r\frac{\partial\varphi_r}{\partial r} + ((rk_L)^2 - n^2)\varphi_r = 0, \tag{15}$$

$$\frac{\partial^2\psi_\theta}{\partial\theta^2} + n^2\psi_\theta = 0,$$

$$r^2\frac{\partial^2\psi_r}{\partial r^2} + r\frac{\partial\psi_r}{\partial r} + ((rk_T)^2 - n^2)\psi_r = 0. \tag{16}$$

The first angular differential equations have solutions if, and only if, n^2 is a positive integer and the second ones are Bessel equations of order n and variable (rk_L) or (rk_T).

B. Resolution of the differential equations for a ring

1. Potentials forms

The circular ring, of external radius R_2, is hollowed in its center by a circular hole of radius R_1. Considering the fact that the disk contains a geometrical and material discontinuity (due to the presence of a hole), the solutions of the Bessel equations will be a combination of first and second kind Bessel functions.[24,25] The general scalar and vectorial potentials are infinite series of n-order products of the angular and radial solutions of Eqs. (15) and (16):[11]

$$\varphi(r,\theta) = \sum_{n=0}^{\infty}[A_n^\varphi J_n(k_L r) + B_n^\varphi Y_n(k_L r)] \times [C_n^\varphi\cos(n\theta) + D_n^\varphi\sin(n\theta)],$$

$$\psi(r,\theta) = \sum_{n=1}^{\infty}[A_n^\psi J_n(k_T r) + B_n^\psi Y_n(k_T r)] \times [C_n^\psi\cos(n\theta) + D_n^\psi\sin(n\theta)], \tag{17}$$

where the sinusoidal parts are the solutions of the angular differential equations, and the functions $J_n(\,)$ and $Y_n(\,)$ are, respectively, the first and second kind n-order Bessel functions.

Considering the spatial symmetry of the excitation supposed in the formulation of the problem, it appears that the following relations are true in all angular and radial positions:

$$\begin{aligned} U_r(r,\theta) &= U_r(r,-\theta), \\ U_\theta(r,\theta) &= -U_\theta(r,-\theta), \end{aligned} \qquad \forall\,(r,\theta)\in[R_1,R_2]\times[0,2\pi]. \tag{18}$$

By taking into account Eqs. (12) and (18) and after calculation, it appears that, for any integer n,

$$D_n^{\varphi} = 0 \text{ and } C_n^{\psi} = 0. \tag{19}$$

The general forms of potentials are simplified and become, after renaming the coefficients,

$$\begin{aligned}\varphi(r,\theta) &= \sum_{n=0}^{\infty} [A_n J_n(k_L r) + B_n Y_n(k_L r)]\cos(n\theta),\\ \psi(r,\theta) &= \sum_{n=1}^{\infty} [C_n J_n(k_T r) + D_n Y_n(k_T r)]\sin(n\theta),\end{aligned} \tag{20}$$

where the unknown coefficients A_n, B_n, C_n, and D_n have to be determined. One can see that the scalar potential, which is the irrotational part of the displacement, and the vectorial potential, representing the rotational part of the movement, are respectively even and odd functions of the variable θ.

2. Boundary conditions

The coefficients A_n, B_n, C_n, and D_n are determined by taking into account the zero stress boundary conditions at the boundaries of the structure. The recurrence relations between Bessel functions and their derivatives[28,29] allow us to write the displacement field (12) under the following form:

$$\begin{aligned}U_r = &\sum_{n=0}^{\infty}\left[A_n\left(\frac{n}{r}J_n(k_L r) - k_L J_{n+1}(k_L r)\right)\right.\\ &\left. + B_n\left(\frac{n}{r}Y_n(k_L r) - k_L Y_{n+1}(k_L r)\right)\right]\cos(n\theta)\\ &+ \sum_{n=1}^{\infty}\left[C_n\frac{n}{r}J_n(k_T r) + D_n\frac{n}{r}Y_n(k_T r)\right]\cos(n\theta),\\ U_\theta = &-\sum_{n=0}^{\infty}\left[A_n\frac{n}{r}J_n(k_L r) + B_n\frac{n}{r}Y_n(k_L r)\right]\sin(n\theta)\\ &-\sum_{n=1}^{\infty}\left[C_n\left(\frac{n}{r}J_n(k_T r) - k_T J_{n+1}(k_T r)\right)\right.\\ &\left. + D_n\left(\frac{n}{r}Y_n(k_T r) - k_T Y_{n+1}(k_T r)\right)\right]\sin(n\theta).\end{aligned} \tag{21}$$

In order to compute numerically a maximum number of eigenmodes contained in the homogeneous solution (21), the condition of zero boundary stress is expressed in the case of an arbitrary thermal distribution (see Sec. II D):

$$\boldsymbol{\sigma}\cdot\mathbf{n}_i = \mathbf{0} \quad \text{at } r = R_i \quad (i = 1,2), \tag{22}$$

where $\mathbf{n}_i$ are the outward unit vectors normal to the boundaries.

By remembering Eq. (3), which relates the stress tensor components to the strains, and the kinematic relation (6), one can rewrite the boundary conditions (22) as follows:

$$\begin{aligned}&\sum_{n=0}^{\infty} f_n^J(k_L R_i,\theta)A_n + \sum_{n=0}^{\infty} f_n^Y(k_L R_i,\theta)B_n + \sum_{n=1}^{\infty} g_n^J(k_T R_i,\theta)C_n\\ &\quad + \sum_{n=1}^{\infty} g_n^Y(k_T R_i,\theta)D_n = (1+\nu)\alpha\Delta T(R_i,\theta),\\ &\sum_{n=0}^{\infty} h_n^J(k_L R_i,\theta)A_n + \sum_{n=0}^{\infty} h_n^Y(k_L R_i,\theta)B_n + \sum_{n=1}^{\infty} l_n^J(k_T R_i,\theta)C_n\\ &\quad + \sum_{n=1}^{\infty} l_n^Y(k_T R_i,\theta)D_n = 0,\end{aligned} \tag{23}$$

with

$$\begin{aligned}f_n^J(k_L r,\theta) = &\left[\frac{n(n-1)(1-\nu)}{r^2}J_n(k_L r)\right.\\ &-\frac{(2n+1+\nu)k_L}{r}J_{n+1}(k_L r)\\ &\left.+k_L^2 J_{n+2}(k_L r)\right]\cos(n\theta), \quad n = 0,\infty,\end{aligned}$$

$$\begin{aligned}g_n^J(k_T r,\theta) = &\left[\frac{n(n-1)(1-\nu)}{r^2}J_n(k_T r_2)\right.\\ &\left.-\frac{n(1-\nu)k_T}{r}J_{n+1}(k_T r)\right]\cos(n\theta), \quad n = 1,\infty\end{aligned}$$

$$\begin{aligned}h_n^J(k_L r,\theta) = &\left[-\frac{n(n-1)}{r^2}J_n(k_L r)\right.\\ &\left.+\frac{nk_L}{r}J_{n+1}(k_L r)\right]\sin(n\theta), \quad n = 0,\infty,\end{aligned}$$

$$\begin{aligned}l_n^J(k_T,r,\theta) = &\left[-\frac{n(n-1)}{r^2}J_n(k_T r)+\frac{nk_T}{r}J_{n+1}(k_T r)\right.\\ &\left.-\frac{k_T^2}{2}J_{n+2}(k_T r)\right]\sin(n\theta), \quad n = 1,\infty,\end{aligned}$$

$$\begin{aligned}f_n^Y(k_L r,\theta) = &\left[\frac{n(n-1)(1-\nu)}{r^2}Y_n(k_L r)\right.\\ &-\frac{(2n+1+\nu)k_L}{r}Y_{n+1}(k_L r)\\ &\left.+k_L^2 Y_{n+2}(k_L r)\right]\cos(n\theta), \quad n = 0,\infty,\end{aligned}$$

$$\begin{aligned}g_n^Y(k_T r,\theta) = &\left[\frac{n(n-1)(1-\nu)}{r^2}Y_n(k_T r_2)\right.\\ &\left.-\frac{n(1-\nu)k_T}{r}Y_{n+1}(k_T r)\right]\cos(n\theta), \quad n = 1,\infty,\end{aligned}$$

$$h_n^Y(k_L r,\theta)=\left[-\frac{n(n-1)}{r^2}Y_n(k_L r)+\frac{nk_L}{r}Y_{n+1}(k_L r)\right]\sin(n\theta),\quad n=0,\infty,$$

$$l_n^Y(k_T r,\theta)=\left[-\frac{n(n-1)}{r^2}Y_n(k_T r)+\frac{nk_T}{r}Y_{n+1}(k_T r)-\frac{k_T^2}{2}Y_{n+2}(k_T r)\right]\sin(n\theta),\quad n=1,\infty,$$

Finally, under the action of a uniformly distributed rise of temperature, the problem of a ring radial vibration is analytically solvable. Because of the uniformity and the symmetry of the excitation, the displacement field does not depend on the angular spatial variable θ and is a purely radial vibration mode (fundamental mode). Mathematically, the displacement field is obtained from expressions (21) for $n=0$. The angular component of the displacement is then zero and the coefficient set to be determined is reduced to the unique coefficients A_0 and B_0, calculated from system (23) which is reduced to two scalar equations. Considering the recurrence relations between Bessel functions,[29] the displacement field is given by the following expression,

$$U_r(r)=\frac{2(1+\nu)\alpha\Delta T(\omega)}{[Y(R_1)J(R_2)-Y(R_2)J(R_1)]}\{[Y(R_1)-Y(R_2)]J_1(k_L r)-[J(R_1)-J(R_2)]Y_1(k_L r)\},\quad (24)$$

where the functions $J(r)$ and $Y(r)$ are defined by

$$J(r)=k_L J_0(k_L r)+\frac{2\nu}{r}J_1(k_L r)-k_L J_2(k_L r),\quad (25)$$

$$Y(r)=k_L Y_0(k_L r)+\frac{2\nu}{r}Y_1(k_L r)-k_L Y_2(k_L r).\quad (26)$$

C. Resolution of the differential equations for a disk

The solution for a disk is treated as a particular case of the ring solution. Here the interior radius R_1 is zero and the exterior radius becomes R. Considering the fact that the disk is full, the displacement in its center is finite and, consequently, the solutions of the Bessel equations can only be of the first kind[24,25] since the second Bessel functions tend to infinity at zero. In addition, since the disk is symmetrically excited with respect to the axis $\theta=0$, the physical conditions on displacement (18) are respected. Finally, the displacement in a disk is simply deduced from expression (21) by setting to zero the coefficients B_n and D_n:

$$\begin{aligned}U_r&=\sum_{n=0}^{\infty}A_n\left(\frac{n}{r}J_n(k_L r)-k_L J_{n+1}(k_L r)\right)\cos(n\theta)+\sum_{n=1}^{\infty}C_n\frac{n}{r}J_n(k_T r)\cos(n\theta),\\ U_\theta&=-\sum_{n=0}^{\infty}A_n\frac{n}{r}J_n(k_L r)\sin(n\theta)-\sum_{n=1}^{\infty}C_n\left(\frac{n}{r}J_n(k_T r)-k_T J_{n+1}(k_T r)\right)\sin(n\theta),\end{aligned}\quad (27)$$

where the coefficients A_n and C_n are determined by taking into account the boundary conditions expressed by the system of equations (23) and adapted to the case of the disk by setting to zero the coefficients B_n and D_n.

As for the case of the ring, the resolution of a disk vibration problem submitted to a uniform temperature rise is analytically possible. The uniform excitation implies a purely radial displacement field corresponding to the fundamental mode ($n=0$). The coefficient A_0 is calculated from the system (23) adapted to the case of disk. Finally, the exact radial displacement field is

$$U_r(r)=\frac{2(1+\nu)\alpha\Delta T(\omega)}{[k_L J_0(k_L R)+(2\nu/R)J_1(k_L R)-k_L J_2(k_L R)]}J_1(k_L r).\quad (28)$$

D. Numerical calculation of eigenfrequencies

Since the thermal field of excitation is unknown, only eigenfrequencies of ring and disk vibrations can be calculated. In the next development, the method of evaluation of ring eigenfrequencies is presented in details. As for the case of disks, this method keep unchanged.

The ring boundaries are discretized in $2N$ points (N points for each border) where the system of equations (23) is verified. The total number of equations is $4N$ and, in order to obtain a square system of equations, the infinite series are truncated until the Nth unknown. The discretized system will take the following form:

$$[M(R_1,R_2,\theta_i,\omega)](A)=(b),\quad (29)$$

where i is the calculation point index which varies from 1 to N, $[M(R_1,R_2,\theta_i,\omega)]$ is a square matrix ($4N\times 4N$) composed of the functions f_n, g_n, h_n, and l_n given in Eq. (23) and evaluated at each point, (A) is a $4N$ vector containing the unknown coefficients A_n, B_n, C_n, and D_n, and (b) is a second member vector, of dimension $4N$, which contains the terms of the unknown thermal excitation.

The second member vector in Eq. (29) is undetermined. However, the ring eigenfrequencies should be determined from the homogeneous system of equations associated to system (23),

$$[M(R_1,R_2,\theta_i,\omega)](A)=(0).\quad (30)$$

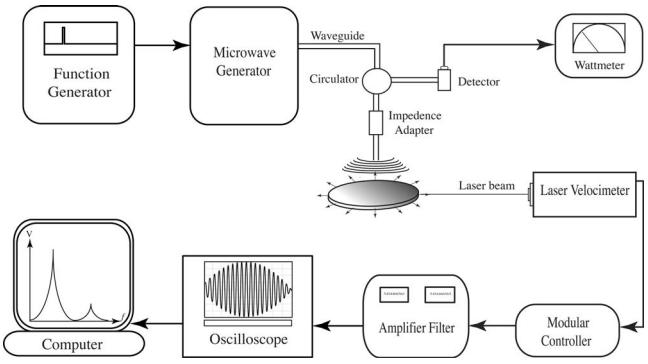


FIG. 1. Experimental setup for exciting and measuring the vibrations of samples submitted to microwaves irradiations.

Since the system (30) admits a solution if, and only if, the determinant of the matrix $[M(R_1,R_2,\theta_i,\omega)]$ is zero, the numerical resolution of the equation,

$$\det[M(R_1,R_2,\theta_i,\omega_k)]=0, \quad k=1,\ldots,\infty, \tag{31}$$

should provide the eigenpulsations of the ring vibration.

Unfortunately, the matrix M is ill-conditioned and the resolution of system (31) is numerically unstable. In order to get rid of this difficulty, the research of eigenfrequencies is performed by the resolution of system (29) corresponding to a *fictitious* nonuniform excitation associated to a *fictitious* displacement field. In addition, this permits us to simulate an excitation that produces the various eigenfrequencies of the disk in addition to those of the purely radial mode. So it is sufficient to calculate the fictitious displacement (or velocity) response on a certain frequency range and to extract the resonance frequencies that appear as peaks in the spectrum.

III. EXPERIMENTAL SETUP

The experimental setup schematized in Fig. 1 allows us to excite the structures by pulsed microwave irradiations. A magnetron, working at the frequency of 9.41 GHz, with a peak power of 5.5 kW, produces a 1–μs electromagnetic pulse or a burst made of n pulses. The number n and the repetition frequency f_{rep} of the pulses are chosen with a function generator (Agilent 33120 A). The temporal evolution of the temperature rise in the irradiated sample and corresponding to one microwave pulse is modeled by the formula of Eqs. (7) and (8). Using a burst, the resonances of the sample can be excited and searched by adjusting the pulses repetition frequency f_{rep} in the burst to particular resonance frequencies.

The microwaves are propagating into a circular waveguide, of internal diameter equal to 25 mm, under the fundamental mode TE_{11}.

In order to obtain a maximum power in the sample, an impedance adaptor is adjusted to get the minimum reflected power measured with a wattmeter. A circulator permits us to direct the microwaves reflected by the sample to the detector and finally to the wattmeter.

The particular velocity on the lateral surface of the sample is measured with a laser velocimeter (Polytec OFV 353) related to a modular controller (Polytec OFV 3001) with a sensitivity of 5 mm/s/V in a frequency bandwidth limited to 250 kHz. The signal is acquired and averaged, through a filtering amplifier (Electronic instrument 3627), by a numerical oscilloscope (LeCroy 9450 A), and sent to a computer to be treated. Then, for obtaining the spectral response, the measured temporal velocities are Fourier transformed.

In order to excite a maximum number of eigenmodes in addition to the purely radial ones, the microwave irradiation is decentered.

Four samples have been designed using epoxy-based resins charged with graphite. Indeed, graphite enhances the microwaves' absorption and the temperature rise in order to maximize the vibration amplitude of the tested samples. The names, the geometrical properties, and the densities of the disks and rings are given in Table I.

TABLE I. Geometric properties and densities of the designed disks and rings.

	Outside radius (mm)	Inside radius (mm)	Density (kg/m^3)
Disk 1 Di_1	31.5	⋯	1150
Disk 2 Di_2	40.0	⋯	1250
Ring 1 Ri_1	31.5	10.0	1150
Ring 2 Ri_2	40.0	15.0	1150

The epoxy resin is viscoelastic. To take into account the dissipative aspect of the wave propagation, a complex Young's modulus is used for which the imaginary part represents the damping of the material. To estimate the Young's modulus, a rod was elaborated in the material having density of 1150 kg/m^3 and submitted to the microwave excitation. Using a one-dimensional model,[19] the Young's modulus was evaluated to $E=(3.11+0.10i)$ GPa with an accuracy around 0.05 GPa for the real part. This technique does not allow the measurement of the Poisson's ratio. However, it will be shown in the following how to estimate it from the disc eigenfrequencies (inverse problem). Then, in the model validation (forward problem), the value of the Poisson's ratio is taken to the estimated value $\nu=0.39$. Finally, the thickness of the designed plates equals 3 mm.

IV. FORWARD PROBLEM, MODEL VALIDATION

Comparing the theoretical results with the experimental ones will first validate the model.

In order to validate the hypothesis of the plane stress state, 3D finite element simulations have been performed on a disk, the thickness of which is 3 mm, excited by a thermal source. Compared to the theoretical bi-dimensional model, the simulations have shown that the eigenfrequency modifications due to three-dimensional effects are less than the experimental errors tied to the reproducibility of the measurements. Consequently, one can model the mechanical behavior of the tested plates assuming a plane stress state. In addition, considering the polymeric nature of the tested materials and the thickness of the designed plates (see Sec. III), one can estimate the temperature variation[16] in the thickness of the plates to less than 0.5%.

TABLE II. Experimental and theoretical vibrational eigenfrequencies of disks (frequencies in bold correspond to the purely radial mode).

Disk Di_1		Disk Di_2	
Experimental (kHz)	Theoretical (kHz)	Experimental (kHz)	Theoretical (kHz)
11.9	11.8	9.8	9.7
14.1	14.1	11.5	11.5
19.3	**19.3**	**15.7**	**15.7**
22.1	21.9	19.7	19.4
24.2	23.7	23.8	23.6
29.0	28.9	26.0	26.0
31.7	31.8	28.9	28.7
35.4	35.2	32.5	32.3
39.3	39.6	38.1	37.9
46.4	46.5	**40.3**	**40.3**
49.1	**49.5**	44.0	43.9
57.3	57.4	46.6	46.8
63.1	63.3	51.5	51.4
67.0	66.7	55.4	55.7
78.4	**78.4**	**63.4**	**64.0**

TABLE III. Experimental and theoretical vibrational eigenfrequencies of rings (frequencies in bold correspond to the purely radial mode).

Ring Ri_1		Ring Ri_2	
Experimental (kHz)	Theoretical (kHz)	Experimental (kHz)	Theoretical (kHz)
7.6	7.8	5.4	5.4
14.4	**14.5**	**10.7**	**10.7**
16.1	16.5	11.7	11.9
21.2	21.6	17.1	17.2
23.2	23.9	22.7	23.4
29.1	29.8	27.4	28.2
34.9	35.2	29.6	29.8
38.1	38.1	34.9	34.8
40.3	40.8	36.0	36.7
42.2	42.6	**38.1**	**38.4**
46.2	**45.8**	44.7	45.2
57.7	57.1	53.2	53.3
66.6	67.8	...	...

Tables II and III give respectively the disk and ring eigenfrequencies obtained experimentally and by the model. These frequencies correspond to different eigenmodes of the samples. One can observe the good correlation between the experimental and the theoretical frequencies. The small differences can be due to the truncation error in the infinite series expressing the displacement field, and to the frequential discretization in the spectra. In addition, the sample supports affect their vibrations and the higher the eigenfrequency, the less reliable the measurements are. The frequencies presented in bold in Tables II and III correspond to the purely radial eigenfrequencies obtained from Eqs. (28) and (24), and experimentally by a centered microwave excitation.

Figure 2 represents the experimental and theoretical particular velocity spectra at the boundary of the two disks submitted to a centered microwave irradiation. The excitation produces a sudden rise of temperature that permits the excitation of a large spectral range. The analytical velocity is obtained by derivation, in the frequency domain, of the radial displacement given by Eq. (24) or (28):

$$\dot{U}_r(r,\theta) = i\omega U_r(r,\theta). \tag{32}$$

Figures 2(a) and 2(b) show the good correlation between the theoretical and the experimental velocity spectra for the disks. The relative amplitudes of particular velocities obtained analytically and experimentally do not perfectly coincide because of the indetermination in the exact distribution of the temperature rise which is, experimentally, axisymmetric rather than uniform. In addition, the residual noise contained in the experimental signals introduces an inaccuracy on the measured amplitudes. However, it is important to notice that the eigenfrequencies found by a nonuniform axisymmetric excitation correspond to those of the purely radial modes obtained by the model in the case of a uniform excitation.

Figures 3(a) and 3(b) shows the responses of the rings and a satisfying coincidence between the experimental and theoretical spectra. The spectrum of the ring Ri_2 is noisier

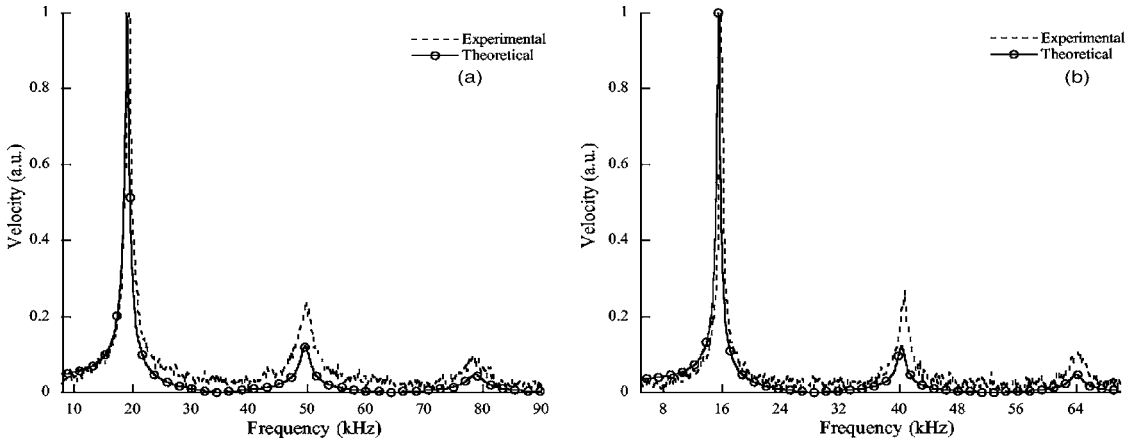


FIG. 2. Comparison of theoretical and experimental radial velocity spectra of disks excited by a centered microwave impulsion. (a) Di_1. (b) Di_2.

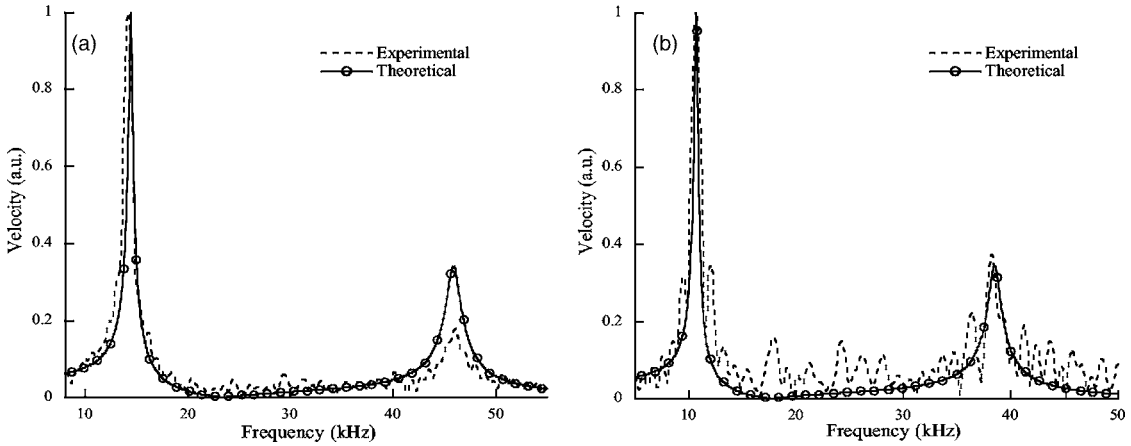


FIG. 3. Comparison of theoretical and experimental radial velocity spectra of rings excited by a centered microwave impulsion. (a) Ri_1. (b) Ri_2.

than the other responses because the diameter of the electromagnetic waveguide, used to excite the plate, is not large enough to irradiate a sufficient material area.

V. INVERSE PROBLEMS

In this section, the model will be exploited for the nondestructive evaluation of the characteristics of the material that constitutes the samples and their geometries.

In a first attempt to check the model and the experiment, characterization methods are applied to disks and rings. They could also be applied to cylinders to evaluate, for instance, the inside radius of the tube, which is not accessible by a direct measurement.

A. Material evaluation

Since the vibration of disks and rings depends on the properties of their constitutive material, a procedure is proposed in order to evaluate the real parts of Young's modulus E' and Poisson's ratios ν of isotropic viscoelastic materials. The choice of the eigenfrequencies used in the material evaluation is important. Indeed, the use of the two first eigenfrequencies of the purely radial mode in the inversion method detailed later reveals that it is not possible to evaluate simultaneously E' and ν with a good accuracy (around 10%).

There are probably various procedures to choose the appropriate frequencies to invert the data. Among them, it was found that the experimental frequency f_0 represented in Figs. 4 and 5 for disks Di_1 and Di_2 is not only sensitive to Poisson's ratio and Young's modulus, but is also easily identifiable. Indeed, this frequency appears just before the first radial frequency f_r in the frequency spectrum and is easily generated when the electromagnetic waveguide is located at the boundary of the excited disk. On the other hand, the first radial frequency f_r is easily identified by means of a test for which the electromagnetic waveguide is centered (see Fig. 2).

The eigenfrequencies f_r (19.3 and 15.7 kHz) and f_o (14.1 and 11.5 kHz) of disks Di_1 and Di_2 are given in Table II. Using the model for estimating the radial and the nonradial eigenfrequencies, one can search the pair values in the plane formed by the real part of Young's modulus E' and Poisson's ratio ν, which provide the experimental values of f_r and f_o/f_r.

The results of these simulations are given in Figs. 6 and

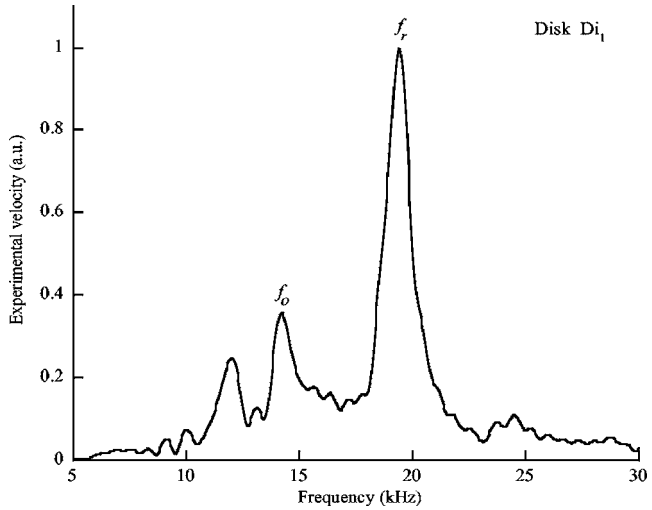


FIG. 4. Experimental velocity spectrum for a noncentered excitation (disk Di_1).

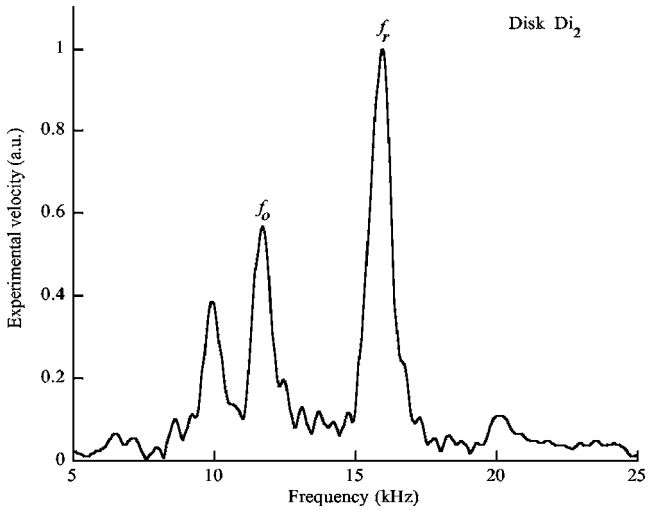


FIG. 5. Experimental velocity spectrum for a noncentered excitation (disk Di_2).

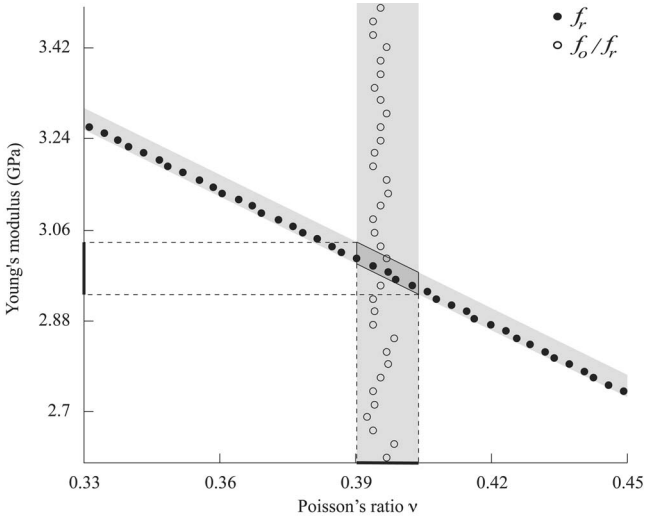


FIG. 6. Solution area giving f_r=19.3 kHz and f_o/f_r=0.73 for the disk Di_1.

7, respectively, for the disks Di_1 and Di_2. Starting from the accuracy on the measured frequencies f_r and f_o evaluated at ±0.05 kHz, one can estimate the minimum and maximum values that f_r and f_o/f_r can reach. The simulations corresponding to these limit values are performed in order to evaluate the accuracy of the evaluated material characteristics. It has been verified that all values of f_r and f_o/f_r between their minima and maxima, given by their accuracies, provide solutions inside the gray area shown in Figs. 6 and 7. The values of Young's modulus and Poisson's ratio are given in Table IV with the estimated accuracy given by the sizes of the intersection areas shown in Figs. 6 and 7.

As suggested by Figs. 6 and 7, the ratio f_o/f_r does not depend on Young's modulus and on the geometry of the disk, but depends on Poisson's ratio. In order to clarify this point, Fig. 8 represents the variation of the ratio f_o/f_r versus Poisson's ratio in three cases. In the two first cases, the geometry is not modified, R=31.5 mm, while the material corresponds to material 1 (ρ=1150 kg/m^3, E'=2.98 GPa) or material 2 (ρ=2700 kg/m^3, E'=69.0 GPa). In the third case, the disk is made of material 1 and has a radius equal to 50.0 mm. The

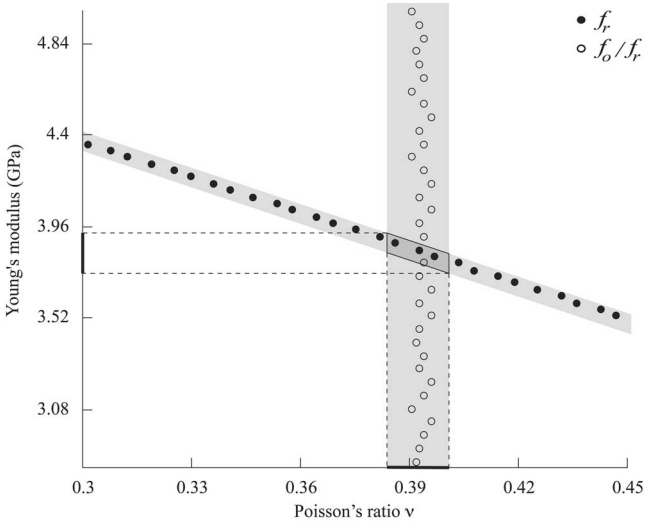


FIG. 7. Solution area giving f_r=15.7 kHz and f_o/f_r=0.73 for the disk Di_2.

TABLE IV. Estimated values of Young's modulus and Poisson's ratio of the disks' material.

Disk	Di_1	Di_2
Young's modulus E (GPA)	3.00±0.05	3.90±0.10
Poission's ratio v	0.40±0.01	0.39±0.01

superposition of the three curves shows that the ratio f_o/f_r does not depend on the density, Young's modulus, and the geometry of disks. In fact, the ratio f_o/f_r depends only on Poisson's ratio. In conclusion, one can use the curve of Fig. 8 to estimate directly Poisson's ratio of an unknown isotropic material starting from its f_o and f_r eigenfrequencies independently from the disk geometry and the material stiffness. Considering the complexity of the model equations, this independence has not been explained theoretically. However, the procedure described in Figs. 6 and 7 permits us to estimate simultaneously Young's modulus and Poisson's ratio of an unknown material with good accuracy.

B. Geometry evaluation

Figure 9 shows the variations of the first three radial eigenfrequencies of a ring versus its inside radius. These eigenfrequencies are calculated with Eq. (24), the outside radius being 40 mm. Figure 9 reveals that the first eigenfrequency decreases when the inside radius increases. Inversely, the second and third frequencies increase with respect to the inside radius. The decrease of the first resonance frequency can be related to the global stiffness of the ring, while the increasing of the second and third eigenfrequencies is due to the decrease of the ring thickness.

From these observations, a procedure for estimating the geometrical characteristics of rings is suggested. Figure 10 represents all the pairs, among two sets of inside and outside radii, which produce the first two experimental eigenfrequen-

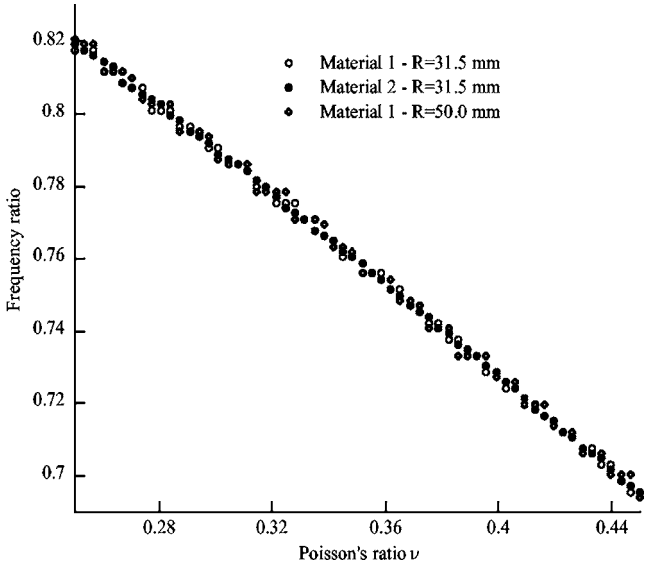


FIG. 8. Comparison of the variations of the frequencies ratio f_o/f_r versus the Poisson's ratio values for different materials and geometries.

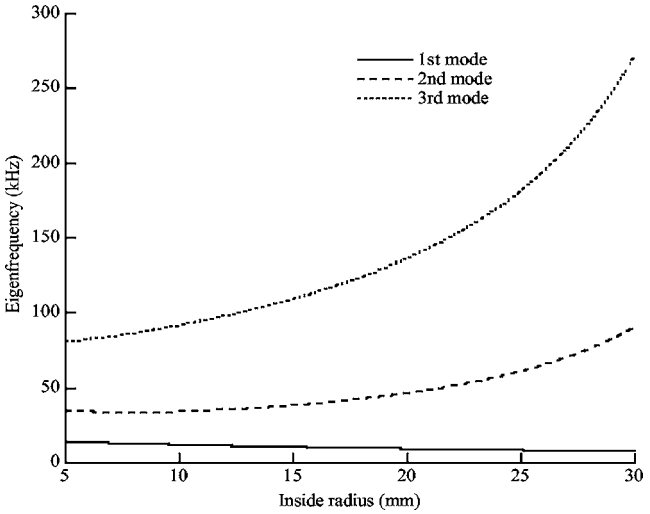


FIG. 9. Variations of the first three eigenfrequencies of radial modes in a ring versus the inside radius.

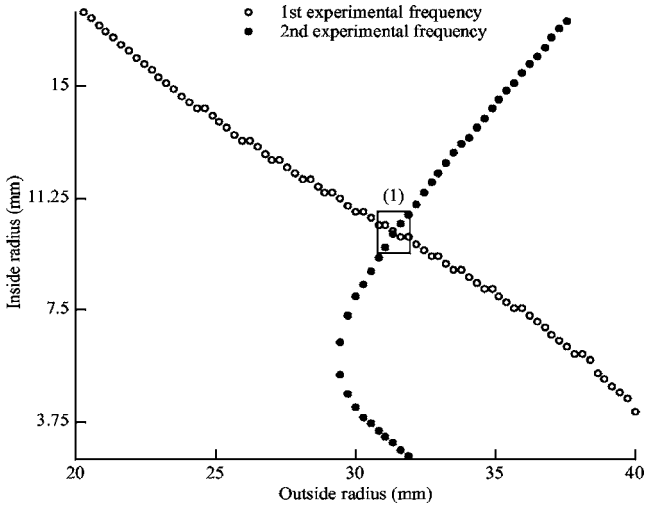


FIG. 10. Solutions for outside and inside radii giving $f_1=14.4$ kHz and $f_2=46.2$ kHz, Ri_1: $R_1=10.0$ mm, $R_2=31.5$ mm.

cies ($f_1=14.4$ kHz and $f_2=46.2$ kHz) of a ring. The intersection point 1 furnishes the actual researched dimensions of the ring Ri_1 ($R_2=31.5$ mm and $R_1=10.0$ mm).

Thus, the knowledge of the two first eigenfrequencies of the radial mode allows us to determine the inside and outside radii of a ring with good accuracy. The use of these two radial eigenfrequencies has the advantage of making the inversion process fast since it uses the simple model (24) to evaluate the eigenfrequencies.

VI. CONCLUSIONS

The free vibration of disks and rings made of an isotropic and viscoelastic material excited by pulsed microwave irradiations has been studied. Depending on the spatial distribution of microwave fields, a semi-analytical model has been presented for estimating the eigenfrequencies of disk and rings and for calculating their displacement fields. The comparison between the theoretical results and the experimental ones has permitted the validation of the model. The simultaneous use of this model and the microwaves' excitation gives access to a rich set of theoretical and experimental eigenfrequencies, in order to perform modal analysis of rings and disks.

It has been shown that this information can serve to solve geometrical and material inverse problems in disks and rings. So the microwave excitation becomes a nondestructive and contactless method for the evaluation of these structures.

For that, the choice of some eigenfrequencies, corresponding to radial and nonradial modes, is crucial to reach a unique and precise solution. The ability to only excite the purely radial eigenmode by the microwave technique is very important in order to identify these modes. Otherwise, it has been observed that the use of a specific frequency ratio is independent of the disk geometry and the material stiffness. The theoretical explanation of this observation is open to investigations. In the future, the full set of frequencies could be exploited for the estimation of more parameters, like in anisotropic structures, for instance.

The model and the experimental procedure presented in this paper for plane cylindrical structures could be adapted to other geometries. As mentioned in the paper, the adaptation to cylinders could be straightforward.

ACKNOWLEDGMENT

The authors are very grateful to Dr. M. Deschamps for his very helpful advice in achieving the present work.

[1]L. Pochhammer, "Ueber die fortpflanzungsgeschwindigkeiten kleiner schwingungen in einem unbegrenzten isotropen kreiscylinder," J. Reine Angew. Math. **81**, 324–336 (1876).
[2]C. Chree, "Longitudinal vibrations of a circular bar," Q. J. Math.**21**, 287–298 (1886).
[3]H. Lamb, Proc. R. Soc., London **98**, 205 (1921).
[4]H. Carrington, "The frequencies of vibration of flat circular plates fixed at the circumference," Philos. Mag. **50**, 1261–1264 (1925).
[5]Lord Rayleigh, *Theory of Sound*, Vol. **I** (Dover, New York, 1945).
[6]T. Wah, "Vibration of circular plates," J. Acoust. Soc. Am. **34**(3), 275–281 (1962).
[7]G. Martincek, "The determination of poisson's ratio and the dynamic modulus of elasticity from the frequencies of natural vibration in thick circular plates," J. Sound Vib. **2**(2), 116–127 (1965).
[8]S. N. Rasband, "Resonant vibrations of free cylinders and disks," J. Acoust. Soc. Am. **57**(4), 899–905 (1975).
[9]H. Lee and R. Singh, "Acoustic radiation from out-of-plane modes of an annular disk using thin and thick plate theories," J. Sound Vib. **282**(1-2), 313–339 (2005).
[10]G. Ambati, J. F. W. Bell, and J. C. K. Sharp, "In-plane vibrations of annular rings," J. Sound Vib. **47**(3), 415–432 (1976).
[11]S. S. H. Chen and T. M. Liu, "Extensional vibration of thin plates of various shapes," J. Acoust. Soc. Am. **58**, 828–831 (1975).
[12]N. H. Farag and J. Pan, "Modal characteristics of in-plane vibration of circular plates clamped at the outer edge," J. Acoust. Soc. Am. **113**, 1935–1946 (2003).
[13]E. A. G. Shaw, "On the resonant vibrations of thick barium titanate disks," J. Acoust. Soc. Am. **28**, 38–50 (1956).
[14]D. S. Moseley, "Contribution to the theory of radial extensional vibrations in thin disks," J. Acoust. Soc. Am. **32**, 991–995 (1960).
[15]B. Hosten and P. Alain Bernard, "Ultrasonic wave generation by time gated microwaves," J. Acoust. Soc. Am. **70**, 1577–1585 (1998).
[16]C. Bacon, B. Hosten, and P. A. Bernard, "Acoustic wave generation in viscoelastic rods by time-gated microwaves," J. Acoust. Soc. Am. **106**, 195–201 (1999).
[17]C. Bacon, B. Hosten, and E. Guilliorit, "One dimensional prediction of the acoustic waves generated in a multilayer viscoelastic body by microwave irradiation," J. Sound Vib. **238**(5), 853–867 (2000).

[18]C. Bacon, E. Guilliorit, B. Hosten, and D. Chimenti, "Acoustic waves generated by pulsed microwaves in viscoelastic rods: Modeling and experimental verification," J. Acoust. Soc. Am. **110**(3), 1398–1407 (2001).
[19]B. Hosten and C. Bacon, "Measurement of complex Young moduli of composite materials by time gated microwaves," Rev. Prog. Quant. Nondestr. Eval. **19**(1), 1113–1120 (2000).
[20]E. Guilliorit, C. Bacon, and B. Hosten, "Prediction of the generation of acoustic waves due to the penetration of pulsed microwaves in multilayer media," J. Acoust. Soc. Am. **112**(1), 65–74 (2002).
[21]E. Guilliorit, B. Hosten, C. Bacon, and D. E. Chimenti, "Microwave excitation of ultrasound in graphite-fiber reinforced composite plates," Ultrasonics **41**, 97–103 (2003).
[22]C. Bacon, E. Guilliorit, and B. Hosten, "An acoustic-microwave method for the study of the mechanical influence of moisture content in materials," J. Appl. Mech. **70**, 268–274 (2003).
[23]A. R. Hadj Henni, C. Bacon, and B. Hosten, "Acoustic generation in piezoelectric materials by microwave excitation," J. Acoust. Soc. Am. **118**(4), 2281–2288 (2005).
[24]J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).
[25]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vols. **1** and **2**.
[26]H. Batard and G. Quentin, "Acoustical resonances of solid elastic cylinders: Parametric study and introduction to the inverse problem," J. Acoust. Soc. Am. **91**, 581–590 (1992).
[27]P. Heyliger, A. Jilani, H. Ledbetter, R. G. Leisure, and C.-L. Wang, "Elastic constants of isotropic cylinders using resonant ultrasound," J. Acoust. Soc. Am. **94**, 1482–1487 (1993).
[28]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions—with formulas, graphs and mathematical tables* (Dover, New York, 1965).
[29]M. R. Spiegel, *Mathematical Handbook of Formulas and Tables* (McGraw-Hill, New York, 1974).

Vibrational modes of partly filled wine glasses

Gregor Jundt, Adrian Radu, Emmanuel Fort, Jan Duda, and Holger Vach[a)]
Laboratoire de Physique des Interfaces et des Couches Minces, CNRS UMR-7647 - Ecole Polytechnique, 91128 Palaiseau, France

Neville Fletcher
Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

(Received 11 August 2005; revised 4 March 2006; accepted 29 March 2006)

Time-average holographic interferometry has been employed to study how the vibrational modes of a singing wine glass change when it is filled with a liquid. While the liquid clearly lowers the resonance frequencies, it does not change the vibrational mode structure in a first approximation. A more detailed analysis, however, reveals that the presence of the liquid causes the simultaneous excitation of two orthogonal modes that are well resolved for the empty glass. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2198183]

PACS number(s): 43.40.Ey, 43.75.Kk, 43.20.Tb, 43.40.At [LPF] Pages: 3793–3798

I. INTRODUCTION

It is a common game. You take a wine glass and you rub your moistened finger around its rim. Besides enjoying a generally rather pure tone emitted by your singing wine glass, you might even take pleasure in observing some ripples on the liquid surface that follow your rotating finger. On a less playful level, many famous composers such as Mozart, Berlioz, or Saint-Saëns have written master music pieces for instruments based on glass vibrations, also called "musical glasses." The ethereal sound they produce has been much appreciated for many centuries. These instruments can be divided into two main groups depending on the way the vibrations are produced.[1] The first group is composed of percussion instruments like bells, cymbals, and balaphones made of glass. The second group consists of bowed instruments, like the glass harmonica and seraphim, that are excited by a "stick and slip" technique which is also at the origin of the sound production for our singing wine glass. The pitch can simply be tuned by pouring some liquid in the glass. This technique is, for instance, successfully used to tune seraph instruments. Another way to change the pitch is to use other glasses with a different thickness or diameter; glass harmonicas are commonly based on this second technique. The growing interest in glass musical instruments has already led to several papers on the acoustics of wine glasses.[2–5] In this paper, we present experimental results obtained with simple wine glasses filled to various levels. We first show how the glass resonance frequencies vary with liquid quantity. Using time-average holographic interferometry, we then investigate the question whether the vibrational mode structure also becomes influenced by the presence of the liquid. We demonstrate that the latter one actually remains unchanged in a first approximation. A more detailed analysis, however, reveals that the liquid introduces an overlap of the two, previously well resolved, orthogonal mode frequencies leading to their simultaneous excitation.

In the next section, we describe our experimental setup. In Section III, we present a simple analytical model to explain the pitch lowering with increasing liquid quantity. Section IV is devoted to our experimental results and discussion.

II. EXPERIMENTAL SETUP

The experimental setup for holographic interferometry is shown in Fig. 1. The optical configuration we use for holographic recording is a modified Mach-Zehnder configuration. The laser beam originating from a 7-mW multi-mode He-Ne laser is split into two beams with equal intensities. The reference beam is filtered and expanded through a spatial filter consisting of a $\times 40$ microscope lens and a 10 μm pin hole. The object beam is expanded by a diffuse glass and its length is chosen to equal the optical path length of the reference beam at the holographic plate in order to guarantee maximal temporal coherence. The position of the diffuser is chosen in such a way that the intensity of the reference beam is about three times higher than that of the object beam at the place of the holographic plate. In order to minimize vibrations the whole optical part is set on a properly isolated optical table. We determined the optimal exposure time to be about 50 s for our ABSYS BB-640 holographic plates. More experimental details concerning the holographic setup can be found in Refs. 6 and 7.

We placed a microphone approximately 2 cm below the glass rim and at a distance of about 2 mm from the glass. This microphone was then connected to a computer to record frequency spectra and to detect the resonance frequencies of the glass. The impulse response of the glass-liquid system was obtained by hitting the glass gently with a metal rod. The selective excitation of a particular vibrating mode was obtained by using a stabilized low frequency generator connected to the input of a 50-W amplifier with a loudspeaker. The accurate resonance frequency was obtained by comparing the amplitudes of the acoustic signal obtained for a

[a)]Electronic mail: vach@leonardo.polytechnique.fr

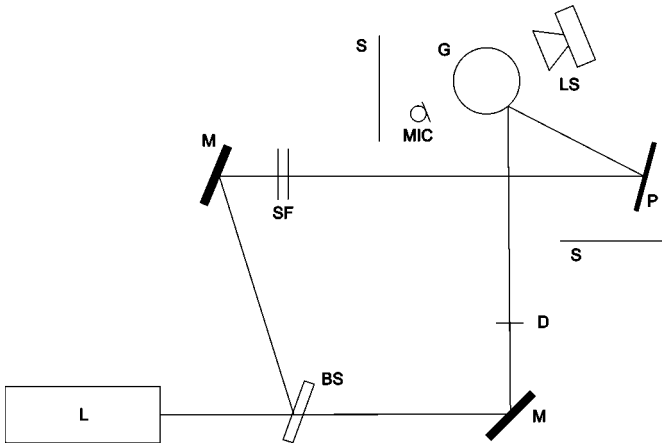


FIG. 1. Experimental setup for time-integrated holographic interferometry. L: laser, BS: beam splitter, M: mirror, SF: spatial filter, D: diffuse glass, MIC: microphone, LS: loudspeaker, P: holographic plate, G: wine glass, and S: beam stop against stray light.

clamped and an unclamped glass. Subtracting the acoustic response of the glass and the loudspeaker system from the one produced by the loudspeaker alone excited by a modulated sinusoidal signal gives an alternative method to obtain the acoustic response of the glass-liquid system. Both methods result in identical experimental resonance frequencies. Finally, the excitation of the quasi-degenerate (2,0) quadrupole modes[4] can also be produced by the simple finger rubbing technique. In this case, however, the vibrational structure of the glass follows the exciting finger. Consequently, it becomes impossible to use time-average holographic interferometry with this excitation scheme. We chose, therefore, the excitation method with the metal rod only for the impulse response measurements while loudspeaker excitation was used for all the other experiments.

III. THEORY

A rigorous treatment of the vibration of liquid-loaded wine glasses is a complex matter for several reasons. Even for an empty glass, the profile is not simple and is different for every glass type: red-wine glasses, chardonnay glasses, champagne glasses, etc. In addition, the wall thickness varies progressively from relatively thick at the base to thin at the top edges. A solution for mode shape and frequency could, of course, be obtained for an arbitrarily shaped partly filled glass by using finite-element numerical analysis, but this would lead to specific rather than general understanding. The aim of the present analysis, in contrast, is to provide a simple treatment in which the underlying physics is made clear.

A. Fundamental mode

The shape of the wine glass under study is intermediate between that of a paraboloid of revolution clamped at its peripheral center, a cone clamped at its vertex, and a cylinder clamped around its base. Even the solution of these idealized cases is very complex,[8] but our initial concern is only with the fundamental mode which is a simple elliptical deformation of the glass cross section, so that general approximations based on very simple considerations[9] can be made. Suppose that the tangential coordinate, measured from the base of the cup, is s, that the azimuthal coordinate is ϕ, and that the total length of the cup surface is L. The fundamental mode can then be written as

$$\psi(s,\phi,t) = f(s)\cos 2\phi \sin \omega t, \tag{1}$$

where ω is the angular frequency of the vibration and $f(s)$ is an unknown function of the coordinate s. To determine an approximation to this function we proceed as follows. The clamping constraint at the base requires that

$$f(0) = 0, \qquad \left.\frac{\partial f}{\partial s}\right|_{s=0} = 0, \tag{2}$$

while the fact that the rim is free requires that

$$\left.\frac{\partial^2 f}{\partial s^2}\right|_{s=L} = 0, \qquad \left.\frac{\partial^3 f}{\partial s^3}\right|_{s=L} = 0. \tag{3}$$

Since we are concerned here with the fundamental mode, which has no nodal rings, it is a good approximation to expand $f(s)$ as a power series and to retain only the first term, so that

$$f(s) \approx \frac{s^2}{L^2}, \tag{4}$$

which satisfies the condition (2) though not (3). Because of the other approximations necessarily involved in the analysis to follow, however, this will be regarded as adequate.

It is now necessary to convert the tangential coordinate s into a vertical coordinate z, and this depends in detail upon the shape of the glass. If the glass profile is cylindrical, then $z=s$ and there is no problem. Similarly for a conical glass $z=As\cos\theta$ where θ is the cone semi-angle and A is a constant. For a glass with parabolic profile, which is actually a closer approximation to the vessel studied, $z=As^2$ close to the base where the wall slope is small and more nearly $z=As$ higher up the glass where the slope is large. Since the real wine glass profile is somewhere between these extremes, and the glass thickness is not constant, an adequate approximation for the present purpose is to write the mode function $f(s)$ of (1) as z^β so that (1) becomes

$$\psi(z,\phi,t) \approx z^\beta \cos 2\phi \sin \omega t. \tag{5}$$

This approximate result can be compared with experiment by examining the positions of the fringes on the holograms of the vibrating glass surface and correcting for the local slope of the surface in the observation direction. This comparison is shown in Fig. 5, which will be discussed in detail in Sec. IV. The data are taken from two holograms made of a glass vibrating in the first and second modes, respectively. β does not depend on the scaling of the glass height. The plot indicates a best-fit value of $\beta \approx 2.1$ for the fundamental mode and $\beta \approx 3.6$ for the hexapolar (3,0) mode, which has three nodal diameters and no nodal circles.

In the next stage of the analysis it is assumed that addition of liquid to the glass does not have a significant effect upon the vibrational mode shape. This is the standard assumption of first-order perturbation theory. In a more accurate second-order perturbation treatment, the effect of the

liquid loading upon the mode shape would be included.[10] This refinement is not attempted in the present case since the treatment already involves major approximations, so that undue mathematical refinement is not appropriate. It is, however, discussed briefly later.

In any simple harmonic vibration, the mean kinetic energy and elastic strain energy are equal. If it is assumed as a first approximation that the vibrational mode shape of the fundamental mode is independent of the liquid loading, then the elastic strain energy for a given amplitude can be assumed to have the constant value E and the kinetic energy of vibration of the glass walls the value $K\omega^2$, where K is a constant and ω is the vibration frequency. The kinetic energy of the contained liquid can be written $W\omega^2$, where W depends upon the depth h of the liquid in the glass and is proportional to the density of the liquid. From this it then follows that

$$\omega^2 = \frac{E}{K + W(h)}. \tag{6}$$

The form of the kinetic energy function $W(z)$ of the liquid depends upon the glass shape and the shape of the vibrational mode, both of which are complicated. To a reasonable approximation for the fundamental mode, however, this mode shape can be assumed to follow the mode shape of the glass walls as given by (5) with an additional two-dimensional internal flow of proportional amplitude. The mass of co-moving liquid at a given height z is proportional to the cross section $S(z)$ of the glass at that height, and so depends upon glass shape. The general expression for $W(h)$ is then

$$W(h) \propto \int_0^h S(z) z^{2\beta}\, dz \tag{7}$$

and we must evaluate the integral for each glass shape. For a cylinder, $S(z)$ is a constant independent of z, for a paraboloid of revolution $S(z) \propto z$, and for a cone $S(z) \propto z^2$, so that the integral in (7) is easily evaluated and has the form Bh^n where B is a constant depending upon the glass dimensions and $n = 2\beta+1$ for a cylinder, $n=2\beta+2$ for a hemisphere, and $n = 2\beta+3$ for a cone. Substituting this into (6) gives the result

$$\omega^2 = \frac{\omega_0^2}{1 + \alpha h^n}, \tag{8}$$

where ω_0 is the mode frequency for the empty glass, α is a constant proportional to the liquid density and also depending upon glass shape and wall thickness, and n is determined as discussed above.

Since the shape of the wine glass studied is complex, the theory can only predict approximate values for the parameter n. For a cylindrical glass $3 \leqslant n \leqslant 5$, for a paraboloidal glass $4 \leqslant n \leqslant 6$, and for a conical glass $5 \leqslant n \leqslant 7$, with the upper limit in each case being most nearly appropriate for a glass of uniform wall thickness. Since the shape most closely conforming to that of the experimental glass is paraboloidal, we expect to find that $n \approx 6$ gives the best fit to the experimental data, and indeed a value of $n=5.5$, as shown in Fig. 2, does give an excellent fit.

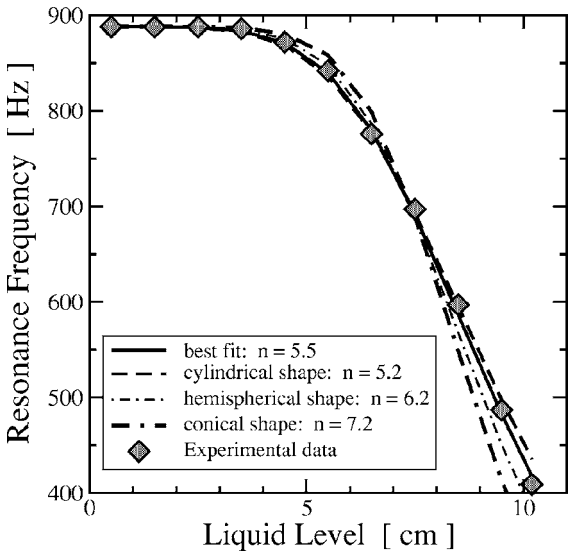


FIG. 2. Experimentally measured eigenfrequencies for the quadrupolar mode (2,0) as a function of water height. The dashed gray line results from the model for a cylindrical shaped shell while the black dotted line is the best-fit spherical curve and the gray continuous line the best-fit spherical-cap curve. The black continuous line shows the least-mean-square fit.

B. Higher modes

The analysis above considers only the fundamental mode which has two nodal diameters and no nodal circles. It is normal to refer to this as a (2,0) mode. A higher mode (p,q) with p nodal diameters and q nodal circles has the general form

$$\psi_{p,q}(s,\phi,t) = f_{pq}(s) \cos p\phi \sin \omega_{p,q} t, \tag{9}$$

and f_{pq} will generally increase about as s^p near the origin and then become oscillatory for larger values of s if $q \neq 0$. Such oscillations associated with the presence of nodal rings greatly complicate the analysis and experiment. If, however, there are no nodal rings, then the form of the mode function in (9) suggests that $\beta \approx p$ so that, since n in (8) lies between $2\beta+1$ and $2\beta+3$ and is actually about 5.5 from the experimental results for the (2,0) mode, we should expect values of about 7.5 and 9.5, respectively, for (3,0) and (4,0) modes.

There is another complication, particularly if many nodal diameters are involved, and this is that vibrations in the liquid will be increasingly localized near to the glass surface, so that the moving-mass term will increase more nearly as the glass perimeter rather than as its area. This has not been taken into account in the simplified analysis above.

C. Limitations of the present theory

Despite the success of this analysis in explaining the experimental results, it is useful to mention briefly the limitations of the predictions. In the limit of a fluid of infinite density, the second-order perturbation leading to changes in mode shape cannot be neglected, since the walls of the glass would be effectively clamped below the surface of the liquid. For the case of a partly filled glass this would lead to an increase in mode frequency with increasing liquid depth rather than a decrease. For this reversal of behavior to occur, it is necessary that the mass of the vibrating fraction of the

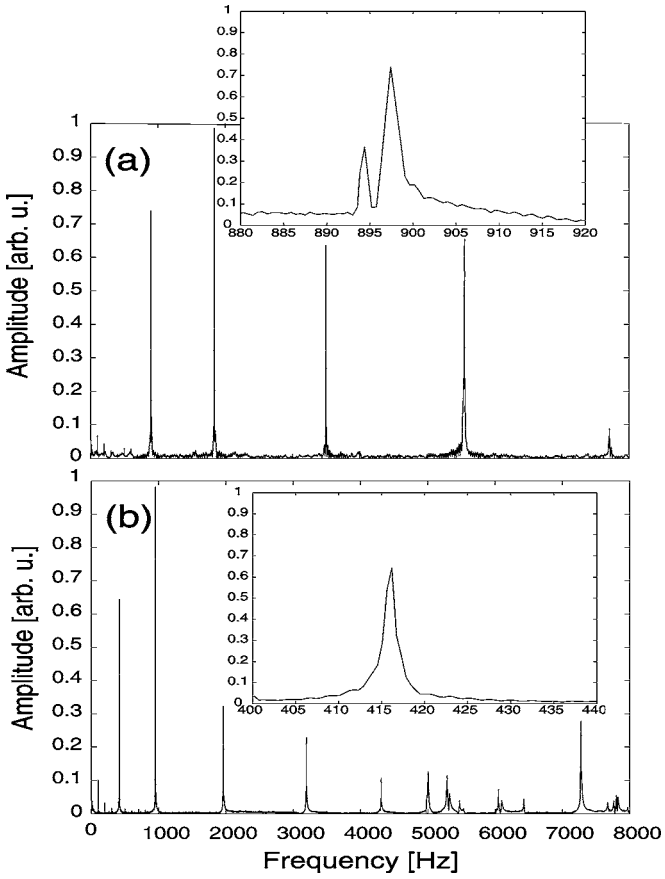


FIG. 3. Amplitude spectrum of the impulse response of an empty (a) and a full (b) glass. Insets: the associated orthogonal (2,0) modes.

liquid be very large compared with that of the vibrating walls. The determination of the threshold density above which this process reversal will take place certainly presents a challenging task for both experiments and theory, but clearly goes beyond the scope of our present work.

IV. RESULTS AND DISCUSSION

Figure 3 shows the typical spectra associated with an impulse response for the empty glass (a) and a glass filled with water (b). The insets show the spectra of the two orthogonal quadrupolar modes for each system. For the empty glass, only a few resonance frequencies are visible in the displayed spectrum. They can be attributed to the flexural $(m,0)$ modes with $m=2$ to 6.[3] For a filled glass, the resonances are shifted toward lower frequencies. This is also true for higher modes. Therefore more resonance peaks appear within the frequency window of Fig. 3.

In the inset (a), the frequencies of the two orthogonal quadrupolar modes are clearly resolved for the empty glass, although the splitting is only about 4 Hz. This splitting is due to small imperfections in the glass structure (variation of thickness, presence of impurities, etc). In the case of a full glass [see inset of Fig. 3(b)], these two modes become degenerate. The reasons for this interesting result will be discussed in more detail below.

In Fig. 2, we present the measured resonance frequency for the (2,0) quadrupolar mode versus the water height in the

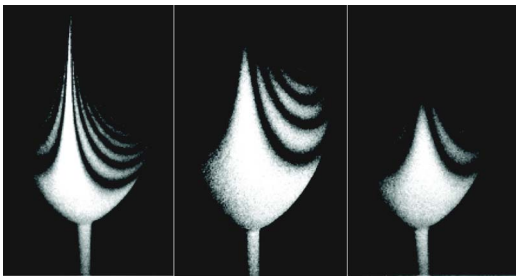


FIG. 4. Photographs of the holographic images of the vibrating wine glass: (a) empty, (b) half-filled with water, and (c) full. In all cases, the glass was excited by a loud speaker at the eigenfrequency of its quadrupolar (2,0) mode.

glass. While a careful parameter fitting might begin with $n \approx 6$ near the base and progress through $n \approx 5$ at medium heights to $n \approx 4$ near the top of the glass, such a refinement is not really justified, and we find a fit with $n \approx 5.5$ as being the most appropriate.

Figure 4 shows the photographs of the holographic images of the vibrating glass for three different filling levels: (a) empty, (b) half-full, and (c) full. The glass is always excited with the loudspeaker at the eigenfrequency of its quadrupolar $(2,0)$ mode determined as described in Sec. II. The brightest fringe corresponds to the nodal line of the eigenmode in each hologram. Skeldon *et al.* showed glasses being intentionally broken by excitation (Fig. 10 in Ref. 4). The crack route most probably corresponds to an antinodal line where the bending stress is greatest. This line is displaced 90° from the nodal line for the case of the $(2,0)$ vibration visible in our holograms.

From the number of fringes readily counted in Fig. 4, it is possible to deduce the vibrational amplitudes. Let the point M of the glass vibrate sinusoidally with an amplitude A and a pulsation ω in a direction toward the holographic plate. Its displacement is thus given by $x(M,t)=A(M)\sin(\omega t)$.

Since the exposure time $T \gg 1/\omega$, the intensity $I(M)$ of the reconstructed hologram is given by

$$I(M) \propto I_S(M) J_0^2\left(\frac{4\pi}{\lambda} A(M)\right) \tag{10}$$

where $I_S(M)$ is the intensity of the reconstructed hologram scattered by the point M when it is not vibrating, J_0 is the Bessel function of the first kind of order zero, and λ is the light wavelength.[6] Dark fringes are consequently centered at each point on the object surface where the Bessel function becomes zero for a given vibrational amplitude $A(M)$.

Figure 5 shows the vibrational amplitude normal to the surface of the empty glass f versus the height z. The data are calculated from the holograms taking into account the geometry of the incident beam and the observation direction. This analysis gives a regression of $z=Af^{0.47}$ for the fundamental mode. This implies that $\beta \approx 2.1$ which gives a value of n between 5 and 7, these extreme values being those for a cylinder and for a cone, respectively. Examining the shape of the glass used in the experiment, it is clear that, except for very small liquid levels such that $h<0.3H$, where H is the total height of the glass cup, the glass profile is much more nearly cylindrical than conical, so that it is to be expected

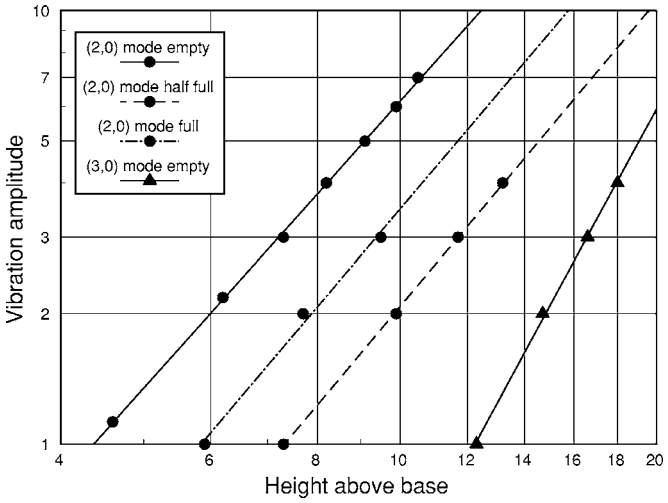


FIG. 5. Vibration amplitude $\psi(z)$ in units of $\lambda/2$, as measured by the progression of the bright fringes corrected for wall angle, plotted against height z above the base (in arbitrary units as measured from an enlarged photograph) for the fundamental (2,0) mode with different liquid levels and for the (3,0) mode. The straight lines are power-law regressions $\psi = Az^{\beta}$ as in Eq. (5). Note that the slope of all the (2,0) mode regressions is the same, indicating that addition of the liquid does not significantly modify the mode shape. Differing heights for corresponding bands for the (2,0) mode arise from different excitation efficiencies.

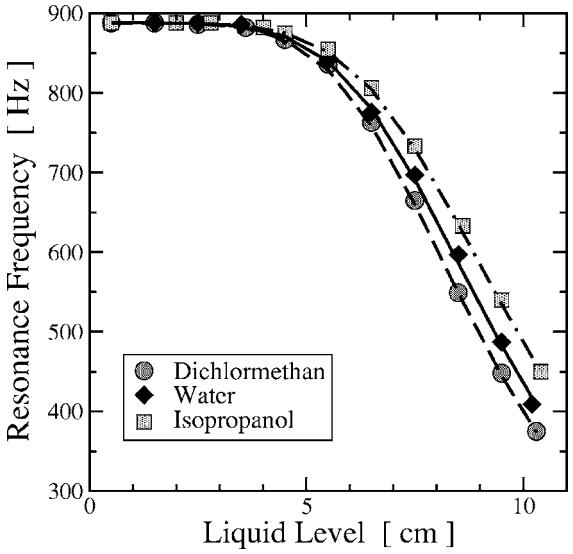


FIG. 6. The plot shows the resonances of the glass filled with dichloromethane, water, and isopropanol, respectively. The density of the liquid in use is proportional to the parameter α.

that the best-fit value of n will be closer to 5 than to 7. As shown in Fig. 2, the best-fit value derived from the microphone experiments is in fact about $n \approx 5.5$. The same analysis has been done for the (3,0) mode which leads to $\beta \approx 3.6$ and therefore $8.2 \leqslant n \leqslant 10.2$. This is in good agreement with the value $n \approx 8$ taken from the respective hologram.

About 25 dark fringes are visible in Fig. 4(a) which gives an approximate amplitude of $A_n \approx 4\ \mu$m; for Figs. 4(b) and 4(c), $A_n \approx 1\ \mu$m (seven dark fringes) considering the employed recording geometry. The amplitude clearly decreases with the liquid level, presumably due to the increased inertia and the less efficient coupling of the loudspeaker to the glass as the frequency decreases.

The parameter α in Eq. (8) is proportional to the density of the liquid filled into the glass. In order to check formula (8) we measured the resonance frequencies for liquids with different densities. Figure 6 shows the data for liquids with three different densities, dichloromethane (1.32 kg/l), water (1 kg/l), and isopropanol (0.79 kg/l). From a fit according to formula (8) we obtain the ratios $\alpha_{\mathrm{dichl}} : \alpha_{\mathrm{water}} : \alpha_{\mathrm{iso}} = 1 : 0.80 : 0.59$, which is in good agreement with the real density ratios of $1 : 0.75 : 0.60$.

We can directly conclude from Fig. 4 that the presence of the liquid does not markedly modify the vibrational shape of the quadrupolar eigenmodes. It is, for instance, impossible to determine from the photographs up to which level the glass was filled. A plot of the positions of the bright fringes in the patterns of (b) and (c) of Fig. 4 for a partly and a completely filled glass using the logarithmic scale of Fig. 5 in fact gives straight lines closely parallel to that drawn for the $(2,0)$ mode of an empty glass, showing that the assumption of little change in the vibrational mode function is justified. The main difference arising from the liquid is the fact that the nodal line no longer reaches the rim of the glass. In spite of the relatively high vibrational amplitude for the empty glass, the bright white nodal line clearly reaches the rim of the glass, while in the case of the half-full glass the upper part appears entirely dark. In the case of a completely filled glass, this phenomenon is even more pronounced. We attribute this effect to the fact that the coupling of the two degenerate modes should lead to an energy exchange between them. The resulting modulation of the vibrational modes, which is expected to have a frequency of a few Hz, however, darkens the image of the glass completely over the integration time of 50 s if the modulation amplitude is larger than $\lambda/2$.

Splitting of the resonances in degenerate modes, as shown in Fig. 3(a), is almost certainly due to asymmetry in the thickness of the glass so that the sine and cosine versions of Eq. (1) have slightly different frequencies. When liquid is added to the glass, this reduces the asymmetry because there is an equal liquid mass added to all parts of the wall. Consequently, the additional mass appears to shift the two resonance peaks to a point where the two nearly degenerate modes cannot be resolved anymore [compare Fig. 3(b)], leading to their simultaneous excitation. This phenomenon can also be nicely demonstrated by the use of the rotating finger excitation: The amplitude of the signal emitted by the empty glass as recorded with our microphone is strongly modulated since the excitation follows the finger. As the node passes in front of the microphone the amplitude becomes nearly zero when a sufficiently directional microphone is employed. When filled with water, however, the amplitude never vanishes entirely. Tentatively, we attribute this change in modulation amplitude to the simultaneous excitation of the two orthogonal quadrupolar modes which becomes possible due to the presence of the liquid. As a consequence, the coupled liquid-glass system clearly behaves differently from an empty glass. A similar behavior has already been reported for cylindrical steel tanks filled with water.[11–13]

V. CONCLUSION

In conclusion, we showed that the presence of a liquid does not change the vibrational structure of a singing wine glass in a first approximation in spite of the obvious lowering of the resonance frequency, i.e., the vibrational movement of the glass continues nearly undisturbed below the level of the liquid. A more detailed analysis of our experimental results reveals, however, that the nodal line becomes modified near the glass rim: using time-average holographic interferometry, we showed how two well-resolved orthogonal quadrupolar modes of an empty glass become degenerate due to the presence of the liquid causing their simultaneous excitation.

ACKNOWLEDGMENTS

We would like to address our special thanks to Julien Masson and Professor Xavier Boutillon for fruitful discussions. We are grateful to the "Centre de Travaux Experimentaux de Physique - Ecole Polytechnique" where the holograms were recorded.

[1]T. D. Rossing, "Wine glasses, bell modes, and Lord Rayleigh," Phys. Teach. **28**, 582–585 (1990).

[2]A. P. French, "In Vino Veritas: A study of wineglass acoustics," Am. J. Phys. **51**, 688–694 (1983).

[3]T. D. Rossing, "Acoustics of the glass harmonica," J. Acoust. Soc. Am. **95**, 1106–1111 (1994).

[4]K. D. Skeldon, V. J. Nadeau, and C. Adams, "The resonant excitation of a wineglass using positive feedback with optical sensing," Am. J. Phys. **66**, 851–860 (1998).

[5]A. Chaigne, "Le chant des verres de vin," Sciences et Avenir **100**, 52–57 (1995).

[6]O. Haeberlé, B. Sapoval, K. Menou, and H. Vach, "Observation of vibrational modes of irregular drums," Appl. Phys. Lett. **73**, 3357–3359 (1998).

[7]F. Pinard, B. Laine, and H. Vach, "Musical quality assessment of clarinet reeds using optical holography," J. Acoust. Soc. Am. **113**, 1736–1742 (2003).

[8]W. Soedel, *Vibrations of Shells and Plates* (Marcel Dekker, New York, 1993).

[9]T. D. Rossing and N. H. Fletcher, *Principles of Vibration and Sound*, 2nd ed. (Springer-Verlag, New York, 2004).

[10]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. **II**, Chap. 9.

[11]P. G. Bentley and D. Firth, "Acoustically excited vibrations in a liquid-filled cylindrical tank," J. Sound Vib. **19**, 179–191 (1971).

[12]C. R. Fuller and F. J. Fahy, "Characteristics of wave propagation and energy distributions in cylindrical elastic shells filled with fluid," J. Sound Vib. **81**, 501–518 (1982).

[13]M. Amabili, "Vibrations of circular tubes and shells filled and partially immersed in dense fluids," J. Sound Vib. **221**, 567–585 (1999).

Prediction method for tire air-pumping noise using a hybrid technique

Sungtae Kim[a)]
School of Mechanical and Aerospace Engineering, Seoul National University, Building 301-1214, Seoul 151-742, Korea

Wontae Jeong[b)]
Applied Technology Research Department, Hyundai MOBIS, 80-10, Mabook-Dong, Giheung-Gu, Yongin-Shi Gyunggi-Do 449-910, Korea

Yonghwan Park[c)]
School of Mechanical and Aerospace Engineering, Seoul National University, Building 301-1214, Seoul 151-742, Korea

Soogab Lee[d)]
Institute of Advanced Aerospace Technology, School of Mechanical and Aerospace Engineering, Seoul National University, Building 301-1303, Shilim-Dong, Gwanak-Gu, Seoul 151-742, Korea

(Received 22 September 2005; revised 8 February 2006; accepted 5 April 2006)

Air-pumping noise from a car tire is investigated with a hybrid technique composed of three stages: (1) small-scale air-pumping noise generation process is modeled as a piston-like movement of the base-side of the tire groove and then numerically simulated; (2) the flow properties in the tire groove are used as air-pumping sources and noise propagation is simulated with emphasis placed on scattering process with full tire/road geometry; (3) the far-field acoustic pressure is predicted from a Kirchhoff integral method by using unsteady flow data in space and time which is provided by the computational fluid dynamics (CFD) calculation of full tire-road domain. The comparison of predicted results shows that the nonlinearity of the air-pumping noise generation mechanism affects not only noise characteristics in frequency domain but also in the directivity pattern. It seems that this approach can overcome the weakness of the acoustic monopole theory which stems from the usual assumption of a small amplitude acoustic wave equation while using nonlinear governing equation for the CFD calculation. Furthermore, through the use of a computational domain which covers tire and road surface, the geometric effects on air-pumping noise generation and propagation are taken into account in the source modeling process. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200140]

PACS number(s): 43.50.Lj, 43.50.Ed, 43.55.Ka, 43.28.Ra [DKW] Pages: 3799–3812

I. INTRODUCTION

In general, tire/road noise generation mechanisms can be divided into vibration-related and aerodynamically-related groups according to the media in which noise occurs and its effects.[1] The examples of the noise of vibration-related mechanism are tread impact noise, texture impact noise and stick/slip or stick/snap noise which result from the impact of tire tread blocks or pattern elements on the road surface, the stroke of road surface texture on the tire tread, and the tread element motion relative to the road surface, respectively. On the other hand, turbulence noise around rolling tire and air-pumping noise caused by the air displacement into/out of cavities in or between the tire tread and the road surface are classified as the noise of aerodynamically-related mechanism. Moreover, pipe resonance in the grooves of tire tread pattern and Helmholtz resonance in the tire tread and the road surface can be categorized into the aerodynamically-related mechanism as a special case of the air-pumping mechanism.[2]

Since the first investigation of the air-pumping noise generation mechanism by Hayden[3] in 1971, most studies have characterized the air-pumping noise source as the volume change of air cavities in car tires. Besides, it has been assumed that the air displaced from the compressed groove feeds an acoustic monopole. In these approaches, the Euler equations are applied for an air-pumping noise source which is modeled as vibrating point monopole source. Here, the sound pressure is related to the second derivative of the volume changes of the squeezed cavities. Until now, lots of studies based on the monopole theory have been performed and some experimental investigations have shown satisfactory agreement with this approach.[4,5]

However, the monopole theory has an intrinsic weakness because it assumes small amplitude fluctuations of pressure and density in the tire groove or around the tire/road contact points for using linearized governing equations. Although

[a)]Electronic mail: yelorei7@snu.ac.kr
[b)]Electronic mail: wtjeong@mobis.co.kr
[c)]Electronic mail: qwer1104@snu.ac.kr
[d)]Author to whom correspondence should be addressed. Electronic mail: solee@plaza.snu.ac.kr

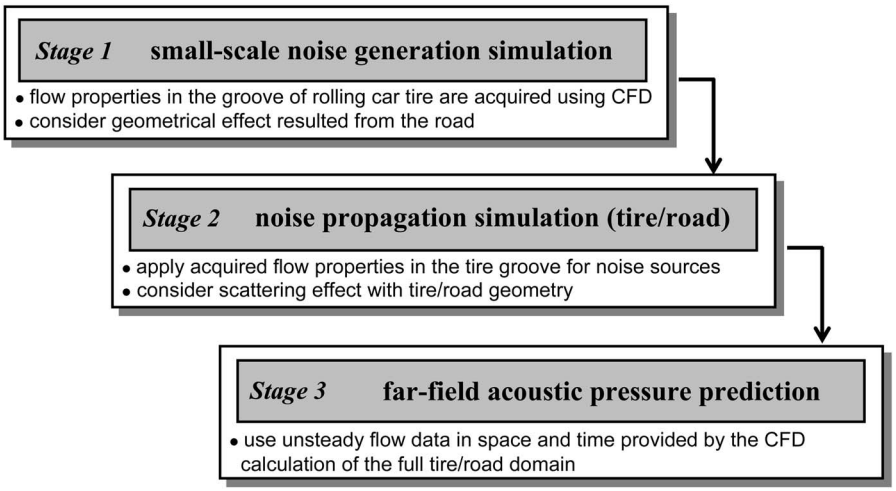


FIG. 1. Schematic diagram of the prediction procedure for the air-pumping noise (combined CFD/Kirchhoff method).

this approach models the air movement as a piston motion moving upward and downward in the cavity, it equates local air movements exactly with the volume changes of the cavity. In the monopole theory, therefore, the movement of the piston does not cause any compression or expansion of the adjacent air. As a consequence, the piston applied in the monopole theory is not directly related with the air-pumping noise generation but is a simple apparatus for the volume changes of the cavity.

Gagen[6] argued that, considering a typical period over which a cavity undergoes squeezing, the volume decrease of the cavity may occur before the air is able to fully evacuate from it. Furthermore, the volume loss of the cavity can cause potentially large density fluctuation. Thus, this invalidates the use of the usual acoustic wave equation which is based on the assumption of small amplitude fluctuation. In order to consider this nonlinear effect, he derived a "squeezed acoustic wave equation" suitable for application to squeezed fluid systems. He considered tire grooves in a rectangular geometry and used the Euler equations for fluid flow together with the moving boundary conditions specifying the cavity under compression. By obtaining an approximate analytic solution, he demonstrated that the usual assumption of the small amplitude acoustic wave equation and the acoustic monopole theory derived from this equation is incorrect for the squeezed system.

However, even in the derivation of the squeezed acoustic wave equation, geometrical effects on the air-pumping noise generation and propagation are not considered. In that the squeezed system doesn't have any wall boundary condition to take into account the contact road surface near the cavity exit, the compressed air may evacuate from the cavity during the whole period of the compression process. In real compression process of the air-pumping mechanism, however, due to the time delay between the volume changes and resultant air movement, some portion of compressed air may be trapped in the cavity. After that, as the cavity moves away from the contact patch, the trapped air is pressed away at the beginning of expansion process and interrupts suction of outside air into the cavity. (Several researchers modeled this noise generation mechanism as a Helmholtz resonance phenomenon, and investigated separately from the air-pumping mechanism.[7,8] Under the consideration of an air-pumping mechanism with the compressible flow properties, however, Helmholtz resonance phenomenon must not be separated.) Besides, when the cavity undergoes compression or expansion (volume decrease or recovery) process, the tire and the road surface form exponentially widening acoustic "horn" geometry close to the leading and trailing edges of the contact patch. And this horn geometry helps to radiate the generated noise and amplifies far-field noise amplitude. Obviously, the variations of these external boundary conditions may change the flow properties in the cavity. As a result, the generation mechanism and the propagation characteristics of the air-pumping noise are different from those of single squeezed system. Therefore, to examine the main features of the air-pumping noise and to find the hidden noise generation mechanism, the air-pumping noise source model and the noise prediction method which can consider the nonlinear noise generation and the geometrical scattering are needed.

In this article, a new method to predict the air-pumping noise in a rolling tire is proposed. This method is based on a hybrid technique which combines a computational fluid dynamics (CFD) technique with a Kirchhoff integral method. First, a Navier-Stokes finite volume solver is executed to solve the squeezed cavity system, which is newly modeled with piston/sliding-door/cavity geometry. Then the resultant flow properties in the grooves of the rolling tire are transferred to the flow simulation of the full tire/road domain as an air-pumping noise source. Finally, in order to predict the far-field acoustic pressure, the flow properties around the rolling tire and the road surface are transferred to the linear Kirchhoff integral method. Figure 1 summarizes the schematic of the proposed prediction procedures.

The outline of the remainder of this article is as follows. First the procedure of the proposed air-pumping noise prediction method is introduced in Sec. II, and numerical formulation is employed follow in Sec. III. The predicted results are provided in Sec. IV, which includes discussions about air-pumping noise generation mechanism in time and frequency domains. In Sec. V, the effect of ventilation on the air-pumping noise is investigated, and predicted results are compared with previous experimental data. The conclusion of this work is then presented in Sec. V.

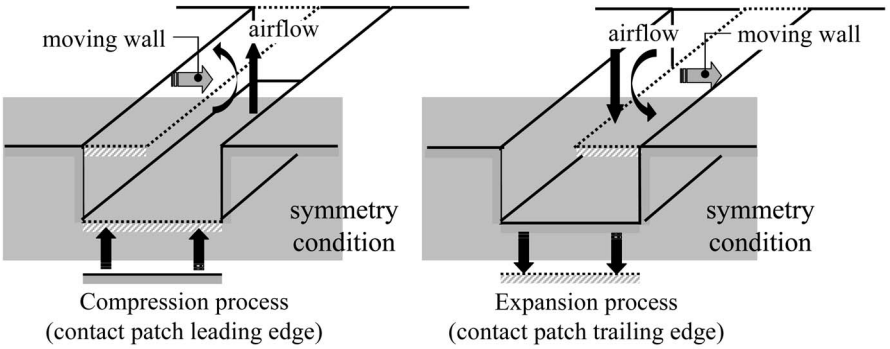


FIG. 2. Schematic diagrams of the piston/sliding-door/cavity source model and boundary conditions of the small scale noise generation simulation.

II. PREDICTION METHOD FOR THE AIR-PUMPING NOISE

A. Small-scale noise generation simulation with the piston/sliding-door/cavity model

When a tire rolls, a tire groove is squeezing as it enters the contact patch between the tire groove and the road surface. As the tire groove progresses through the contact patch, it closes up which results in the air which is trapped within it being compressed. After that, as the tire groove has been released from the contact patch, this compressed air is released and pressed away from it. This squeezing and releasing process of the tire groove is similar to that of pumping noise and known as air-pumping noise generation process.[2]

In the present study, to model this air-pumping noise generation process with considering the effect of the road surface, the piston/sliding-door/cavity model is proposed. The tire groove is modeled as a hexahedral cavity, which has a baffled cavity-cover that acts as a sliding-door and a base-side that moves like a piston. This model is shown schematically in Fig. 2. The squeezing/releasing and resulting volume decreasing/increasing process of the tire groove is modeled by the piston-like movement of the base-side of the cavity. In addition, the relative motion of the contact patch to the tire groove is modeled as the closing/opening motion of the sliding-door at the cavity open end, which has the same period as the compression/expansion motion of the piston. Here, it is assumed that there are no more volume decrease of the tire groove after the tire groove is fully covered by the road surface. Therefore, at the moment that the hexahedral cavity is closed by the sliding-door, the piston is paused and kept in its position for some time period which is determined by the tire rolling speed. After that, as the tire groove moves away from the contact patch trailing edge, the opening motion of the sliding-door and the expansion motion of the piston are started.

B. Near-field noise propagation simulation with the full tire/road geometry

The unsteady flow data obtained in the previous stage are used as air-pumping noise sources in the second stage, where noise propagation is simulated with emphasis placed on the scattering process with the full tire/road geometry. The flow properties in the compression process of the base-side piston are used as an air-pumping noise source in the volume decreasing process of the tire groove at the contact patch leading edge. Also, the properties in the expansion process are used as a noise source in the volume increasing process at the contact patch trailing edge. The schematic diagram of the full tire/road geometry and boundary conditions for the near-field noise propagation simulation are shown in Fig. 3.

In the present study, it is assumed that the tire has a cylindrical geometry. Moreover, the deformation of the tire side-wall is not considered. Thus, the full tire/road geometry has the same conceptual configuration as "the undeformed slick-treaded tyre model" in the TINO project.[9] In order to take into account the tire rolling motion, the angular velocity of the tire is imposed for the wall boundary condition of the cylinder, and the traveling speed of the tire is imposed for the wall boundary condition of the road surface. Further, at the nearest cells to the edges of the contact patch, the air-pumping noise sources are located as line sources. These source terms are determined by the mass flow rate and velocity vector data at the cavity open end, which are acquired in the numerical simulation of the piston/sliding-door/cavity model. The resulting source terms are then added to the governing equations.

C. Far-field acoustic pressure prediction from a Kirchhoff integral method

A linear Kirchhoff integral method is applied to predict the air-pumping noise of a rolling tire. The near-field flow

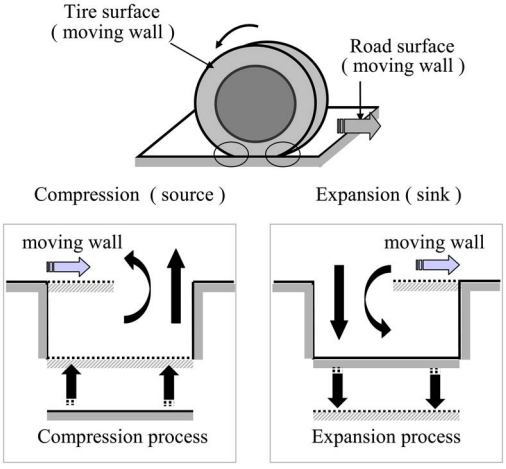


FIG. 3. (Color online) Schematic diagrams of the full tire/road geometry and boundary conditions of the near-field noise propagation simulation.

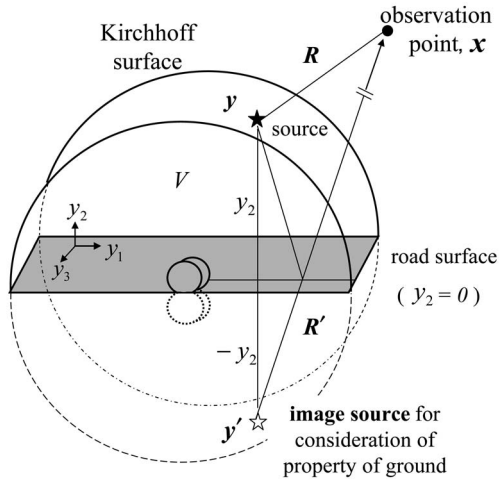


FIG. 4. Schematic diagram of the Kirchhoff surface for the prediction of the air-pumping noise and the coordinate system for half space Green's function (method of images)

data obtained by the CFD calculation of the full tire/road domain with modeled air-pumping noise sources are used as an input to the Kirchhoff integral formula. In order to obtain the coordinate system which is coincident with that of general laboratory road-wheel facility test,[4] the traveling speed of the tire is imposed for the wall boundary condition to the road surface. In this coordinate system, the tire position and the observation point become fixed in space. As a result, the Kirchhoff integral formula for a stationary control surface is acceptable. Then, this Kirchhoff integral formula can be regarded as the analytic expression of Huygens' principle.[10]

The integration is performed at the control surface, called the Kirchhoff surface, containing the flow field information. In the proposed method, the cylindrical surface that is magnified the tire surface can be considered as the control surface. Furthermore, in order to consider reflection effect by the road surface, the method of images[11] is applied by imposing the image sources under the road surface. Figure 4 shows the schematic of the image source and the Kirchhoff surface under the consideration.

III. NUMERICAL FORMULATION

A. Monopole theory

Though the monopole theory is not used in the proposed method, the authors will briefly introduce the monopole theory for later comparison with the proposed method.

The monopole theory applies the acoustic wave equation to a vibrating point source modeled as a small spherical emitter of radius, a, undergoing small oscillations much less than the radius. In case of a spherical wave centered at the coordinate origin, all the flow parameters are functions of the radial distance, r, and t only. Since the pressure gradient acts radially, the fluid velocity can only be in the radial direction. Thus, the wave equation reduced to

$$\frac{1}{c^2}\frac{\partial^2 p'}{\partial t^2} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial p'}{\partial r}\right), \tag{1}$$

and the general solution of Eq. (1) is as follows:

$$p'(r,t) = \frac{f(r-ct)}{r}, \tag{2}$$

where c is speed of sound, and f can be any functions.

Suppose that the radial velocity at radius, a on the surface of the spherical radiator is specified to be u_a. And, if the source is much smaller than the wavelength of sound, this surface vibration will generate the centered wave like

$$p'(r,t) = \frac{a^2}{r}\rho_0\frac{\partial u_a}{\partial t}\left(t - \frac{r-a}{c}\right), \tag{3}$$

where, ρ_0 is the density of air.

Meanwhile, the resulting volume change is related to the radial velocity on the surface of the sphere taken on the following form:

$$\frac{\partial V}{\partial t} = 4\pi a^2\frac{\partial a}{\partial t} = 4\pi a^2 u_a. \tag{4}$$

In turn, this allows us to substitute Eq. (4) back into Eq. (3) to obtain

$$p'(r,t) = \frac{1}{4\pi r}\rho_0\frac{\partial^2 V}{\partial t^2}\left(t - \frac{r-a}{c}\right), \tag{5}$$

which is relating the pressure wave emitted by a monopole source to the second derivative of the volume change of that source.[12]

B. CFD techniques

An unsteady, compressible, and three-dimensional Navier-Stokes equation is solved to analyze the flow field around a rolling tire and in the squeezed cavity system modeled as the piston/sliding-door/cavity geometry. The system of equations consists of a local time derivative term, three convective, and three viscous flux vectors. In physical coordinates, the governing equation written in the nondimensionalized conservative form is

$$\frac{\partial \mathbf{Q}}{\partial t} + \frac{\partial \mathbf{E}}{\partial x} + \frac{\partial \mathbf{F}}{\partial y} + \frac{\partial \mathbf{G}}{\partial z} = \frac{1}{\mathrm{Re}}\left(\frac{\partial \mathbf{E}_v}{\partial x} + \frac{\partial \mathbf{F}_v}{\partial y} + \frac{\partial \mathbf{G}_v}{\partial z}\right), \tag{6}$$

where conservative variables, $\mathbf{Q}$, and flux vectors are

$$\mathbf{Q}=\begin{pmatrix}\rho\\ \rho u\\ \rho v\\ \rho w\\ \rho e\end{pmatrix},\quad \mathbf{E}=\begin{pmatrix}\rho u\\ \rho u^2+p\\ \rho uv\\ \rho uw\\ (\rho e+p)u\end{pmatrix},$$

$$\mathbf{F}=\begin{pmatrix}\rho v\\ \rho uv\\ \rho v^2+p\\ \rho vw\\ (\rho e+p)v\end{pmatrix},\quad \mathbf{G}=\begin{pmatrix}\rho w\\ \rho uw\\ \rho vw\\ \rho w^2+p\\ (\rho e+p)w\end{pmatrix},$$

$$\mathbf{E}_v=\begin{pmatrix}0\\ \tau_{xx}\\ \tau_{xy}\\ \tau_{xz}\\ (e_v)_5\end{pmatrix},\quad \mathbf{F}_v=\begin{pmatrix}0\\ \tau_{yx}\\ \tau_{yy}\\ \tau_{yz}\\ (f_v)_5\end{pmatrix},\quad \mathbf{G}_v=\begin{pmatrix}0\\ \tau_{zx}\\ \tau_{zy}\\ \tau_{zz}\\ (g_v)_5\end{pmatrix}.$$

With the perfect gas assumption for the Newtonian fluids, shear stress can be written as

$$\tau_{ij}=\mu\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right)+\lambda\left(\frac{\partial u_k}{\partial x_k}\right)\delta_{ij}. \tag{7}$$

Here, from the local thermodynamic equilibrium state of fluid, viscous coefficients, λ, μ satisfy the Stokes hypothesis, $3\lambda+2\mu=0$. And, energy terms of viscous flux are represented as follows:

$$\begin{aligned}(e_v)_5&=u\tau_{xx}+v\tau_{xy}+w\tau_{xz}+\frac{\mu}{(\gamma-1)\mathrm{Pr}}\frac{\partial c^2}{\partial x},\\ (f_v)_5&=u\tau_{yx}+v\tau_{yy}+w\tau_{yz}+\frac{\mu}{(\gamma-1)\mathrm{Pr}}\frac{\partial c^2}{\partial y},\\ (g_v)_5&=u\tau_{zx}+v\tau_{zy}+w\tau_{zz}+\frac{\mu}{(\gamma-1)\mathrm{Pr}}\frac{\partial c^2}{\partial z}.\end{aligned} \tag{8}$$

To complete the set of compressible equations, the perfect gas equation of state is defined as

$$p=\rho(\gamma-1)\left\{e-\tfrac{1}{2}(u^2+v^2+w^2)\right\}, \tag{9}$$

where γ represents the specific heat ratio.

The solution algorithm is based on the PISO method,[13] and the fully implicit scheme[14] is used for time discretization and the quadratic upstream interpolation for convective kinematics (QUICK) scheme[15] is used for spatial discretization. The QUICK scheme is a higher order differencing scheme, and uses a three point upstream weighted quadratic interpolation for cell face values. The formulation of the QUICK scheme for uniform grid spacing is given by following expression:

$$\begin{aligned}f_r&=\tfrac{1}{2}(f_i+f_{i+1})-\tfrac{1}{8}(f_{i+1}-2f_i+f_{i-1}),\\ f_l&=\tfrac{1}{2}(f_i+f_{i-1})-\tfrac{1}{8}(f_i-2f_{i-1}+f_{i-2}).\end{aligned} \tag{10}$$

Based on an analysis of the truncation error the properties of the QUICK scheme for $\partial f/\partial x$ can be expressed as

$$\left.\frac{\partial f}{\partial x}\right|_i=\frac{f_r-f_l}{\Delta x}. \tag{11}$$

In the numerical simulation of the piston/sliding-door/cavity geometry, an additional equation called the geometric conservation law[16] (GCL) is solved for the moving coordinate velocity. The integral form of GCL is derived from the conservation of mass by setting the density of fluid, $\rho=1$ and the velocity of fluid flow, $v=0$. It can be written as follows:

$$\frac{d}{dt}\int_V dV=\int_S \mathbf{u}_s\cdot dS, \tag{12}$$

where $\mathbf{u}_s$ represents the local velocity of cell boundary, V is the volume of element, and S is the surface area of element. Equation (12) means that the change in volume of each control volume between two time instants, t^n and t^{n+1}, must be equal to the volume swept by the cell boundary during that time $\Delta t=t^{n+1}-t^n$. By a transformation from the Cartesian coordinate (x,y,z) to the body-fitted coordinate (ξ,η,ζ), the following form of the integral statement is acquired:

$$\frac{d}{dt}\int_V Jd\xi d\eta d\zeta=\int_V(\nabla\cdot\mathbf{u}_s)Jd\xi d\eta d\zeta. \tag{13}$$

Here, J represents the volume element in the transformed coordinate. Therefore, the computed value of J must be consistent with the value of ΔV implied by the numerical scheme used for solving the flow equation. Expanding the right-hand side of Eq. (13) and after performing necessary manipulations, the following form of the differential statement of GCL is obtained:

$$J_t+(\xi_t)_\xi+(\eta_t)_\eta+(\zeta_t)_\zeta=0, \tag{14}$$

where, ξ_t, η_t, ζ_t are the metric terms given by

$$\begin{aligned}\xi_t&=-[\dot{x}(y_\eta z_\zeta-y_\zeta z_\eta)+\dot{y}(z_\eta x_\zeta-z_\zeta x_\eta)+\dot{z}(x_\eta y_\zeta-x_\zeta y_\eta)],\\ \eta_t&=-[\dot{x}(y_\zeta z_\xi-y_\xi z_\zeta)+\dot{y}(z_\zeta x_\xi-z_\xi x_\zeta)+\dot{z}(x_\zeta y_\xi-x_\xi y_\zeta)],\\ \zeta_t&=-[\dot{x}(y_\xi z_\eta-y_\eta z_\xi)+\dot{y}(z_\xi x_\eta-z_\eta x_\xi)+\dot{z}(x_\xi y_\eta-x_\eta y_\xi)].\end{aligned} \tag{15}$$

Here, $\dot{x}$, $\dot{y}$, $\dot{z}$ are the grid velocities in the x, y, and z directions, respectively. Equation (14) is solved numerically to update the Jacobian values at each time step. The difference equation for the fully implicit scheme is given by

$$J^{n+1}=J^n-\Delta t[(\xi_t)_\xi^{n+1}+(\eta_t)_\eta^{n+1}+(\zeta_t)_\zeta^{n+1}], \tag{16}$$

where the metric coefficients are calculated from Eq. (15) at the $(n+1)$th time step and the grid velocities are calculated as follows:

$$\dot{x}^{n+1}=\frac{x^{n+1}-x^n}{\Delta t},\quad \dot{y}^{n+1}=\frac{y^{n+1}-y^n}{\Delta t},\quad \dot{z}^{n+1}=\frac{z^{n+1}-z^n}{\Delta t}. \tag{17}$$

C. Kirchhoff integral method and method of images

The Kirchhoff integral method applies the wave equation between a known acoustic surface, S, and a point in space within a uniform field.[17] The surface is assumed to contain all nonlinear flow effects and noise sources. This method allows propagation of a known sound field to a single point in space, and is useful when acoustic near field is already known.

The linear Kirchhoff integral formula is as follows:

$$p'(\vec{x},t) = \frac{1}{4\pi}\int_S \int \left[\frac{\cos\theta}{r^2}p' - \frac{1}{r}\frac{\partial p'}{\partial n} + \frac{\cos\theta}{cr}\frac{\partial p'}{\partial \tau} \right] \times dS(\vec{y},\tau). \tag{18}$$

Here, p' is the perturbed pressure, $(\vec{x},t)$ are the observer's location and time, $(\vec{y},\tau)$ are the source location and retarded time variables, θ is the angle between the normal vector $(\vec{n})$ on the surface and the radiation vector $(\vec{r})$, and r is the distance between an observer and a source at the retarded time. Note that pressure and its derivatives are calculated at the retarded time, τ.

To consider the road surface effect the method of images can be applied in the Kirchhoff integral formula.[11] Suppose that the mean flow is zero and let V be the region $y_2 \geqslant 0$ (shown in Fig. 4). It is constructed that the free space Green's function whose normal derivative vanishes on the boundary $y_2=0$. Since

$$G^0(\mathbf{y},\tau|\mathbf{x},t) = \frac{1}{4\pi R}\delta\left(\tau - t + \frac{R}{c_0}\right) \tag{19}$$

is a solution to the homogeneous wave equation when $\mathbf{y} \neq \mathbf{x}$, and since this equation is invariant under the transformation $y_2 \rightarrow -y_2$, a solution of the homogeneous wave equation that satisfies the causality condition is

$$h = \frac{1}{4\pi R'}\delta\left(\tau - t + \frac{R'}{c_0}\right), \tag{20}$$

where $R' = |\mathbf{x}-\mathbf{y}'|$ and $\mathbf{y}' = y_1\mathbf{i} - y_2\mathbf{j} - y_3\mathbf{k}$. Then, as $\mathbf{y}'$ is never in V, h is nonsingular in this region, the Green's function satisfies the boundary condition of the road surface in the region V is

$$G(\mathbf{y},\tau|\mathbf{x},t) = \frac{1}{4\pi R}\delta\left(\tau - t + \frac{R}{c_0}\right) + \frac{1}{4\pi R'}\delta\left(\tau - t + \frac{R'}{c_0}\right). \tag{21}$$

IV. PREDICTION OF AIR-PUMPING NOISE

The tire is modeled as a cylindrical geometry which has the dimension of 165/70 SR 13 tire and the traveling speed is 100 km/h (27.8 m/s). Thus, it has a dimension of 346.64 mm in radius ($0.5D$) and a width (W) of 165 mm. The contact patch length is assumed to be 189 mm, which is reported in previous experiment[18] in case of the tire being loaded by two passengers. In addition, the tread pitch of the tire is adjusted to 55.6 mm so that the fundamental tread impact frequency at the speed of 100 km/h is 500 Hz. The tire groove is assumed to be a hexahedral cavity which has a dimension of 6 mm in width (w), 6 mm in depth (d), and 165 mm in length (l). In Fig. 5, the schematic diagram of the tire groove geometry is shown. The tire groove is assumed to be a transverse pocket-type where all sides of the hexahedral cavity are closed except for one open end. Moreover, in order to consider the relative motion of the contact patch, the sliding door is located at the open end. The adoption of the transverse pocket-type groove has an advantage in investigating the nonlinearity of the air-pumping noise generation mechanism through the reduction of the air flow in the direction of tire side-wall. As reported in previous research,[19] the decrease in volume of the tire groove is assumed to be 10% of the total volume. In addition, the volume changes are assumed to be simple harmonic. The dimensions and numerical settings are listed in Table I.

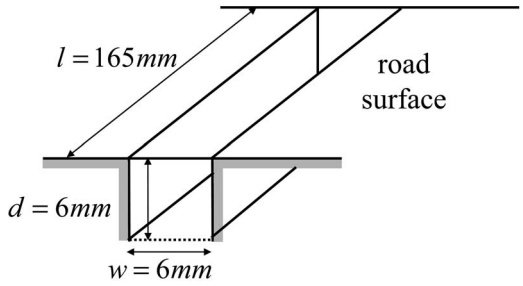


FIG. 5. Schematic diagram of the tire groove geometry (transverse pocket-type groove).

As mentioned earlier, when the monopole theory is applied to predict the air-pumping noise, directional characteristics of the air-pumping noise and the geometrical effects on noise generation and propagation cannot be investigated. Instead of using the monopole theory, therefore, the linear source model which equates the air movement exactly with the volume changes of the tire groove is introduced. Through the comparison of the linear and the nonlinear source model, the nonlinear characteristics of the air-pumping noise are investigated. In case of predicting air-pumping noise with the linear source model, the air velocity at the open end of the cavity is assumed to be identical to the velocity of the piston. In addition, the air density is to be constant in ambient value. Therefore, the input in numerical simulation of the full tire/road domain is acquired in algebraic manner and a numerical simulation of small-scale noise generation is not necessary.

TABLE I. Dimensions and numerical settings.

Tire radius (165/70 SR 13)	346.64 m
Adjustment of the tread pitch	55.6 mm
Car speed	100 km/h (27.8 m/s)
Contact patch length	189 mm
Tread passing frequency	500 Hz
Cavity compression/expansion period (T_p)	0.000 215 83 s
Total air-pumping period (T)	0.002 s
Total volume change	$165\times6\times0.6\times10^{-9}=594\times10^{-9}$ m^3
Variation of cavity depth in compression process	$d-d\times0.1\times\{\frac{1}{2}\times(1-\cos(\pi t/T_p))\}$

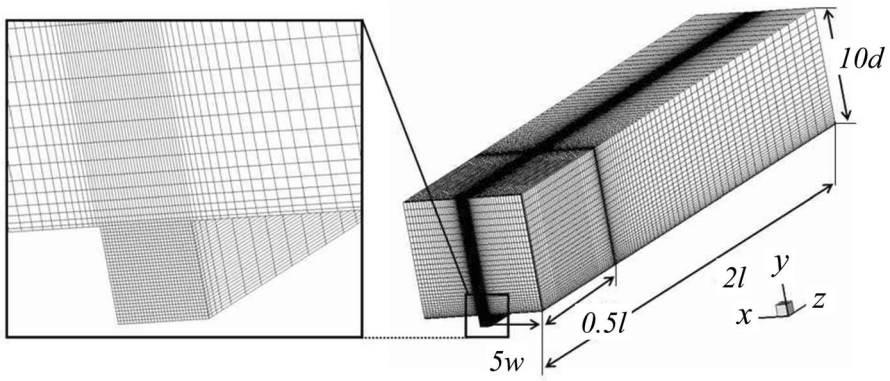


FIG. 6. Computational mesh for the simulation of the piston/sliding-door/cavity source model.

A. Small-scale noise generation simulation

The small-scale air-pumping noise generation process modeled with the piston-sliding-door/cavity geometry is numerically simulated. The grid system is characterized by a hexahedral free field domain and a cavity domain, which represents the tire groove. Figure 6 shows the computational mesh for the air-pumping noise source model simulation. The free field domain has a dimension of 10w in width, 10d in depth, and 2l in length. Total number of nodes is 406 172 and total number of cells is 384 000. In order to separate the air-pumping noise source to the turbulent noise source, it is assumed that there is no mean flow. Besides, the road surface is assumed to be perfectly reflective. Therefore, no-slip boundary condition is imposed at the cavity surfaces, the sliding-door surface, and the road surface. For the reduction of computational cost, a symmetric boundary condition is taken on the tire's plane of symmetry and the flow analysis is performed in a half of the tire groove.

Figure 7 shows the contours of the perturbed pressure in the air-pumping noise generation process. Each sectional diagram shows the perturbed pressure distributions on the planes parallel to the XY plane, located at distances 20, 40, 60, and 80 mm from the tire's plane of symmetry. Also, Fig. 8 shows streamlines on the tire's plane of symmetry.

At the contact patch leading edge, the cavity is gradually covered by the sliding-door and, at the same time, the air in it is compressed by the piston. Thus, the air flow in the cavity evacuates with a bias, and a vortical flow is generated outside the cavity. This vortical motion of the evacuating flow is suppressed by the road surface in real situation and by the wall boundary condition in the second stage flow simulation of the full tire/road domain. However, at the contact patch trailing edge, there exists a vortical flow in the cavity rather than the outside it. This vortical flow generation in the tire groove results from the collision of the compressed outgoing air flow and the incoming air flow suctioned by the piston. Judging from a standpoint of the monopole theory, therefore, the compressed air suppresses the volume increase of noise source. Thus, it is expected that the characteristics of the air-pumping noise are not harmonic even if the volume changes of the cavity are. That is to say, the compressibility of air and the relative motion of road surface to the tire groove will cause the difference between the spectral characteristics of air-pumping noise and that of noise source.

In addition, as seen in Fig. 7, even though the tire has transverse pocket-type grooves, there exists an air flow around the tire sidewall, or near the side of the cavity. The air is compressed and pressed away to this region at the contact patch leading edge, and expanded and sucked into the cavity from it at the contact patch trailing edge. Therefore, this perturbed pressure around the tire sidewall region will generate the far field air-pumping noise in this direction. However, in case of using the linear source model, even in the Helmholtz resonance approach, this air flow cannot be taken into account due to the assumption of equating air velocity to that of the moving piston.

B. Near-field noise propagation simulation

In the second stage, the unsteady flow data acquired by the small-scale noise generation simulation are used as air-pumping noise sources in the near-field noise propagation simulation with the full tire/road geometry. In Fig. 9, the computational mesh for the full tire/road geometry is shown. The computational domain has a dimension of 10D in radius and 10W in width. Total number of nodes is 250 650 and total number of cells is 240 044. As the piston/sliding-door/cavity simulation, a pressure prescribed boundary condition is applied at far boundary and a symmetric boundary condition is taken on the tire's plane of symmetry. A no-slip boundary condition is imposed on the tire and the road surface, which is adjusted by the angular velocity of the tire for the cylinder wall and the traveling speed for the road surface.

Figures 10 and 11 show the results of the simulation of the full tire/road domain with the linear source model (model 1) and with the proposed piston/sliding-door/cavity model (model 2), respectively. First, it is found that the tire/road geometry has an effect on not only the air-pumping noise generation process but also the air-pumping noise propagation. As the compression waves which are generated by the air-pumping noise sources at the contact patch edges propagate through the hornlike region, wave fronts are deformed normal to the surfaces around the tire and the road. This wave front deformation is observed both two model cases. Thus, it is inferred that the cause of this deformation of the wave form is not the nonlinear characteristics of the air-pumping noise generation mechanism but the horn-like geometry.

Though the geometrical effect is observed in both two model cases, it is easily seen the difference between model 1

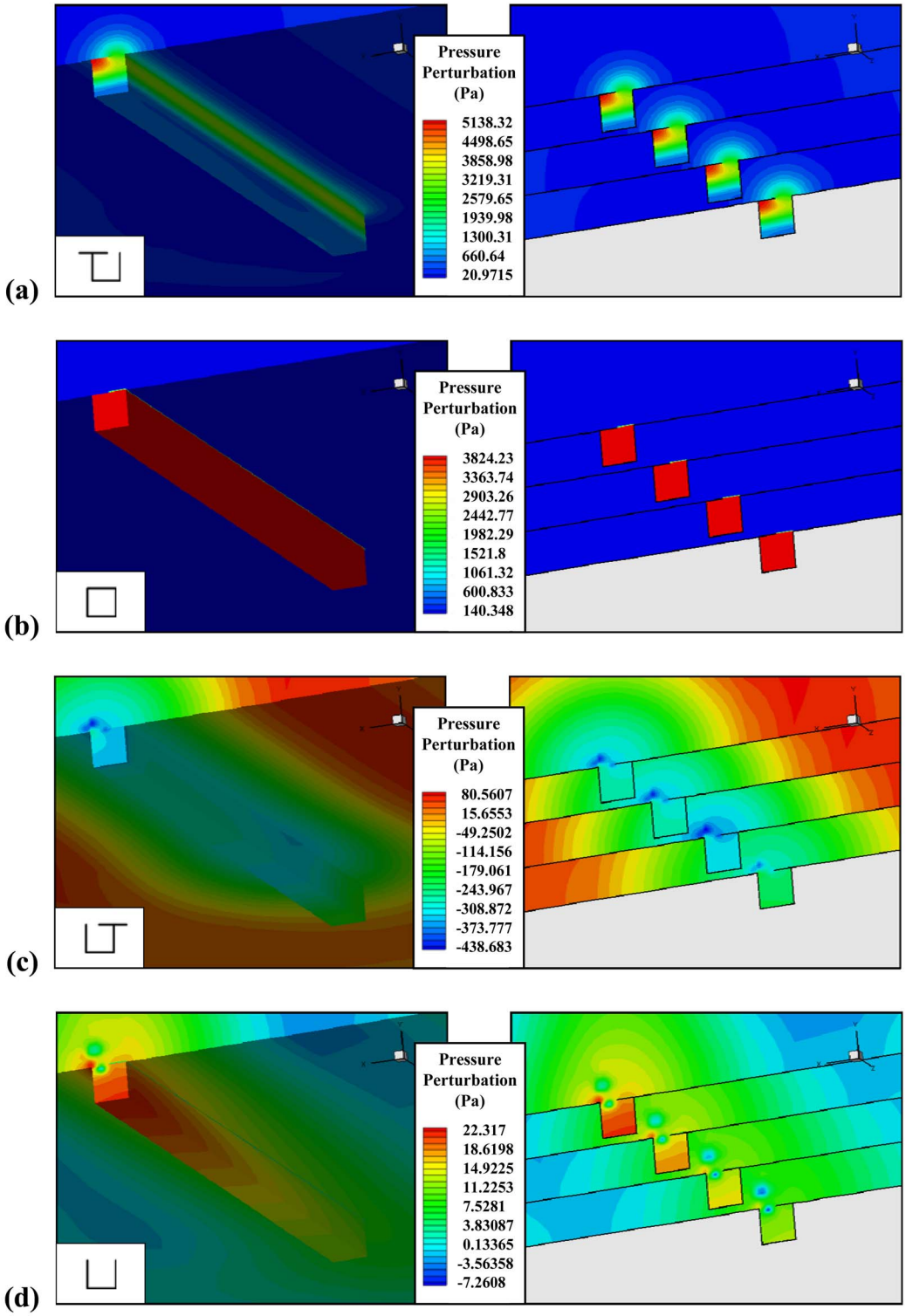


FIG. 7. (Color online) Pressure perturbation contours of the nonlinear (piston/sliding-door/cavity) source model in air-pumping noise generation process at (a) $t=T_p/2$ (contact patch leading edge), (b) $t=T$ (contact patch leading edge), (c) $t=T_p/2$ (contact patch trailing edge), and (d) $t=T$ (contact patch trailing edge).

and model 2. In case of model 1, the perturbed pressure is distributed almost uniformly along the tread area of the tire and weak pressure perturbation is observed on the tire sidewall. However, in case of model 2, at the initial stage of air-pumping, there exists a strong pressure perturbation near the tire sidewall, especially around the contact patch leading and trailing edge. This pressure perturbation on the tire sidewall is due to the air flow in the region of the side of the tire groove, which is observed in Fig. 7. As mentioned before, the pressure perturbation in the region of the tire side-wall will cause the far-field noise propagation in the direction of the side of a rolling tire and, as a result, the different directivity pattern from that of the air-pumping noise with the linear source model.

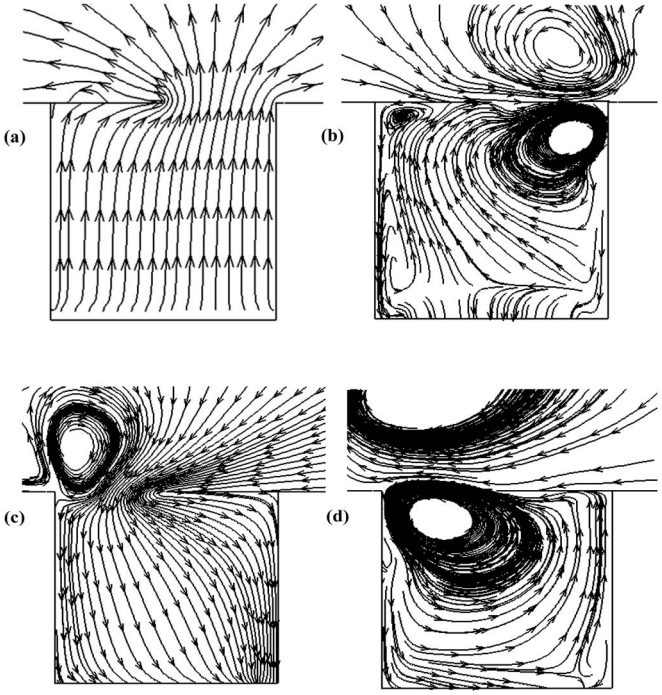


FIG. 8. Streamlines on the plane of symmetry of the tire at (a) $t=T_p/2$ (contact patch leading edge), (b) $t=T$ (contact patch leading edge), (c) $t=T_p/2$ (contact patch trailing edge), and (d) $t=T$ (contact patch trailing edge).

C. Far-field acoustic pressure prediction

The far-field acoustic pressure is predicted by use of the Kirchhoff integral formula with the near-field flow data obtained in the second stage. In the present study, the cylindrical surface which is magnified the tire surface five times is considered as the control surface. In addition, to consider perfectly reflective ground surface, the image sources of sound are imposed under the road surface.

Figure 12 shows the predicted acoustic pressure variations for several angles from the direction of travel during one air-pumping period at the position 7.5 m distance from the origin of the tire, the measurement point recommended in ISO R 362. From Eq. (5), the predicted acoustic pressure by use of the monopole theory is given by

$$p'(r,t) = \frac{1}{4\pi r}\rho_0 \frac{\partial^2 V}{\partial t^2}\left(t - \frac{r-a}{c}\right)$$

$$\approx 1.0772 \cos(14556t - 318.28)\ \text{Pa}. \qquad (22)$$

Because the monopole theory models the air-pumping noise source as a point monopole source, the anti-phasic behavior of the air-pumping noise sources at the contact patch leading and trailing edges is not considered. Further, the cancellation process through the acoustic reflection by the tire and the road surfaces is not taken into account, either. Therefore, the predicted acoustic pressure level of the monopole theory is higher than that of the proposed method, which is shown in Fig. 12.

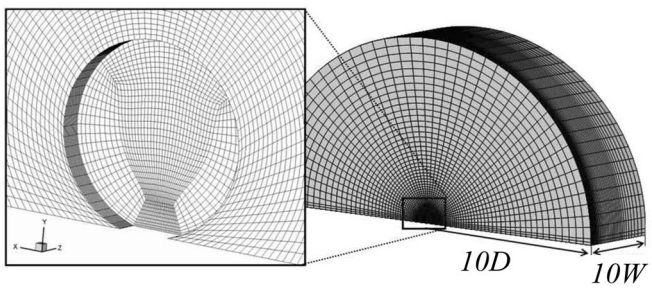


FIG. 9. Computational mesh for the simulation of the full tire/road domain.

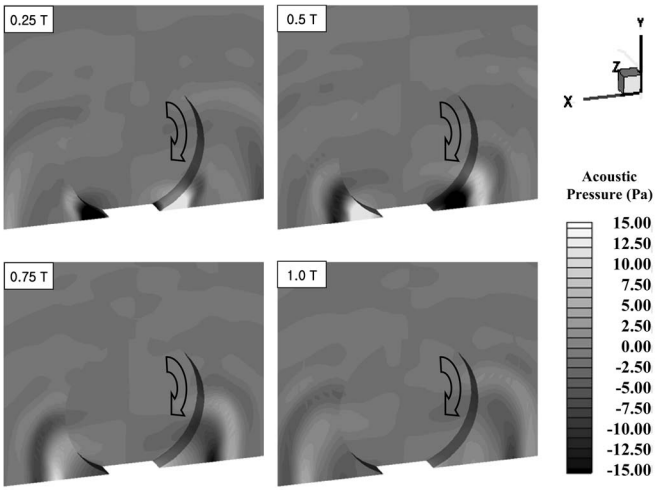


FIG. 10. Pressure perturbation contours of the full tire/road domain simulation with the linear source model in air-pumping noise generation process at (a) $t=T/4$, (b) $t=T/2$, (c) $t=3T/4$, and (d) $t=T$.

By comparison between the time history of the air-pumping noise generated with the linear source model (model 1) and that with the nonlinear source model (model 2), the nonlinear characteristics of the air-pumping noise can be ascertained. In case of model 1, the acoustic pressure amplitude has a symmetric distribution to the *YZ* plane (the axis of rotation of the tire, 90° Fig. 12) and its phase has an antisymmetric distribution. Moreover, there is little acoustic propagation in the direction of the axis of rotation of the tire. However, in case of model 2, contrary to the case of model 1, noise propagation in the direction of the axis of rotation of

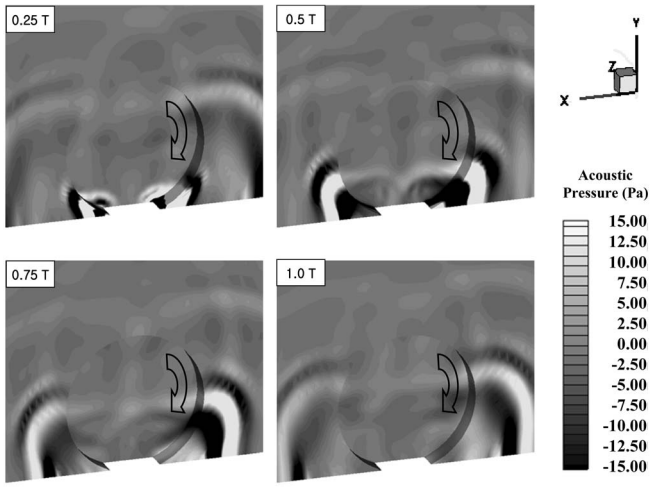


FIG. 11. Pressure perturbation contours of the full tire/road domain simulation with the nonlinear (piston/sliding-door/cavity) source model in air-pumping noise generation process at (a) $t=T/4$, (b) $t=T/2$, (c) $t=3T/4$, and (d) $t=T$.

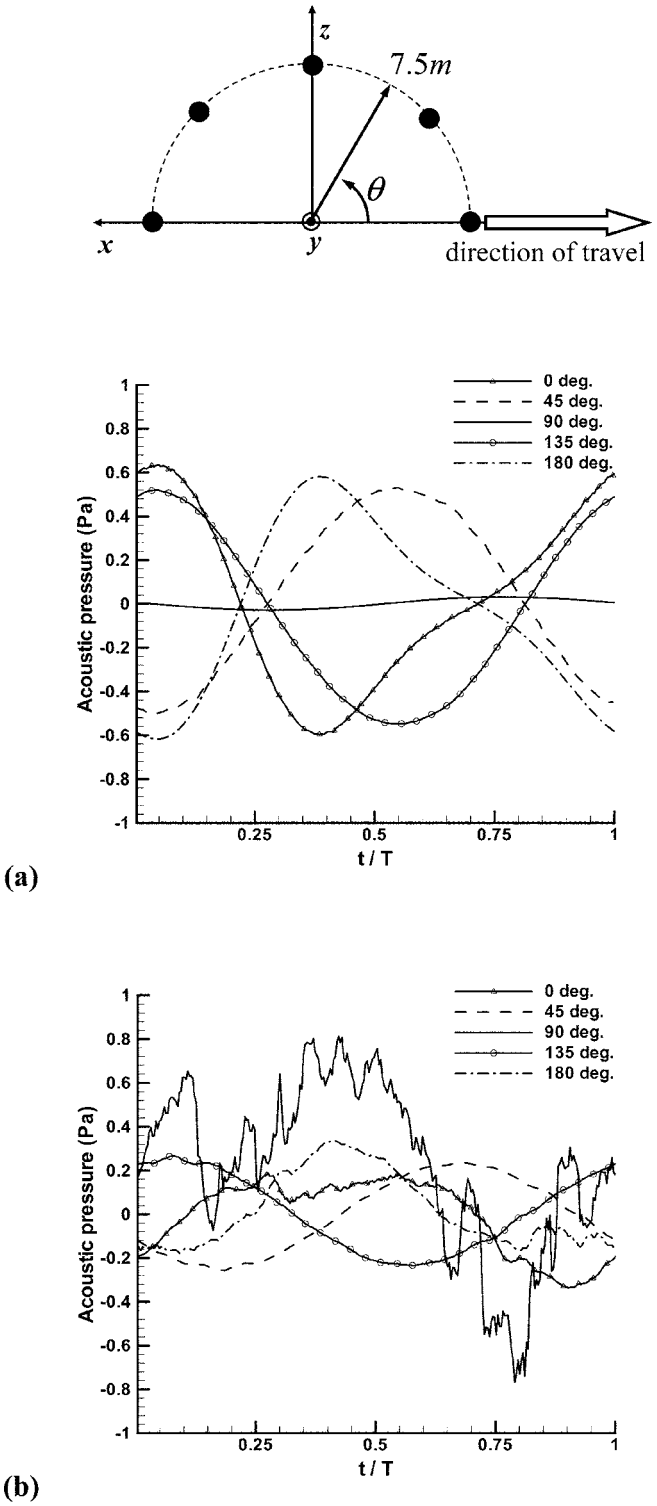


FIG. 12. Predicted acoustic pressure variations at the position 7.5 m distance from the origin of the tire (z=7.5 m) with (a) the linear noise source model and (b) the nonlinear (piston/sliding-door/cavity) source model during one period of air-pumping.

the tire is dominant. In addition, the level of the acoustic pressure of model 2 in the direction of travel of the tire (0° or 180°) is slightly lower than that of model 1. It seems that this disparity between model 1 and 2 results from the nonlinear behavior of the air-pimping noise source. In case of model 1, the phases of the air-pumping sources behavior at the contact patch leading and trailing edges are perfectly anti-phase each other. In addition, the nonmoving tire road geometry is symmetric to the axis. As a result, there exists a strong cancellation in the direction of the axis of rotation of the tire where retarded times from the contact patch leading edge and that from the trailing edge are nearly same each other. Thus, the level of the air-pumping noise is gradually decreasing as the observation point moves from the direction of travel (0° or 180°) to the direction of the axis of rotating of the tire (90°). But, in case of model 2, as mentioned before, the compressed air in the cavity is pressed away at the beginning of expansion process and interrupts suction of outside air into the cavity. Therefore, the phases of the air-pumping sources behavior at the contact patch leading and trailing edges are not perfectly anti-phase each other. Thus, there exists a weak cancellation. In addition, as seen in Fig. 7, a strong pressure perturbation is generated in the region of the side of the tire groove, or near the tire side-wall. Thus, this pressure perturbation may increase the noise propagation in the direction of the axis of rotation of the tire. Similar result is shown in the ISO-sound-intensity contour diagrams for crossbar tread tire on smooth surface reported in the experimental data by Donovan and Oswald.[20] The experimental data shows that the acoustic intensity level in the tire side-wall area (in the direction of the axis of tire rotation) is significant, although not as high as that at the contact patch leading edge region.

In order to investigate the directivity pattern of the air-pumping noise more clearly, the acoustic pressure distribution on the surface which has a radius of 7.5 m from the origin of the tire is shown in Fig. 13. As seen in Fig. 12, in case of the linear source model, the directivity patterns have an antisymmetric configuration to the *YZ* plane. Whereas, in case of the nonlinear source model, it is apparently seen that the noise radiation toward the direction of the axis of rotation of the tire is dominant and the level of the air-pumping noise in the forward direction is higher than that in the backward direction. This forward directivity results from the relative motion of the road surface to the tire groove and the compressibility of air. At the contact patch leading edge, the closing movement of the sliding-door (i.e., road surface) helps the air in the tire groove compressed more highly. At the contact patch trailing edge, however, the compressed air in the tire groove and the opening movement of the sliding-door prevent the tire groove from sucking the outside air. Therefore, the noise emission of the air-pumping sources at the contact patch leading edge becomes more efficient. Similar forward directivity is shown in the intensity level contour diagrams reported in the Technical University of Gdansk experimental data.[2] The experimental data show that the high-emission region locates at the leading edge of the footprint in the sound intensity maps for a summer tire tested at 90 km/h on a very smooth surface for frequency within the range 1000–3150 Hz, which is the generally reported frequency range of the air-pumping noise.

In order to investigate the characteristics of the air-pumping noise in frequency domain, the spectral analysis is performed with the acoustic pressure time history which is presented in Fig. 12. In Fig. 14, the predicted A-weighted air-pumping noise spectrum which is generally used in the measurement of the tire/road noise is presented at one-third

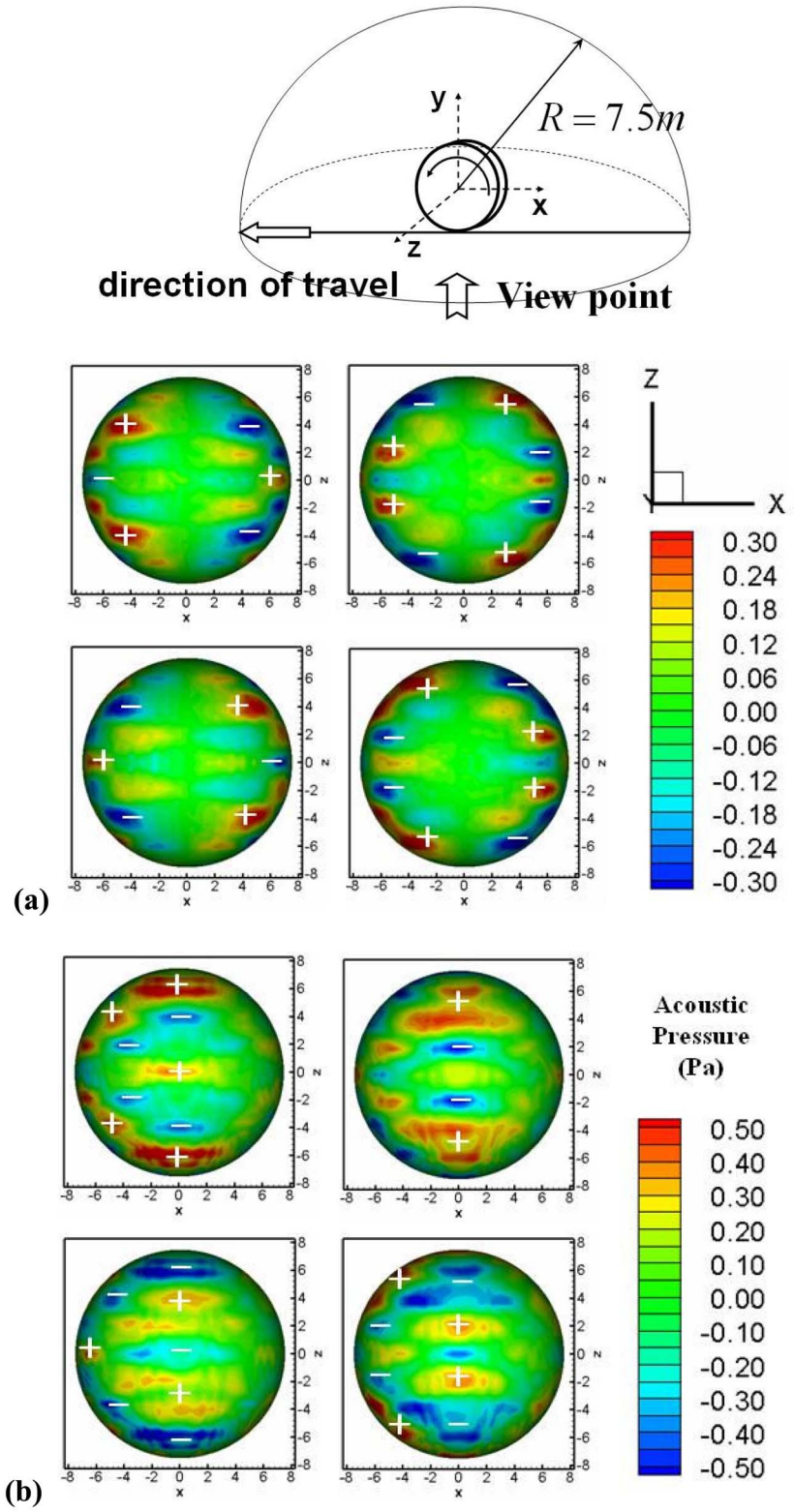


FIG. 13. (Color online) Predicted acoustic pressure distribution with (a) the linear noise source model and (b) the nonlinear (piston/sliding-door/cavity) source model during one period of air-pumping.

octave band frequencies. In the present study, the tread pattern periodicity was 55.6 mm, giving an expected excitation frequency of 500 Hz, which also appears in the spectra as a peak in both model cases. However, in case of model 2, the second peak at frequency approximately 1 kHz has a higher level than that of model 1. This higher level of the second peak frequency seems to result from the pipe resonance of the tire groove. The resonance frequency of the air in a pipe closed in both end is given by

$$f_r = \frac{c}{2l}, \tag{23}$$

where f_r is the resonance frequency (Hz), c is the speed of sound (m/s), and l is the length of the pipe (m). In the present study, the pipe resonance frequency is 1039.39 Hz. Thus, it is shown that the pipe resonance effect is predictable in case of model 2 in which the air-pumping noise source is modeled with the piston/sliding-door/cavity geometry and

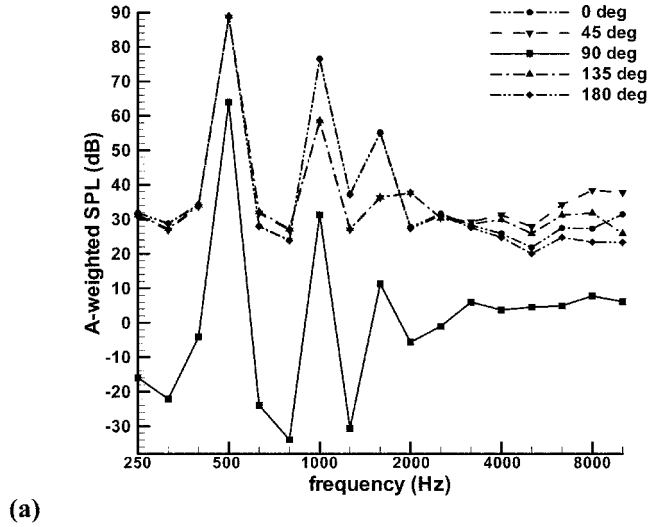


FIG. 14. Predicted A-weighted air-pumping noise spectra with (a) the linear noise source model and (b) the nonlinear (piston/sliding-door/cavity) source model at one-third octave band frequencies.

numerically simulated. Further, in case of model 2, the predicted noise levels within the range 2–8 kHz where it is reported that the air-pumping noise has most of its energy concentrated are increased, especially in the direction of the axis of rotation of the tire (90°). As mentioned before, the increase of noise emission in this direction is caused by the strong pressure perturbation near the tire sidewall, which results from the compressibility of the air and the relative motion of the road surface to the tire groove. Therefore, it is expected that the consideration of the nonlinear effect will be indispensable to study air-pumping noise generation and propagation.

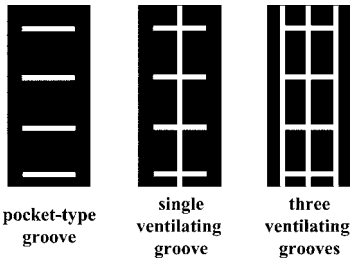


FIG. 15. Experimental tread patterns used in the study of the influence of pocket groove (Ref. 4).

V. INFLUENCE OF VENTILATION ON AIR-PUMPING NOISE

The air-pumping mechanism is believed to be one of the most important in tire/road noise generation and many researches for investigating the air-pumping noise generation mechanism are in progress. However, the tire/road noise generation mechanism is an extremely complicated mixture of dynamic mechanisms and related phenomena, so it is difficult to identify the air-pumping noise. Furthermore, there are few research results that can be quantitatively compared with the predicted result in the present study. Therefore, to validate the proposed method, predicted results are qualitatively compared with experimental results of the influence of the pocket-type grooves on air-pumping noise spectra performed by Ejsmont *et al.*[4] In their experiment, the influence of pockets in the tread was observed on a hand-cut experimental tire. The tread patterns used in the experiment are shown in Fig. 15. In their experiment the tire geometry, the traveling speed of the tire, and the tread pitch are same as the numerical settings of the present study. The length and depth of the experimental tire groove are not reported, however, so the tire groove is assumed to have the same dimension as that of the present study. In addition, the ventilating cavity would have the dimension as the tire groove width and a half of the tire groove depth. In Fig. 16, the schematic diagram of the ventilated piston/sliding-door/cavity source model is shown. In order to ventilate the tire groove, an additional hexahedral

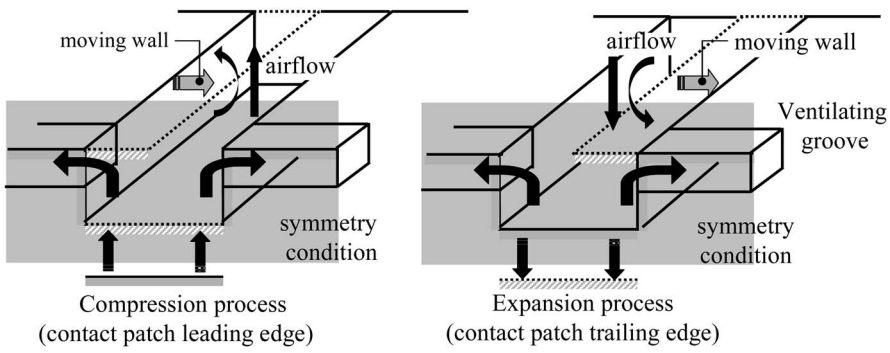


FIG. 16. Schematic diagram of the ventilated piston/sliding-door/cavity source model.

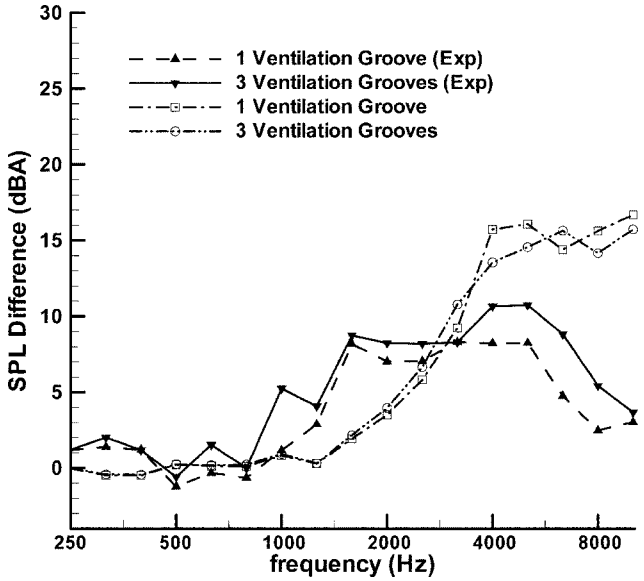


FIG. 17. Influence of the ventilating pocket-type groove on air-pumping noise spectrum. SPL difference between pocket type groove and ventilated pocket type groove at one-third octave band frequencies at the position 62.82° from the direction of travel on *XZ* plane.

cavity adjacent to the original piston/sliding-door/cavity model is equipped. As a result, though the tire groove is closed by the sliding-door in the air-pumping process, the air compressed by the piston can evacuate through the ventilating cavity.

Figure 17 shows the SPL difference between the pocket type groove and the ventilated pocket type groove at one-third octave band frequencies at the position 62.82° from the direction of travel on *XZ* plane. In the experimental result, the ventilation of the pockets by circumferential groove decreases the levels of frequencies within the range 2–6.5 kHz. However, the decrease of the predicted noise level occurs for frequencies within the range 3–8 kHz. This discrepancy in frequency range seems results from the difference of the tire groove and the ventilating groove geometry with the experimental tire. Even though the discrepancy is somewhat large in the low frequency domain, it shows that the predicted results showing a decreasing tendency of the noise levels at frequencies affected by the nonlinearity of air-pumping noise agree well with the experimental results. Through experimental results of the influence of the groove width[4] which has a nonlinear dependence of emitted sound intensity upon it, the groove width has been found to have dominant effects on the air-pumping noise level at the frequencies about 1–4 kHz. Meanwhile, as seen in Fig. 14, the frequency range in which the noise level is affected by the nonlinearity of the air-pumping noise is about 2–8 kHz in the present study. Thus, it can be inferred that the ventilation of the pocket-type groove has dominant effects on the noise sources which have nonlinear characteristics, and it can be checked in both experimental and predicted results.

Further, it is also found in both cases that a single ventilating groove eliminates the air-pumping noise level, which has nonlinear characteristics, almost as much as by three ventilating grooves.

VI. CONCLUDING REMARKS

In order to investigate the air-pumping noise from a car tire, a novel method based on a hybrid technique that combined a CFD technique with a Kirchhoff integral method is proposed. The air-pumping noise source behavior is obtained by the small-scale noise generation simulation with the piston/sliding-door/cavity model, which takes into account the relative motion of the road surface to the tire in the air-pumping noise generation process. The obtained flow properties in the tire grooves are then transferred to the flow simulation of the full tire/road domain as an air-pumping noise source. Finally, the far-field acoustic pressure is predicted by the linear Kirchhoff integral method through the use of the acquired flow properties around rolling tire and road surface geometry.

Through the comparison of the linear and the nonlinear source model simulations, it is found that the nonlinearity of the air-pumping noise source affects not only noise characteristics in frequency domain but also in the directivity pattern. While the sound field of the linear air-pumping source is symmetric to the axis of rotation of the tire, the directivity pattern of the nonlinear air-pumping source is asymmetric, i.e., the nonlinearity of the air-pumping noise source increases the emission of noise in the direction of the axis of rotation and in the direction of traveling of the tire. In the frequency domain, it is found that an increase in noise emission occurs for frequencies within the range of 2–8 kHz where it is reported that the air-pumping noise has most of its energy concentrated. Further, through a qualitative comparison of the ventilating effect with the measured data, it is also found that this noise peak level at frequencies 3–8 kHz is suppressed by a ventilation of the tire groove. It shows that the predicted results showing a decreasing tendency of the noise levels at frequencies affected by the nonlinearity of air-pumping noise agree well with the experimental results.

By use of the three-stage hybrid technique based on a numerical method with nonlinear governing equations, the proposed prediction method can take into account the effect of the tire/road geometry and the nonlinearity of the air-pumping noise source. Thus, the proposed method can cover the important physical processes of the air-pumping mechanism: the small-scale nonlinear noise generation, the mid-scale acoustic scattering process and the far-field noise propagation. From an engineering viewpoint, the proposed methodology is capable of providing us with an efficient method for further noise reduction and for the tire/road noise control.

ACKNOWLEDGMENT

This work was supported by Grant No. R01-2001-000-00401-0 from Korea Science and Engineering Foundation.

[1]M. Heckl, "Tyre noise generation," Wear **113**, 157–170 (1986).
[2]J. A. Ejsmont and U. Sandberg, *Tyre/Road Noise Reference Book* (INFOMAX, Harg, SE-59040 Kisa Sweden, 2002).
[3]R. E. Hayden, "Roadside noise from the interaction of a rolling tire with road surface," in Proceedings of the Purdue Noise Conference, West Lafayette, IN, 1971, pp. 62–67.
[4]J. A. Ejsmont, U. Sandberg, and S. Taryma, "Influence of tread pattern on tire/road noise," in *Transactions of the Society of Automotive Engineers*

(Society of Automotive Engineers, Warrendale, 1984), pp. 1–9.
[5]M. Jennewein and M. Bergmann, "Investigations concerning tyre/road noise sources and possibilities of noise reduction," Proc. Inst. Mech. Eng., Part C: Mech. Eng. Sci. **199**, 199–205 (1985).
[6]M. J. Gagen, "Novel acoustic sources from squeezed cavities in car tires," J. Acoust. Soc. Am. **106**, 794–801 (1999).
[7]J. F. Hamet, C. Deffayet, and M. A. Pallas, "Air-pumping phenomena in road cavities," in International Tire/Road Noise Conference 1990, pp. 19–29.
[8]N. A. Nilsson, "Air resonant and vibrational radiation—possible mechanisms for noise from cross-bar tires," in International Tire/Road Noise Conference 1979, pp. 93–109.
[9]P. Guisset and F. Augusztinovicz, "TINO noise emission: Analysis and prediction models," in First International Colloquium on Vehicle Tyre and Road Interaction, Paper 99.06, 1999.
[10]B. B. Bakerand and E. T. Copson, *The Mathematical Theory of Huygens' Principle* (Clarendon, Oxford, 1953).
[11]M. E. Goldstein, *Aeroacoustics* (McGraw-Hill, New York, 1976).
[12]K. J. Plotkin and E. Stusnick, "A unified set of models for tire/road noise generation," Wyle Research Report No. WR 81–26, Wyle Laboratories, Arlington, 1981.
[13]R. I. Issa, "Solution of the implicitly discretised fluid flow equations by operator-splitting," J. Comput. Phys. **62**, 40–65 (1986).
[14]R. D. Richtmeyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd Ed. (Wiley-Interscience, New York, 1967).
[15]B. P. Leonard, "A stable and accurate convective modeling procedure based on quadratic upstream interpolation," Comput. Methods Appl. Mech. Eng. **19**, 56–98 (1979).
[16]R. Kamakoti and W. Shyy, "Evaluation of geometric conservation law using pressure-based fluid solver and moving grid technique," Int. J. Numer. Methods Heat Fluid Flow **14**, 851–865 (2004).
[17]R. R. Mankbadi, *Transition, Turbulence, and Noise: Theory and Applications for Scientists and Engineers* (Kluwer Academic, Boston, 1994).
[18]N. A. El-Sebai, M. Watany, and A. Saad, "Tire/road interface airborne noise characteristics generation," in *Proceedings of the 1999 Noise and Vibration Conference* (Society of Automotive Engineers, Warrendale, 1999), pp. 605–613.
[19]I. D. Wilken, L. J. Oswald, and R. Kickling, "Research on individual noise source mechanisms of truck tires: Aeroacoustic source," in *SAE Highway Noise Symposium* (Society of Automotive Engineers, Warrendale, 1976), 762024, pp. 181–186.
[20]P. R. Donovan and L. J. Oswald, "The identification and quantification of truck tire noise sources under on-road operating conditions," Research Publication No. GMR-3380, General Motors Research Laboratories, Warren, MI, 1980.

A linearized Euler finite-difference time-domain sound propagation model with terrain-following coordinates

Dietrich Heimann[a] and Regina Karle
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre Oberpfaffenhofen, 82234 Weßling, Germany

(Received 19 July 2005; revised 4 April 2006; accepted 5 April 2006)

The equations of a linearized Euler finite-difference time-domain model for sound propagation through an inhomogeneous and moving atmosphere above absorbing ground are adapted to terrain-following coordinates. With the transformed numerical grid the effect of gentle orography is fully considered and the shortcomings of the representation of the orography by blocked-out grid cells in an untransformed grid are avoided. Comparisons with an analytical method to predict the interference patterns of a monochromic sound field above an idealized hill with rigid ground show good agreement. Further applications include absorbing ground and wind. A three-dimensional application also is discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200139]

PACS number(s): 43.50.Vt, 43.28.Fp, 43.28.Js, 43.28.En [DKW] Pages: 3813–3821

I. INTRODUCTION

During the past five years linearized Euler finite-difference time-domain (FDTD) models have become increasingly common for modeling outdoor sound propagation (Blumrich and Heimann, 2002; Salomons *et al.*, 2002; Heimann, 2003; Van Renterghem and Botteldooren, 2003; Heimann and Blumrich, 2004; Blumrich and Heimann, 2004; Ostashev *et al.*, 2005; Heutschi *et al.*, 2005). Also FDTD models were coupled with frequency-domain parabolic equation (PE) models (Van Renterghem *et al.*, 2005, 2006). One reason is that this type of numerical model has, in principle, no limitations with respect to the source-receiver geometry and the complexity of ground, topography and meteorology. The other reason consists in the fact that faster and more powerful computers became available such that simulations of outdoor sound propagation with realistic assumptions are now feasible. FDTD models have the potential to become a standard tool in outdoor sound propagation modeling provided the computer performance continues to grow. A weakness so far has been the difficulty of representing smooth orography, i.e., nonplane ground like hills and valleys, in orthogonal numerical grids of FDTD models. In this study the model equations are adapted to a terrain-following coordinate system which allows one to take into account orography without the drawbacks of a stair-step representation of the ground surface.

The consideration of orography in outdoor fluid dynamics by a transformed terrain-following coordinate system goes back to Phillips (1957). Mahrer and Pielke (1975) introduced terrain following coordinates in early meteorological mesoscale modeling. Since then the use of terrain-following coordinates with height, pressure or entropy as the vertical coordinate, is the standard method to include the effect of hilly or mountainous terrain in meteorological numerical simulations and numerical weather prediction. Coordinate transformations were also used in seismic wave propagation modeling to consider the topography (Hestholm and Ruud, 1994; Hestholm, 1999). With respect to acoustical modeling, but not with the focus on outdoor sound propagation, Botteldooren (1994) proposed a FDTD model which uses a quasi-Cartesian grid to incorporate the effects of tilted and curved boundaries. Instead of modifying the whole computational grid, Tolan and Schneider (2003) introduced a locally conformal method near slanted or curved surfaces. In the context of outdoor sound propagation of an orographically structured terrain, Sack and West (1995) transferred the concept of terrain-following coordinates to a PE frequency-domain sound propagation model. Now, the FDTD model of Blumrich and Heimann (2002) with its extension to a ground layer by Salomons *et al.* (2002) is adapted to terrain following coordinates to broaden its range of applicability. This model also includes the effect of meteorology.

The untransformed model equations (in orthogonal coordinates) are presented in Sec. II. These equations consider an inhomogeneous moving atmosphere and a ground layer. Section III deals with the terrain-following coordinate system. In Sec. IV the transformed equations are applied to sound propagation over idealized hills in two and three dimensions. The results are compared with analytical solutions. Some tests are performed with and without wind and with rigid and absorbing ground. However, it is the intention of this paper to keep the applications as simple as possible because they should mainly show the hill effect without a simultaneous influence of too many other effects. Finally, conclusions are drawn in Sec. V.

II. MODEL EQUATIONS

The prognostic FDTD model equations are defined in a Cartesian coordinate system with the position vector $\mathbf{r}=x\cdot\mathbf{i}+y\cdot\mathbf{j}+z\cdot\mathbf{k}$ ($\mathbf{k}$ pointing into the vertical direction). They provide the tendencies of the particle velocity vector $\mathbf{u}=u\cdot\mathbf{i}+v\cdot\mathbf{j}+w\cdot\mathbf{k}$ and the sound pressure p under the influence of three-dimensional (3D) distributions of wind $\mathbf{u}_{\mathrm{met}}=u_{\mathrm{met}}\cdot\mathbf{i}+v_{\mathrm{met}}\cdot\mathbf{j}+w_{\mathrm{met}}\cdot\mathbf{k}$ and air temperature T_{met}. The air tempera-

[a]Electronic mail: d.heimann@dlr.de

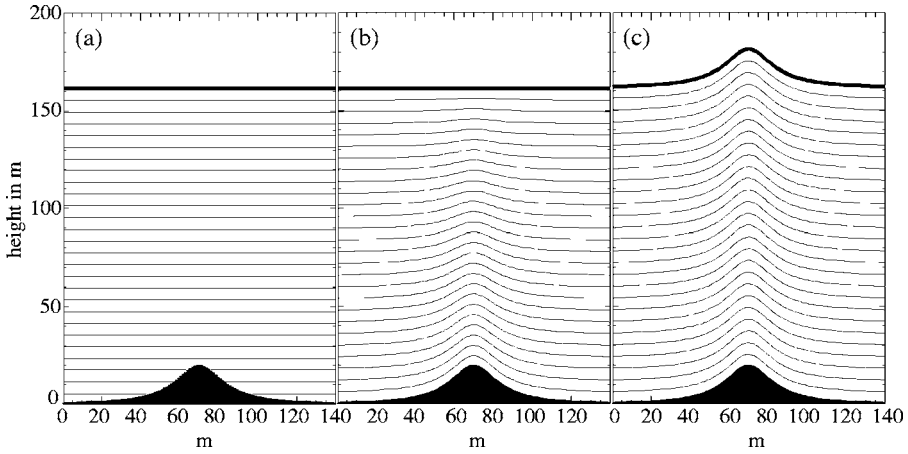


FIG. 1. Vertical cross section of the numerical grid levels for an untransformed grid (a), for a grid transformed with Eq. (9) (b), and for a grid transformed with Eq. (10) (c).

ture is linked with the air density ρ_{met} and the air pressure p_{met} through the equation of state $R_d T_{\text{met}} = p_{\text{met}}/\rho_{\text{met}} = c^2/\kappa$ with the gas constant of dry air $R_d = 287\ \text{J kg}^{-1}\ \text{K}^{-1}$ and the ratio of the specific heats of air at constant pressure and constant volume $\kappa = c_p/c_v = 1.40$. The sound speed in resting air is denoted by c. All meteorological parameters are assumed to be stationary during the propagation time between source and receiver although they may vary spatially throughout a 3D volume.

The model equations for sound propagation in dry air (moving medium) are given, e.g., by Blumrich and Heimann (2002). In this study these equations are used in an approximated form. A justification for these approximations is given in Appendix A. The approximations reduce the number of terms and therefore keep the computation time of the numerical solution within feasible limits. The model equations as they are used in this paper correspond to Eqs. (A5) and (A6) of Appendix A and are here written in the component form:

$$\frac{\partial u}{\partial t} = -u_{\text{met}}\frac{\partial u}{\partial x} - v_{\text{met}}\frac{\partial u}{\partial y} - w_{\text{met}}\frac{\partial u}{\partial z} - \frac{1}{\rho_{\text{met}}}\frac{\partial p}{\partial x}, \tag{1}$$

$$\frac{\partial v}{\partial t} = -u_{\text{met}}\frac{\partial v}{\partial x} - v_{\text{met}}\frac{\partial v}{\partial y} - w_{\text{met}}\frac{\partial v}{\partial z} - \frac{1}{\rho_{\text{met}}}\frac{\partial p}{\partial y}, \tag{2}$$

$$\frac{\partial w}{\partial t} = -u_{\text{met}}\frac{\partial w}{\partial x} - v_{\text{met}}\frac{\partial w}{\partial y} - w_{\text{met}}\frac{\partial w}{\partial z} - \frac{1}{\rho_{\text{met}}}\frac{\partial p}{\partial z}, \tag{3}$$

$$\frac{\partial p}{\partial t} = -u_{\text{met}}\frac{\partial p}{\partial x} - v_{\text{met}}\frac{\partial p}{\partial y} - w_{\text{met}}\frac{\partial p}{\partial z} - \kappa p_{\text{met}}\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right]. \tag{4}$$

The model equations for sound propagation in the ground (porous medium) are based on Zwikker and Kosten (1949). According to Salomons *et al.* (2002) they read:

$$\frac{\partial u}{\partial t} = -\frac{\Omega}{c_s}\frac{1}{\rho_{\text{met}}}\left[\frac{\partial p}{\partial x} + \sigma u\right], \tag{5}$$

$$\frac{\partial v}{\partial t} = -\frac{\Omega}{c_s}\frac{1}{\rho_{\text{met}}}\left[\frac{\partial p}{\partial y} + \sigma v\right], \tag{6}$$

$$\frac{\partial w}{\partial t} = -\frac{\Omega}{c_s}\frac{1}{\rho_{\text{met}}}\left[\frac{\partial p}{\partial z} + \sigma w\right], \tag{7}$$

$$\frac{\partial p}{\partial t} = -\kappa\frac{p_{\text{met}}}{\Omega}\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right]. \tag{8}$$

The ground properties are described by the flow resistivity σ (SI-unit: Pa s m^{-2}), the porosity Ω (dimensionless), and a structure parameter c_s (dimensionless).

Body sources of acoustic waves are omitted from the previous equations. In the subsequent numerical examples, acoustic wave fields are generated by imposing a time-varying sound pressure at a source point either within or on the boundary of the domain.

The model is numerically implemented on a staggered orthogonal grid. The grid corresponds to "Type C" after Arakawa and Lamb (1977) with the pressure defined in the center of the grid cells and the velocity components in the centers of the respective cell sides. Figure 1(a) shows the model levels in a vertical cross section. Rigid obstacles (buildings, barriers) can be considered by "blocking out" the respective grid cells, i.e., by setting the normal components of the particle velocity to zero at the interfaces of these grid cells. Topographical features with sloped sides (dams, embankments) or orographic elements (hills, mountains, valleys) can also be introduced by "blocked out" cells. However, in this case the stair-step representation of slopes gives rise to spurious reflections. As another drawback, the consideration of large hills or deep valleys has the consequence that a greater number of grid cells are underground and therefore computer storage is wasted. The introduction of terrain following coordinates overcomes these problems.

III. COORDINATE TRANSFORMATION

Terrain-following coordinates applied in meteorological mesoscale modeling (Pielke, 2002) are often based on the transformation rule described by Gal-Chen and Sommerville (1975)

$$\hat{x} = x; \quad \hat{y} = y; \quad \hat{z} = H\frac{z - h(x,y)}{H - h(x,y)}, \tag{9}$$

where $\hat{x}$, $\hat{y}$, and $\hat{z}$ denote the transformed spatial coordinates. H is the height of the top of the model domain and $h(x,y)$ describes the height of the terrain. The transformation rule of Eq. (9) causes the model levels to change gradually from terrain following at the lower boundary to plane at the upper boundary of the model domain. The model levels are therefore condensed above elevated regions [Fig. 1(b)]. This type of coordinate transformation has distinct advantages in meteorological simulation. However, in the modeling of outdoor sound propagation between near-ground sources and receivers only a rather shallow atmospheric layer is involved. Therefore it is more convenient to apply a coordinate transformation by which the upper boundary of the domain is parallel to the ground. An appropriate transformation rule is given by

$$\tilde{x} = x; \quad \tilde{y} = y; \quad \tilde{z} = z - h(x,y). \tag{10}$$

This coordinate system preserves the spatial resolution within the model domain even if the domain does not extend to high altitudes [Fig. 1(c)].

The transformation of the model equations [Eqs. (1)–(4) in the air, Eqs. (5)–(8) in the ground] is explained in Appendix B. The transformed model equations are given in Appendix B by Eqs. (B1)–(B4) (in the air) and Eqs. (B5)–(B8) (in the ground). The system of differential equations is integrated in time on a staggered grid in the transformed coordinates. Again Type C after Arakawa and Lamb (1977) is used. First-order forward-in-time and centered-in-space finite differences are employed. For further details of the numerical scheme and the treatment of boundaries we refer to Blumrich and Heimann (2002) and Heimann and Blumrich (2004).

Eqs. (B1)–(B8) (see Appendix B) can be used to simulate sound propagation over gentle terrain. However, steep slopes with $|\nabla h| > 1$ may lead to numerical instability unless the time step is correspondingly adapted (see Appendix B). In this case it is therefore appropriate to use an untransformed orthogonal numerical grid with blocked-out grid cells. The same applies to vertical or overhanging terrain features (e.g., rocks) and obstacles of complex shape (e.g., buildings, roof overhangs, balconies, viaducts) which cannot be modeled by transformed coordinates as defined in Eq. (10).

IV. APPLICATION TO SOUND PROPAGATION OVER A HILL

To demonstrate the suitability of the coordinate transformation, the model was first applied in two dimensions (vertical cross section) to sound propagation over an idealized hill with rigid ground. A calm atmosphere ($u_{\mathrm{met}} = w_{\mathrm{met}} = 0$) with $p_{\mathrm{met}} = 101\,325$ Pa and $\rho_{\mathrm{met}} = 1.20$ kg m^{-3} is assumed. This corresponds to $T_{\mathrm{met}} = 20\,^{\circ}$C and $c = 343.4$ m s^{-1} throughout the domain.

In the following we apply the identities for the coordinates $\tilde{x} = x$, $\tilde{y} = y$ and the spatial increments $\Delta\tilde{x} = \Delta x$, $\Delta\tilde{y} = \Delta y$, and $\Delta\tilde{z} = \Delta z$ [Eq. (10)]. The tilde is used only to indicate the relative height above ground ($\tilde{z}$) and the vertical velocity components in the terrain-following coordinates ($\tilde{w}, \tilde{w}_{\mathrm{met}}$). The model domain extends over $0 \leqslant x \leqslant 162$ m and $0 \leqslant \tilde{z} \leqslant 162$ m with a grid interval of $\Delta x = \Delta z = 0.3$ m. A time step of $\Delta t = 4 \times 10^{-4}$ s ensures the fulfillment of the Courant-Friedrichs-Lewy criterion for numerical stability (see Appendix B).

A "Witch of Agnesi" profile was chosen for the hill as the lower boundary condition

$$h(x) = h_{\mathrm{top}}\left[1 + \frac{(x - x_0)^2}{\ell^2}\right]^{-1} \tag{11}$$

with the height $h_{\mathrm{top}} = 20$ m, the width $\ell = 15$ m, and a position of the hill top at $x_0 = 70$ m. The maximum elevation gradient is $(\partial h/\partial x)_{\max} = \pm 0.9$.

A sound source is defined as a time-varying boundary condition at one ground based grid cell at the left domain boundary with the cell center exactly at ($x_s = 0.5\Delta x$; $\tilde{z}_s = 0.5\Delta z$). Here the sound pressure p is prescribed as a steady harmonic pressure oscillation with a frequency of $f = 50$ Hz (period: $f^{-1} = 0.02$ s, wavelength: $\lambda = 6.9$ m). A symmetry boundary condition with $u(x_s - 0.5\Delta x) = -u(x_s + 0.5\Delta x)$ is employed at $x = 0$. A damping layer is not implemented at this boundary.

At the remaining (top and right) boundaries a damping layer of a thickness of four times the wave length is employed in addition to the impedance boundary condition [Eqs. (15) and (16) in Blumrich and Heimann, 2002] to avoid spurious reflections. For details of the damping layer refer to Eq. (5c) in Heimann and Blumrich (2004).

In the following applications the use of a continuous monochromatic source was given preference over an initial broadband sound pulse for the following two reasons. (1) The positions of interference patterns are very sensitive to variations in the shape of ground surface. Therefore, deviations in the position of the interference patterns between numerical and analytical solutions are a good indicator of inaccuracies. In order to visualize interference patterns in the case of a pulse source it would be necessary to store the sound pressure time evolution at all grid points. (2) The use of a monochromatic wave ensures a fixed ratio between wave length (λ) and the grid interval (Δx, Δy, Δz), i.e., the choice of the grid interval can be adapted to the wave length to ensure a numerically sufficient and uniform spatial resolution of the sound waves.

Results are shown in terms of sound pressure level $\mathrm{SPL} = 10\log(p_a^2/p_0^2)$ with $p_0 = 2 \times 10^{-5}$ Pa. p_a is the effective sound pressure amplitude which can be evaluated during the integration by averaging the squared sound pressure at each grid point over the last period of the integration time. The latter is set to 0.75 s to ensure that the turn-on signal has already left the domain and a steady-state sound field has established.

Figure 2(a) illustrates the resulting SPL for the model run with original Cartesian coordinates [Eqs. (1)–(4)]. In this case the hill is represented by 8960 blocked out grid cells. The result shows three major features: (1) An increase of the sound level near the front slope as a consequence of multiple reflection on the concave part of the slope. (2) A series of interference patterns above the hill which are caused by the

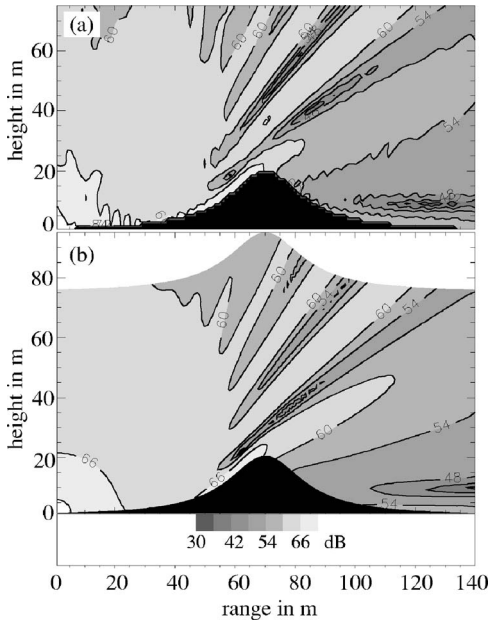


FIG. 2. SPL in dB of simulated 2D sound propagation from a ground based source (50 Hz) over a 20 m high hill with rigid ground: (a) untransformed grid with blocked out terrain, (b) transformed terrain-following grid. The contours are based on every fourth grid point.

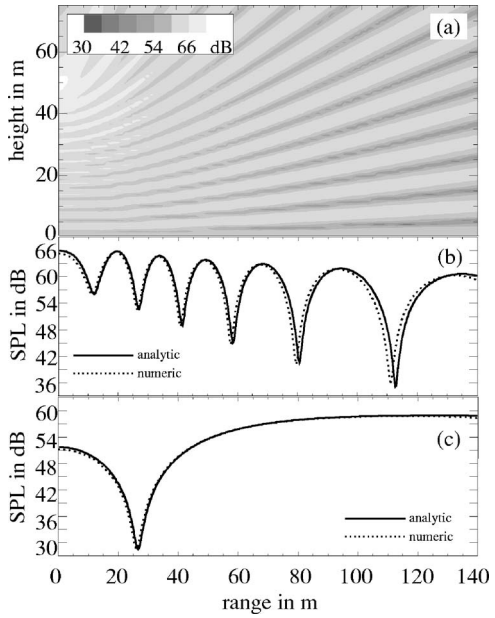


FIG. 3. SPL in dB of 2D sound propagation from a 50 m high source (50 Hz) over flat rigid ground. (a) Cross-section of the numerical FDTD model result (contours based on every fourth grid point). The lower panels show horizontal profiles at $z=30.15$ m (b) and $z=1.65$ m (c) of the analytical solution (solid) and the numerical solution (broken).

superposition of direct sound and sound reflected at the front slope of the hill. (3) A reduced sound level in the back side of the hill as a consequence of shading. Figure 2(b) shows the results for the same situation but based on the transformed Eqs. (B1)–(B4). Basically, the model run with transformed coordinates shows the same features as the run with untransformed coordinates. However, the results based on the transformed equations are much smoother than those based on the orthogonal grid. The ripples in the contour lines of Fig. 2(a) are caused by reflections at the unrealistic stair-step like orography.

The model results obtained with the transformed equations for propagation over rigid ground are now compared with the analytical solution of the superposition of direct and totally reflected sound with N possible reflections:

$$p_{a,r} = p_{a,s}\left| \frac{1}{\sqrt{kd_{sr}}}\exp[-ikd_{sr}] + \sum_{n=1}^{N} \frac{\sqrt{A_{g,n}/A_{r,n}}}{\sqrt{kd_{sg,n}}} \times \exp[-ik(d_{sg,n}+d_{gr,n})] \right|. \quad (12)$$

Equation (12) refers to two-dimensional (2D) cylindrical waves generated by an infinitely long line source. It is a far field approximation. The sound pressure amplitude approaches infinity towards the source point. $p_{a,r}$ is the sound pressure amplitude at the receiver, k is the wave number, and d_{sr} is the ray path length between source and receiver of the direct ray. The sum adds the contributions of all $N \geqslant 1$ possible reflected rays between source and receiver, where $d_{sg,n}$ is the path length between source and the nth reflection point and $d_{gr,n}$ is the path length between the nth reflection point and the receiver. $A_{g,n}$ and $A_{r,n}$ are the ray tube cross sections of the nth reflected ray at the reflection point and the receiver, respectively. According to Chapter 8.8 of Pierce (1981) the cross-section ratio for a 2D cylindrical wave reflected at a 2D cylindrical surface element is given by

$$\frac{A_{r,n}}{A_{g,n}} = 1 + \frac{d_{gr,n}}{d_{sg,n}} + 2d_{gr,n}\frac{\partial^2 h}{\partial x^2}(x_{g,n})\cos\,\theta_n, \quad (13)$$

where θ_n denotes the angle of wave incidence at the nth reflection point $x_{g,n}$. The near field reference pressure amplitude $p_{a,s}$ was adjusted such that the results of the numerical model are matched in the case of free field propagation. The path lengths and the positions $(x_{g,n};\tilde{z}_{g,n})$ of the reflection points of all N possible ray paths between the known positions of the source $(x_s;\tilde{z}_s)$ and the receivers $(x_r;\tilde{z}_r=\text{all grid cell centers})$ were iteratively determined with the following equation which expresses Snell's Law of Reflection:

$$(x_{g,n}-x_s) + (\tilde{z}_{g,n}-\tilde{z}_s)\frac{\partial h}{\partial x}(x_{g,n}) = \left((x_r-x_{g,n}) + (\tilde{z}_r-\tilde{z}_{g,n})\frac{\partial h}{\partial x}(x_{g,n})\right)\frac{d_{sg,n}}{d_{gr,n}}. \quad (14)$$

As Eq. (12) cannot predict the sound level in acoustical shadows, the comparisons are performed for geometries which guarantee that the complete domain is directly illuminated by the source.

Figure 3 shows the comparison of the numerical FDTD result for flat rigid terrain [Eqs. (1)–(4)] and the corresponding analytical solution of Eq. (12). The 50 m high source position (grid cell center exactly at $\tilde{z}_s=166.5\Delta z=49.95$ m; $f=50$ Hz) leads to well pronounced interference patterns [Fig. 3(a)]. The horizontal SPL profile at $\tilde{z}=30.15$ m shows that the positions of the numerically determined maxima and minima are slightly shifted [Fig. 3(b)]. However near the

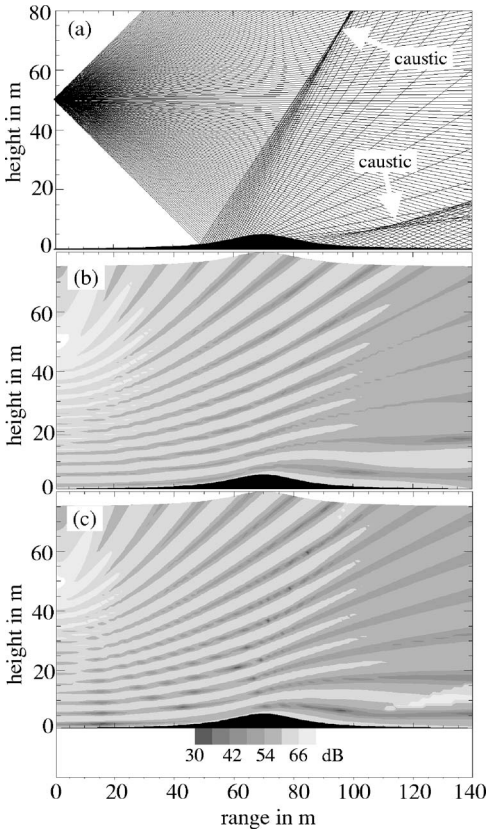


FIG. 4. SPL in dB of 2D sound propagation from a 50 m high source (50 Hz) over a 5 m high hill with rigid ground. (a) Sound rays emitted from the source toward elevation angles from −45° to +45° in steps of 0.5°, (b) numerical FDTD model simulation, (c) analytical solution.

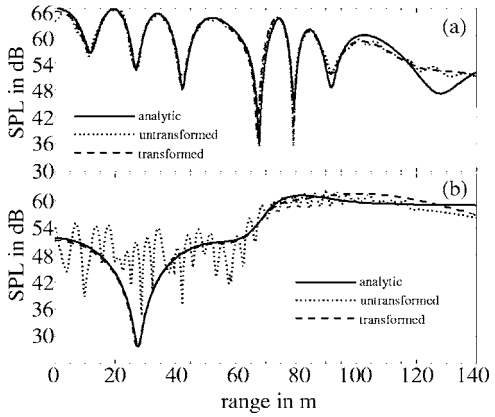


FIG. 5. Horizontal SPL profiles of 2D sound propagation from a 50 m high source (50 Hz) over a 5 m high hill with rigid ground. Results are shown for the analytic solution (solid lines) and for the numerical solutions with untransformed coordinates and blocked out terrain (dots) and transformed coordinates (dashes). The profiles refer to $\tilde{z}=30.15$ m (a) and $\tilde{z}=1.65$ m (b).

ground ($\tilde{z}=1.65$ m) both solutions are almost identical [Fig. 3(c)]. A point-by-point evaluation inside the area ($0\leqslant x\leqslant 140.1$ m; $0\leqslant\tilde{z}\leqslant 3$ m) results in a mean absolute deviation of 0.004 dB and a spatial correlation coefficient of 0.999. Inside the cross section shown in Fig. 3(a) ($0\leqslant x\leqslant 140.1$ m; $0\leqslant\tilde{z}\leqslant 75$ m) the corresponding values are 1.125 dB and 0.960.

For a comparison of a situation with hill and transformed coordinates [Eqs. (B1)–(B4)] a low hill [$h_{\text{top}}=5$ m; $\ell=15$ m; $x_0=70$ m in Eq. (11)] with rigid ground was assumed. The 50 m high source position guarantees that no part of the domain is shaded by the hill [Fig. 4(a)]. The hill causes a distinct distortion of the interference patterns [compare Fig. 4(b) with Fig. 3(a)] which are generally well captured by the numerical model [compare Fig. 4(b) with Fig. 4(c)]. The ray image in Fig. 4(a) shows the highly divergent rays which are reflected at the convex part of the hill top area. It also shows two caustics which are caused by the concave ground surface outside the hill top region. The ray divergence and the caustics are indicated in the analytical and numerical solutions by a fading of the interference patterns and weak secondary maxima, respectively.

For a more quantitative comparison horizontal profiles of the SPL at $\tilde{z}=1.65$ m and $\tilde{z}=30.15$ m above ground are shown in Figs. 5(a) and 5(b). The analytical solution and the solution for the transformed coordinates are in good agreement. Major deviations appear behind the hill for $x>90$ m where the numerical simulation smoothes the interference patterns more than the analytical calculation. The solution for untransformed coordinates with blocked out terrain is unrealistically distorted near the ground as a consequence of reflections at the stair-step like ground surface. At $\tilde{z}=30.15$ m both numerical solutions are of comparable quality. For $x>90$ m the agreement between the numerical solutions is better than that between the numerical and analytical ones. The deficiencies in this area are therefore not caused by the coordinate transformation and its numerical representation. Possible reasons are general numerical inaccuracies according to the choice of the numerical scheme, the time step, and the spatial increment, or the fact that the wave length of $\lambda=6.9$ m is in the order of the spatial scale of the hill and a purely specular reflection as assumed by the ray model is not given. However, this was not further investigated.

To show the improvements achieved by the terrain-following coordinates, quantitative quality measures were evaluated point by point inside the near-ground layer ($0\leqslant x\leqslant 140.1$ m; $0\leqslant\tilde{z}\leqslant 3$ m). Here, the mean absolute deviations of the SPL between the analytical solution and the numerical solutions based on untransformed and transformed coordinates amount to 1.76 and 0.69 dB, respectively. The corresponding spatial correlation coefficients are 0.86 and 0.98. Inside the cross section shown in Figs. 4(b) and 4(c) ($0\leqslant x\leqslant 140.1$ m; $0\leqslant\tilde{z}\leqslant 75$ m) the corresponding values are 1.21 and 1.03 dB for the mean absolute deviation and 0.93 and 0.96 for the correlation coefficient.

Further simulations were performed for a 20 m high hill [$h_{\text{top}}=20$ m; $\ell=15$ m; $x_0=70$ m in Eq. (11)], a source position on the ground (grid center exactly at $x_s=0.5\Delta x$; $\tilde{z}_s=0.5\Delta z$), 50 Hz sound and different assumptions with respect to ground reflectivity and wind: "rigid ground without wind," "absorbing ground without wind," "rigid ground with a logarithmic downwind profile," and "rigid ground with a logarithmic upwind profile." The absorbing ground parameters [$\tilde{\sigma}=50$ kPa s m^{-2}; $\tilde{\Omega}=0.2$; $\tilde{c}_s=0.1$ in Eqs. (B5)–(B8)] are typical of grass covered soil. Equations (B5)–(B8) are solved in a 48 cm deep model ground layer with a vertical grid interval of 6 cm. Only locally reacting ground is taken into account for simplicity, i.e., derivatives in x direction are

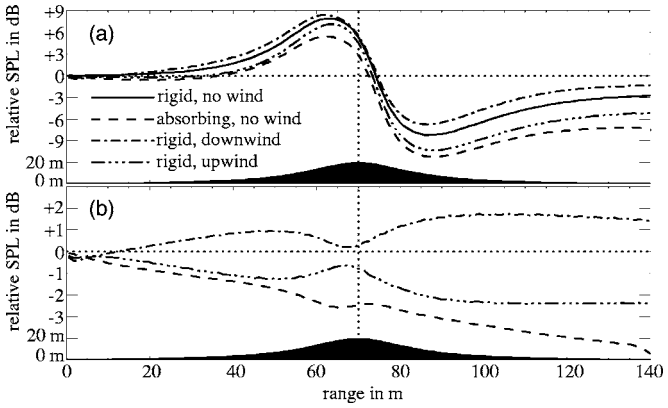


FIG. 6. Simulated SPL profiles at $\tilde{z}=1.65$ m for 2D propagation from a ground based source (50 Hz) over a 20 m high hill: (a) relative to the SPL of propagation over flat rigid ground without wind; (b) relative to the SPL of propagation over a 20 m high hill with rigid ground and without wind.

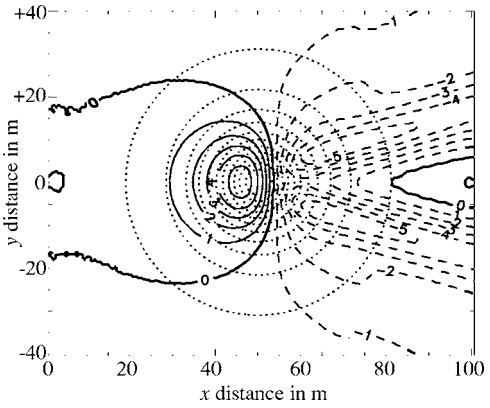


FIG. 7. SPL at $\tilde{z}=1.65$ m for 3D propagation from a ground based source (50 Hz) at $(x=0;y=0)$ over a 10 m high hill with rigid ground and without wind relative to the SPL of propagation over flat rigid ground (in dB). The dotted circles mark the elevation contours of the hill in 1/12 m intervals. The isopleths are based on every fourth grid point.

set to zero and a refinement of the horizontal grid interval is not necessary. The vertical profile of the horizontal wind component $\tilde{u}_{\mathrm{met}}(\tilde{z})=\pm 2.1668\ \mathrm{m\ s^{-1}}\ \ln[(\tilde{z}+\tilde{z}_0)/\tilde{z}_0]$ is calculated with an aerodynamic roughness length of $\tilde{z}_0=0.1$ m. This corresponds to a wind speed of 10 m/s at 10 m above ground. The transformed vertical wind component is set to $\tilde{w}_{\mathrm{met}}=0$ so that wind field is nondivergent and the air always flows parallel to the ground surface. Note that assumed flow over the hill is highly idealized because inertial and gravity effects are not considered. Since a symmetry boundary condition at $x=x_s-0.5\Delta x$ is not applicable in the case of wind, the model domain is extended in the horizontal direction so that the left boundary is situated at $x=-30$ m and a damping layer is invoked in the zone $-30\ \mathrm{m}\leqslant x\leqslant -2.4$ m.

The results of the FDTD model runs are shown in Fig. 6(a) as the SPL relative to propagation over flat, rigid ground for $\tilde{z}=1.65$ m. Generally, the SPL is increased by 5–9 dB in front of the hill due to multiple reflections by the convex slope profile. On the rear side the SPL drops by 7–11 dB. Figure 6(b) presents the SPL relative to the case with hill, rigid ground and without wind. The difference between rigid and absorbing ground (without wind) increases almost steadily with the distance from the source and amounts to more than 4 dB at $x=140$ m. The results for wind are almost symmetrically higher (downwind) or lower (upwind) as compared to the case without wind. Near the hill top the differences nearly vanish because here the direct rays have passed through layers where the vertical wind gradient is already significantly smaller than close to the ground. Behind the hill the diffracted sound grazes over the surface and the wind effect is strongly pronounced.

Finally, a 3D simulation was performed. The 3D grid comprises $400\times400\times200$ grid cells with a grid interval of $\Delta x=\Delta y=\Delta z=0.3$ m. The source is placed at the coordinate origin. It emits a continuous 50 Hz tone. Damping layers are employed at the top boundary and the lateral boundaries except that of $x=0$ where a symmetry condition is set. An axisymmetric Witch of Agnesi hill was assumed with the geometry $h_{\mathrm{top}}=10$ m; $\ell=10$ m; $x_0=50$ m; $y_0=0$ [in the 3D version of Eq. (11)] and rigid ground. Figure 7 presents the SPL at $\tilde{z}=1.65$ m relative to the result for plane and rigid ground. Again the SPL is increased in front of the hill with a maximum on the direct line between source and hill top. On the rear side the SPL is decreased as in the 2D case. However, it attains the value of the case without hill already at $x=80$ m (30 m behind the hill top). Although this area is in the shadow of the hill, the amplification is caused by the constructive interference of sound which is horizontally diffracted to the left- and right-hand side around the hill.

V. CONCLUSION AND DISCUSSION

A linearized Euler finite-difference time-domain model was numerically implemented on a transformed terrain-following coordinate grid to include the effect of gentle orography. The quality of the results is considerably improved with respect to simulations based on an untransformed grid with blocked-out orography, in particular near the ground where reflections caused by the stair-step representation of the orography are most pronounced. The comparison with an analytical solutions shows that the coordinate transformation is capable of reproducing the basic features of the orographical effect. A ground layer can be easily installed because it automatically follows the ground surface. Instead of a ground layer, also alternative methods (e.g. Heutschi *et al.*, 2005) could be implemented with transformed coordinates to incorporate the effects of finite impedance ground.

The model can be operated using a wind field obtained from a prognostic or diagnostic meteorological model. In this case full consistency between orography, wind (and temperature), and sound field can be taken into account. Such a coupling of meteorological and acoustical models as it was already performed for flat terrain with obstacles (e.g., by Blumrich and Heimann, 2002) requires an interpolation of the meteorological data to the acoustical model grid. In the case of non-flat terrain this interpolation is not trivial, in particular if the two models use different coordinate transformations. As already stated in Sec. III the transformation in meteorological models is often based on Eq. (9) whereas for the acoustical model the transformation rule of Eq. (10) is proposed here. It was therefore considered also to base the

acoustical model on the transformation rule of Eq. (9) in order to be more compatible with the meteorological models. However, this idea was rejected for two reasons. (1) A transformation with Eq. (9) results in a higher number of terms in the model equations. (2) Normally, meteorological microscale and mesoscale use larger grid intervals, often different spatial resolutions in horizontal and vertical direction, and their computational domains extend to higher altitudes. As a consequence an interpolation cannot be avoided even if the transformation rule is identical.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the detailed and comprehensive advices of two anonymous referees who substantially helped us to improve the manuscript, to eliminate errors, and to present the transformation of the model equations into terrain-following coordinates in a clearer way.

APPENDIX A

According to Blumrich and Heimann (2002) a set of differential equations based on a linearization of the Euler equations reads for sound propagation in dry air:

$$\frac{\partial \mathbf{u}}{\partial t} = -\mathbf{u}_{\mathrm{met}} \cdot \nabla \mathbf{u} - \mathbf{u} \cdot \nabla \mathbf{u}_{\mathrm{met}} - \frac{1}{\rho_{\mathrm{met}}} \nabla p + \frac{p}{\kappa p_{\mathrm{met}} \rho_{\mathrm{met}}} \nabla p_{\mathrm{met}}, \tag{A1}$$

$$\frac{\partial p}{\partial t} = -\mathbf{u}_{\mathrm{met}} \cdot \nabla p - \mathbf{u} \cdot \nabla p_{\mathrm{met}} - \kappa p_{\mathrm{met}} \nabla \cdot \mathbf{u} - \kappa p \nabla \cdot \mathbf{u}_{\mathrm{met}}. \tag{A2}$$

Besides all nonlinear effects also gravity, earth rotation and diabatic processes are neglected. The diffusion term in the prognostic equation of the particle velocity as it was introduced by Blumrich and Heimann (2002) is omitted here.

The magnitude and thus the relative importance of terms in the above equations can be assessed with the help of a 'scale analysis' (Burger, 1958). By this method the variables are assumed to be composed of characteristic values (scales) and dimensionless quantities near unity:

$$\mathbf{u} = U\mathbf{u}^*, \quad p = Pp^*, \quad \mathbf{u}_{\mathrm{met}} = V\mathbf{u}^*_{\mathrm{met}}, \quad \rho_{\mathrm{met}} = Q\rho^*_{\mathrm{met}},$$

$$p_{\mathrm{met}} = \bar{p}_{\mathrm{met}} + \delta p_{\mathrm{met}} = \bar{p}_{\mathrm{met}} + D\,\delta p^*_{\mathrm{met}}.$$

Here, capital letters are used for the scales whereas the asterisk indicates the dimensionless quantities of near unity magnitude. The air pressure p_{met} is split into a uniform background large-scale pressure $\bar{p}_{\mathrm{met}}$ and a small-scale deviation δp_{met} which may vary along the propagation path. The typical length of propagation paths in near-ground sound propagation (20 m–20 km) belongs to the meteorological microscale or mesoscale (cf. Orlanski, 1975). Therefore spatial pressure variations due to large-scale, synoptic pressure systems are not relevant in this case.

Characteristic time and length scales can be defined as: T=time scale of acoustic waves with the period $\tau = f^{-1} = T\tau^*$; S=length scale of acoustics waves with the wave length $\lambda = S\lambda^*$; and L=length scale of the relevant meteorological motions along the propagation path. S and T are linked because of $\lambda = cf^{-1}$, where c is the sound speed, i.e., $c = ST^{-1}c^*$. The time derivatives can be expressed as $\partial/\partial t = T^{-1}\partial/\partial t^*$, the space derivatives as $\nabla = S^{-1}\nabla^*$ for acoustical parameters, and $\nabla = L^{-1}\nabla^*$ for meteorological parameters. With these definitions and with $c^2\rho_{\mathrm{met}} = \kappa p_{\mathrm{met}}$, Eqs. (A1) and (A2) can be rewritten as

$$\left[\frac{U}{T}\right]\frac{\partial \mathbf{u}^*}{\partial t^*} = -\left[\frac{VU}{S}\right]\mathbf{u}^*_{\mathrm{met}} \cdot \nabla^* \mathbf{u}^* - \left[\frac{UV}{L}\right]\mathbf{u}^* \cdot \nabla^* \mathbf{u}^*_{\mathrm{met}} - \left[\frac{P}{QS}\right]\frac{1}{\rho^*_{\mathrm{met}}}\nabla^* p^* + \left[\frac{T^2PD}{S^2Q^2L}\right]\frac{p^*}{c^{*2}\rho^{*2}_{\mathrm{met}}}\nabla^* \delta p^*_{\mathrm{met}}, \tag{A3}$$

$$\left[\frac{P}{T}\right]\frac{\partial p^*}{\partial t^*} = -\left[\frac{VP}{S}\right]\mathbf{u}^*_{\mathrm{met}} \cdot \nabla^* p^* - \left[\frac{UD}{L}\right]\mathbf{u}^* \cdot \nabla^* \delta p^*_{\mathrm{met}} - \left[\frac{SQU}{T^2}\right]c^{*2}\rho^*_{\mathrm{met}}\nabla^* \cdot \mathbf{u}^* - \left[\frac{PV}{L}\right]\kappa p^* \nabla \cdot \mathbf{u}^*_{\mathrm{met}}. \tag{A4}$$

For typical problems in near-ground outdoor sound propagation the range of magnitude of the characteristic scales is given by $T = 10^{-4}$–10^{-2} s (for frequencies in the order from 100 Hz to 10 kHz); $S = 10^{-2}$–1 m (for the corresponding wavelengths); $P = 10^{-2}$–10^{-1} Pa (for sound pressure levels between 60 and 80 dB); $Q = 1$ kg m^{-3} (typical air density near the ground); $U = PTQ^{-1}S^{-1} = 10^{-1}$–$10^{-6}$ m s^{-1} (because of the relation $u_a = p_a c^{-1}\rho^{-1}_{\mathrm{met}}$ with u_a and p_a denoting the amplitudes of $\mathbf{u}$ and p, respectively); $L = 10^1$–10^2 m (typical length scale of atmospheric structures in the propagation path); $D = 10^1$ Pa (typical meteorological small-scale pressure variations); and $V = 1$–10^1 m s^{-1} (typical wind speeds near the ground).

It is convenient to divide Eq. (A3) by (UT^{-1}) and Eq. (A4) by (PT^{-1}). With $M = VTS^{-1}$ (the Mach number of the air flow) we get the following magnitudes of the terms in Eq. (A1).

left side	1
right side, first term	$M = 10^{-2}$–10^{-1}
right side, second term	$VTL^{-1} = 10^{-6}$–10^{-2}
right side, third term	1
right side, fourth term	$T^2DQ^{-1}S^{-1}L^{-1} = 10^{-7}$–$10^{-4}$

and in Eq. (A2):

left side	1
right side, first term	$M = 10^{-2}$–10^{-1}
right side, second term	$T^2DQ^{-1}S^{-1}L^{-1} = 10^{-7}$–$10^{-4}$
right side, third term	1
right side, fourth term	$VTL^{-1} = 10^{-6}$–10^{-2}

For these magnitudes it turns out that the left-hand sides of Eqs. (A3) and (A4) are mainly balanced by the third terms of the right-hand sides as the governing terms. The second and fourth terms of the right sides of these equations are always at least two orders of magnitude smaller. The first terms of the right-hand sides are at least one order of magnitude smaller than the governing terms but usually larger than the second and fourth terms.

According to the scale analysis it can be concluded that the second and fourth terms of the right-hand side of Eqs. (A1) and (A2) can be neglected in most cases. However, the negligence can be critical for low frequencies (large magnitude of T) in combination with strong meteorological gradients as they appear in heavy turbulence (large magnitude of VL^{-1} and DL^{-1}). The fourth term of the right side of Eq. (A2) always vanishes if the air flow is nondivergent. In a calm atmosphere ($\mathbf{u}_{\mathrm{met}}=0$) the first and second term in Eq. (A1) and the first and fourth term of Eq. (A2) vanish in any case.

Neglect of the second and fourth terms of the right-hand sides of Eqs. (A1) and (A2) leads to the approximated equations which are used in this paper:

$$\frac{\partial \mathbf{u}}{\partial t} = -\mathbf{u}_{\mathrm{met}} \cdot \nabla \mathbf{u} - \frac{1}{\rho_{\mathrm{met}}} \nabla p, \tag{A5}$$

$$\frac{\partial p}{\partial t} = -\mathbf{u}_{\mathrm{met}} \cdot \nabla p - \kappa p_{\mathrm{met}} \nabla \cdot \mathbf{u}. \tag{A6}$$

APPENDIX B

In this appendix the relationship between the original and transformed coordinate system is stepwise deduced for the spatial coordinates, the dependent model variables and their spatial and temporal derivatives. Finally, the terms of the untransformed model equations in the air [Eqs. (1)–(4)] and in the ground [Eqs. (5)–(8)] are replaced by those referring to the transformed terrain-following coordinates.

(a) Transformation of the spatial coordinates: The transformation rule of the spatial coordinates is given by Eq. (10) with $x=\tilde{x}$, $y=\tilde{y}$, and $z=\tilde{z}+h(x,y)$, where $\tilde{x}$, $\tilde{y}$, and $\tilde{z}$ are the terrain-following coordinates. This rule implies that $\tilde{x}=x$, $\tilde{y}=y$, $\tilde{z}=z-h(x,y)$.

(b) Transformation of the dependent model variables: The relationships between untransformed and transformed particle velocity components follow from applying the time derivation to the spatial coordinates and using the chain rule

$$u(x,y,z,t) = \frac{dx}{dt} = \frac{d\tilde{x}}{dt} = \tilde{u}(\tilde{x},\tilde{y},\tilde{z},t),$$

$$v(x,y,z,t) = \frac{dy}{dt} = \frac{d\tilde{y}}{dt} = \tilde{v}(\tilde{x},\tilde{y},\tilde{z},t),$$

$$w(x,y,z,t) = \frac{dz}{dt} = \frac{d}{dt}(\tilde{z}+h(x,y))$$

$$= \tilde{w}(\tilde{x},\tilde{y},\tilde{z},t) + \frac{\partial h}{\partial x}\tilde{u}(\tilde{x},\tilde{y},\tilde{z},t) + \frac{\partial h}{\partial y}\tilde{v}(\tilde{x},\tilde{y},\tilde{z},t),$$

and for the sound pressure

$$p(x,y,z,t) = \tilde{p}(\tilde{x},\tilde{y},\tilde{z},t).$$

(c) Transformation of the medium (air and ground) parameters: Analogously to (b) one gets

$$u_{\mathrm{met}} = \tilde{u}_{\mathrm{met}}, \quad v_{\mathrm{met}} = \tilde{v}_{\mathrm{met}}, \quad w_{\mathrm{met}} = \tilde{w}_{\mathrm{met}} + \frac{\partial h}{\partial x}\tilde{u}_{\mathrm{met}} + \frac{\partial h}{\partial y}\tilde{v}_{\mathrm{met}},$$

$$p_{\mathrm{met}} = \tilde{p}_{\mathrm{met}}, \quad \rho_{\mathrm{met}} = \tilde{\rho}_{\mathrm{met}}, \quad \Omega = \tilde{\Omega}, \quad c_s = \tilde{c}_s, \quad \sigma = \tilde{\sigma}.$$

The dimensionless quantity κ is independent of the coordinate system.

(d) Transformation of the time derivatives of the dependent model variables: The time derivation applied to relationships in (b) leads to

$$\frac{\partial u}{\partial t} = \frac{\partial \tilde{u}}{\partial t}, \quad \frac{\partial v}{\partial t} = \frac{\partial \tilde{v}}{\partial t}, \quad \frac{\partial w}{\partial t} = \frac{\partial \tilde{w}}{\partial t} + \frac{\partial h}{\partial x}\frac{\partial \tilde{u}}{\partial t} + \frac{\partial h}{\partial y}\frac{\partial \tilde{v}}{\partial t},$$

$$\frac{\partial p}{\partial t} = \frac{\partial \tilde{p}}{\partial t}.$$

(e) Transformation of the spatial derivatives of the dependent model variables: The spatial derivation is applied to relationships in (b) using the chain rule and the relationship $\tilde{z}=z-h(x,y)$ in (a). For example, $\partial/\partial x$ applied to $u(x,y,z,t)=\tilde{u}(\tilde{x},\tilde{y},\tilde{z},t)$ results in

$$\frac{\partial u}{\partial x} = \frac{\partial \tilde{u}}{\partial \tilde{x}}\frac{\partial \tilde{x}}{\partial x} + \frac{\partial \tilde{u}}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial x} + \frac{\partial \tilde{u}}{\partial \tilde{z}}\frac{\partial \tilde{z}}{\partial x} = \frac{\partial \tilde{u}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{u}}{\partial \tilde{z}},$$

where $\partial\tilde{x}/\partial x=1$, $\partial\tilde{y}/\partial x=0$, $\partial\tilde{z}/\partial x=\partial z/\partial x-\partial h/\partial x$ with $\partial z/\partial x=0$.

Analogously, the following expressions are obtained for all spatial derivatives:

$$\frac{\partial u}{\partial x} = \frac{\partial \tilde{u}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{u}}{\partial \tilde{z}}, \quad \frac{\partial u}{\partial y} = \frac{\partial \tilde{u}}{\partial \tilde{y}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{u}}{\partial \tilde{z}}, \quad \frac{\partial u}{\partial z} = \frac{\partial \tilde{u}}{\partial \tilde{z}},$$

$$\frac{\partial v}{\partial x} = \frac{\partial \tilde{v}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{v}}{\partial \tilde{z}}, \quad \frac{\partial v}{\partial y} = \frac{\partial \tilde{v}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{v}}{\partial \tilde{z}}, \quad \frac{\partial v}{\partial z} = \frac{\partial \tilde{v}}{\partial \tilde{z}},$$

$$\frac{\partial w}{\partial x} = \left[\frac{\partial \tilde{w}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{w}}{\partial \tilde{z}}\right] + \frac{\partial h}{\partial x}\left[\frac{\partial \tilde{u}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{u}}{\partial \tilde{z}}\right] + \frac{\partial h}{\partial y}\left[\frac{\partial \tilde{v}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{v}}{\partial \tilde{z}}\right] + \frac{\partial^2 h}{\partial x^2}\tilde{u} + \frac{\partial^2 h}{\partial x \partial y}\tilde{v},$$

$$\frac{\partial w}{\partial y} = \left[\frac{\partial \tilde{w}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{w}}{\partial \tilde{z}}\right] + \frac{\partial h}{\partial x}\left[\frac{\partial \tilde{u}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{u}}{\partial \tilde{z}}\right] + \frac{\partial h}{\partial y}\left[\frac{\partial \tilde{v}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{v}}{\partial \tilde{z}}\right] + \frac{\partial^2 h}{\partial x \partial y}\tilde{u} + \frac{\partial^2 h}{\partial y^2}\tilde{v},$$

$$\frac{\partial w}{\partial z} = \frac{\partial \tilde{w}}{\partial \tilde{z}} + \frac{\partial h}{\partial x}\frac{\partial \tilde{u}}{\partial \tilde{z}} + \frac{\partial h}{\partial y}\frac{\partial \tilde{v}}{\partial \tilde{z}},$$

$$\frac{\partial p}{\partial x} = \frac{\partial \tilde{p}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{p}}{\partial \tilde{z}}, \quad \frac{\partial p}{\partial y} = \frac{\partial \tilde{p}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{p}}{\partial \tilde{z}}, \quad \frac{\partial p}{\partial z} = \frac{\partial \tilde{p}}{\partial \tilde{z}}.$$

(f) Transformed model equations (air): The expressions of (b)–(e) are now substituted into the untransformed model Eqs. (1)–(4). This leads to the model equations in the air expressed in the terrain following coordinate system:

$$\frac{\partial \tilde{u}}{\partial t} = -\tilde{u}_{\mathrm{met}}\frac{\partial \tilde{u}}{\partial \tilde{x}} - \tilde{v}_{\mathrm{met}}\frac{\partial \tilde{u}}{\partial \tilde{y}} - \tilde{w}_{\mathrm{met}}\frac{\partial \tilde{u}}{\partial \tilde{z}} - \frac{1}{\tilde{\rho}_{\mathrm{met}}}\left[\frac{\partial \tilde{p}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{p}}{\partial \tilde{z}}\right], \tag{B1}$$

$$\frac{\partial \tilde{v}}{\partial t} = -\tilde{u}_{\mathrm{met}}\frac{\partial \tilde{v}}{\partial \tilde{x}} - \tilde{v}_{\mathrm{met}}\frac{\partial \tilde{v}}{\partial \tilde{y}} - \tilde{w}_{\mathrm{met}}\frac{\partial \tilde{v}}{\partial \tilde{z}} - \frac{1}{\tilde{\rho}_{\mathrm{met}}}\left[\frac{\partial \tilde{p}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{p}}{\partial \tilde{z}}\right], \tag{B2}$$

$$\frac{\partial \tilde{w}}{\partial t} = -\tilde{u}_{\mathrm{met}}\frac{\partial \tilde{w}}{\partial \tilde{x}} - \tilde{v}_{\mathrm{met}}\frac{\partial \tilde{w}}{\partial \tilde{y}} - \tilde{w}_{\mathrm{met}}\frac{\partial \tilde{w}}{\partial \tilde{z}} + \frac{1}{\tilde{\rho}_{\mathrm{met}}}\left[\frac{\partial h}{\partial x}\frac{\partial \tilde{p}}{\partial \tilde{x}} + \frac{\partial h}{\partial y}\frac{\partial \tilde{p}}{\partial \tilde{y}} - \left(\left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2 + 1\right)\frac{\partial \tilde{p}}{\partial \tilde{z}}\right] - \left[\frac{\partial^2 h}{\partial x^2}\tilde{u}_{\mathrm{met}} + \frac{\partial^2 h}{\partial x \partial y}\tilde{v}_{\mathrm{met}}\right]\tilde{u} - \left[\frac{\partial^2 h}{\partial x \partial y}\tilde{u}_{\mathrm{met}} + \frac{\partial^2 h}{\partial y^2}\tilde{v}_{\mathrm{met}}\right]\tilde{v}, \tag{B3}$$

$$\frac{\partial \tilde{p}}{\partial t} = -\tilde{u}_{\mathrm{met}}\frac{\partial \tilde{p}}{\partial \tilde{x}} - \tilde{v}_{\mathrm{met}}\frac{\partial \tilde{p}}{\partial \tilde{y}} - \tilde{w}_{\mathrm{met}}\frac{\partial \tilde{p}}{\partial \tilde{z}} - \kappa \tilde{p}_{\mathrm{met}}\left[\frac{\partial \tilde{u}}{\partial \tilde{x}} + \frac{\partial \tilde{v}}{\partial \tilde{y}} + \frac{\partial \tilde{w}}{\partial \tilde{z}}\right]. \tag{B4}$$

(g) Transformed model equations (ground): The expressions of (b)–(e) are now substituted into the untransformed model Eqs. (5)–(8). This leads to the model equations in the ground expressed in the terrain following coordinate system:

$$\frac{\partial \tilde{u}}{\partial t} = -\frac{\tilde{\Omega}}{\tilde{c}_s \tilde{\rho}_{\mathrm{met}}}\left[\frac{\partial \tilde{p}}{\partial \tilde{x}} - \frac{\partial h}{\partial x}\frac{\partial \tilde{p}}{\partial \tilde{z}} + \tilde{\sigma}\tilde{u}\right], \tag{B5}$$

$$\frac{\partial \tilde{v}}{\partial t} = -\frac{\tilde{\Omega}}{\tilde{c}_s \tilde{\rho}_{\mathrm{met}}}\left[\frac{\partial \tilde{p}}{\partial \tilde{y}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{p}}{\partial \tilde{z}} + \tilde{\sigma}\tilde{v}\right], \tag{B6}$$

$$\frac{\partial \tilde{w}}{\partial t} = -\frac{\tilde{\Omega}}{\tilde{c}_s \tilde{\rho}_{\mathrm{met}}}\left[-\frac{\partial h}{\partial x}\frac{\partial \tilde{p}}{\partial \tilde{x}} - \frac{\partial h}{\partial y}\frac{\partial \tilde{p}}{\partial \tilde{y}} + \left(\left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2 + 1\right)\frac{\partial \tilde{p}}{\partial \tilde{z}} + \tilde{\sigma}\tilde{w}\right], \tag{B7}$$

$$\frac{\partial \tilde{p}}{\partial t} = -\frac{\kappa p_{\mathrm{met}}}{\tilde{\Omega}}\left[\frac{\partial \tilde{u}}{\partial \tilde{x}} + \frac{\partial \tilde{v}}{\partial \tilde{y}} + \frac{\partial \tilde{w}}{\partial \tilde{z}}\right]. \tag{B8}$$

(h) Consequences for the numerical solution: The numerical stability of the model solution with an explicit numerical scheme is controlled by the Courant-Friedrichs-Lewy (CFL) number $\mathrm{CFL}=\Delta s(\Delta t c_{\max})^{-1}$, where Δs is the smallest extent of a grid cell, Δt is the time step, and $c_{\max}$ is the fastest signal velocity which in our case is the sum of the maximum speed of the sound and the maximum wind speed. An appropriate upper limit of the CFL number ($\mathrm{CFL} \leqslant 2^{-0.5}$ in 2D, $\mathrm{CFL} \leqslant 3^{-0.5}$ in 3D) ensures that a signal is not transported farther than the smallest extent of a grid cell during one time step.

In a terrain-following isotropic grid ($\Delta\tilde{x}=\Delta\tilde{y}=\Delta\tilde{z}$) the smallest extent of a grid cell is given by the shortest distance between the model levels. In the transformed system the distance between the model levels is given by $\Delta s=\Delta\tilde{z}[|\nabla h|^2+1]^{-0.5}$. Although the spatial increments $\Delta\tilde{x}$, $\Delta\tilde{y}$, and $\Delta\tilde{z}$ are constant, Δs decreases above slopes as illustrated in Fig. 1(c). Therefore, the allowed maximum time step must decrease with increasing steepness of the terrain. For a maximum steepness of $|\nabla h|=1$ the time step has to be decreased by a factor of $2^{-0.5}=0.707$ relative to that for flat terrain. This factor reduces to 0.0995 for a steepness of $|\nabla h|=10$ so that the number of time steps and thus the computational effort increases by more than a factor of 10.

Arakawa, A., and Lamb, V. R. (**1977**). "Computational design of the basic dynamic processes of the UCLA general circulation model," Methods Comput. Phys. **17**, 173–265.

Blumrich, R., and Heimann, D. (**2002**). "A linearized Eulerian sound propagation model for studies of complex meteorological effects," J. Acoust. Soc. Am. **112**, 446–455.

Blumrich, R., and Heimann, D. (**2004**). "Numerical estimation of atmospheric approximation effects in outdoor sound propagation modelling," Acust. Acta Acust. 90, 24–37.

Botteldooren, D. (**1994**). "Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid," J. Acoust. Soc. Am. **95**, 2313–2319.

Burger, A. (**1958**). "Scale considerations of Planetary motions of the atmosphere," Tellus **10**, 195–205.

Gal-Chen, T., and Sommerville, R. C. J. (**1975**). "On the use of a coordinate transformation for the solution of the Navier-Stokes Equations," J. Comput. Phys. **17**, 209–228.

Heimann, D. (**2003**). "Numerical simulations of wind and sound propagation through an idealised stand of trees," Acust. Acta Acust. 89, 779–788.

Heimann, D., and Blumrich, R. (**2004**). "Time-domain simulations of sound propagation through screen-induced turbulence," Acust. Acta Acust. 65, 561–582.

Hestholm, S. (**1999**). "Three-dimensional finite difference viscoelastic wave modelling including surface topography," Geophys. J. Int. **139**, 852–878.

Hestholm, S., and Ruud, B. (**1994**). "2D finite-difference elastic wave modelling including surface topography," Geophys. Prospect. **42**, 371–390.

Heutschi, K., Horvath, M., and Hofmann, J. (**2005**). "Simulation of ground impedance in finite difference time domain calculations of outdoor sound propagation," Acust. Acta Acust. 91, 35–40.

Mahrer, Y., and Pielke, R. A. (**1975**). "The numerical study of the airflow over mountains using the University of Virginia mesoscale model," J. Atmos. Sci. **32**, 2144–2155.

Orlanski, I. (**1975**). "A rational subdivision of scales for atmospheric processes," Bull. Am. Meteorol. Soc. **56**, 527–530.

Ostashev, V. E., Wilson, D. K., Liu, L., Aldridge, D. F., Symons, N. P., and Marlin, D. (**2005**). "Equations for finite-difference, time-domain simulations of sound propagation in moving inhomogeneous media and numerical implementation," J. Acoust. Soc. Am. **117**, 503–517.

Phillips, N. A. (**1957**). "A coordinate system having some special advantages for numerical forecasting," J. Meteorol. **14**, 184–185.

Pielke, R. A. (**2002**). *Mesoscale Meteorological Modeling*, 2nd ed. (Academic Press, San Diego).

Pierce, A. D. (**1981**). *Acoustics—An Introduction to Its Principles and Applications* (McGraw-Hill, New York).

Sack, R. A., and West, M. (**1995**). "A parabolic equation for sound propagation in two dimensions over any smooth terrain profile: The generalised terrain parabolic equation (GT-PE)," Appl. Acoust. **45**, 113–129.

Salomons, E., Blumrich, R., and Heimann, D. (**2002**). "Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain models," Acust. Acta Acust. 88, 483–492.

Tolan, J. G., and Schneider, J. B. (**2003**). "Locally conformal method for acoustic finite-difference time-domain modeling of rigid surfaces." J. Atmos. Sci. **114**, 2575–2581.

Van Renterghem, T., and Botteldoren, D. (**2003**). "Numerical simulation of the effect of trees on downwind noise barrier performance," Acust. Acta Acust. 89, 764–778.

Van Renterghem, T., Salomons, E., and Botteldooren, D. (**2005**). "Efficient FDTD-PE model for sound propagation in situations with complex obstacles and wind profiles," Acust. Acta Acust. 91, 671–679.

Van Renterghem, T., Salomons, E., and Botteldooren, D. (**2006**). "Parameter study of sound propagation between city canyons with a coupled FDTD-PE model," Appl. Acoust. **67**, 487–510.

Zwikker, C., and Kosten, C. W. (**1949**). *Sound Absorbing Materials* (Elsevier, New York).

Experimental methods for investigating the acoustical interaction between transducers

Boris Aronov
Acoustic Research Laboratory, Advanced Technology and Manufacturing Center and Department of Electrical and Computer Engineering, the University of Massachusetts, Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723 and BTech Acoustics, LLC, 1445 Wampanoag Trail, Suite 115, East Providence, Rhode Island 02915

(Received 17 December 2005; revised 26 March 2006; accepted 28 March 2006)

Experimental methods for investigating the mutual radiation impedances in arrays are considered. One method, called the *Z* method, is based on measurements of the input impedance of the interacting transducers. The second method, called the *V* method, is based on measurements of the acoustomotive force generated on the surface of one of the transducers by an interacting transducer. This method can be used to investigate the mutual impedances between transducers at frequencies far below their resonant frequencies. Theoretical foundations for the methods are discussed and the experimental setups for implementing the methods are described. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2198181]

PACS number(s): 43.58.Bh, 43.38.Fx, 43.38.Pf [AJZ] Pages: 3822–3830

I. INTRODUCTION

Considering the effects of acoustical interaction between transducers populating acoustic underwater arrays is of a great importance for predicting their real performance. Many papers are devoted to the investigation of acoustical interactions. A vast bibliography on the related issues can be found in Ref. 1. For some idealized, yet important, configurations of arrays and transducer radiating surfaces the analytical solutions to the problem of acoustic interaction are obtained.[2–6] In all of these array configurations the radiating surfaces of the transducers were assumed to be flush with the surface of an array. The solutions obtained revealed some general properties of the interaction phenomena and made it possible to obtain a rather simple approximate analytical estimation for the mutual radiation impedances in these cases. However, for real arrays the most practical way to determine mutual impedances between transducers is through experimentation. This especially applies to those populated by transducers having a nonplanar configuration and those furnished with special baffles, which in particular can be intended to change an effect of interaction. Nonetheless, the methods of an experimental investigation of interaction between transducers, and especially between transducers in arrays, have not obtained adequate attention. A number of papers report on the measurement results of the mutual impedances (see, for example, Refs. 1, 7, and 8), but none of them describe the experimental technique and procedures in detail. At the same time, certain peculiarities are inherent in the procedure of measuring mutual impedances and special precautions have to be made in order to get accurate results.

The objective of the paper is to consider experimental methods of measuring the mutual radiation impedances between transducers in an array. Two somewhat different although complementary methods, which are called the *Z* method and the *V* method, are considered. Theoretically they are based on the equivalent circuit representation of a transducer operation, and these methods are introduced in Sec. II. In Sec. III procedures for both methods to measure the mutual impedance between the two transducers vibrating in a free space are illustrated. In Sec. IV peculiarities of measuring the mutual impedances in an array made of three and more transducers are considered. This paper presents the methodology of ongoing theoretical and experimental investigations, preliminary results of which were reported in part at the 143rd ASA meeting.[9] The experimental results of the investigations are in progress and beyond the scope of this paper.

II. MUTUAL RADIATION IMPEDANCE

In an *N*-element array the total radiation impedance of the *i*th transducer, Z_i, can be represented as[1]

$$Z_{aci} = Z_{acii} + \sum_{l \neq i}^{N} Z_{acil}, \tag{1}$$

where Z_{acii} is the self-radiation impedance of the transducer and Z_{acil} are the coupled radiation impedances, which characterize the acoustical coupling between transducers in the array. The coupled radiation impedance can be expressed as

$$Z_{acil} = z_{acil} U_l / U_i, \tag{2}$$

where U_i and U_l are the surface velocities and z_{acil} is the mutual radiation impedance of the *i*th and *l*th transducers. The mutual radiation impedance is inherent in the transducer and array geometry, whereas the coupled impedance depends also on the relative velocities of the transducers vibration, i.e., on a mode of the array operation. The mutual radiation impedance depends only on the separation between transducers, when the transducer surface is flush with the surface of an array, as it is shown in Figs. 1(a) and 1(b), or otherwise on a position of the transducer in array, as it is illustrated in Fig. 1(c). The goal of an experimental investigation is to determine the mutual impedances between transducers for a particular transducer and array configuration.

 0001-4966/2006/119(6)/3822/9/$22.50

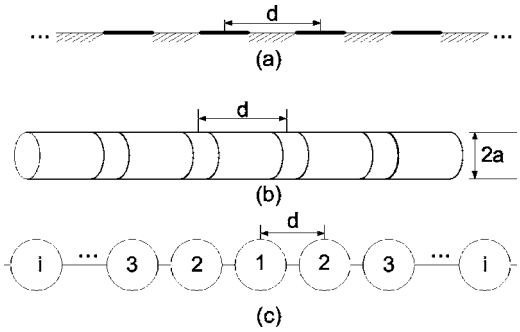


FIG. 1. Configurations of the transducers and arrays: (a) the transducer surface is flush with a rigid plane baffle,[2,3] (b) finite height cylinders in a rigid cylindrical baffle,[4] and (c) array of cylindrical transducers having parallel axes.

Using an equivalent circuit representation of a transducer operating in the transmit mode,[10] the effect of interaction between transducers in an array can be taken into account by introducing the coupled impedances $Z_{ac12}=z_{ac12}U_2/U_1$ into the circuits of interacting transducers, or by introducing the "acoustomotive force" $F_{ac12}=z_{ac12}U_2$ instead of coupled impedance in the circuit for one of transducers, as it is shown in Figs. 2(a) and 2(b) correspondingly (for simplicity only the two-transducer array is considered). The equivalence of both representations follows from the fact that the same acoustic power is associated with the coupled impedance and with the acoustomotive force in the equivalent circuit of transducer 1, namely,

$$\dot{\bar{W}}_{ac12}=Z_{ac12}|U_1|^2=F_{ac12}U_1^*=z_{ac12}U_2U_1^*.$$

It is clear from the equivalent circuit of the transducers in the form of Fig. 2(a) that the mutual impedances can be found by measuring the input impedances of the interacting transducers, a part of which the coupled impedances constitute (Z method).

In an equivalent circuit in Fig. 2(b) the acoustic interaction is represented by force F_{12}, which is produced by one of the transducers operating in the transmit mode on the surface of another transducer operating in the receive mode. By measuring the output voltage of the receiving transducer relative to the input voltage of the radiating transducer the acting force $F_{12}=z_{12}U_2$ and the mutual impedance z_{12} can be evaluated (V method).

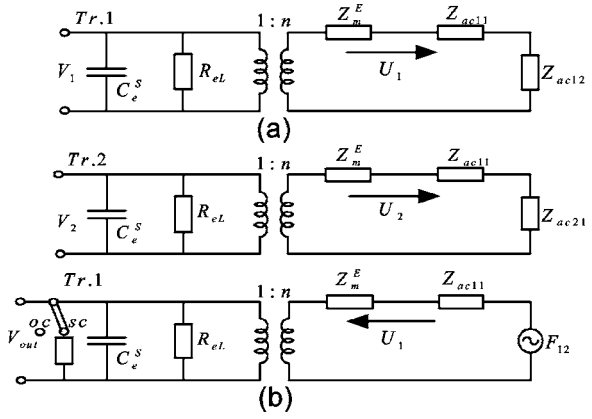


FIG. 2. The equivalent circuits for application of the experimental methods for investigating the mutual radiation impedances: (a) the circuit of transducer 1 in the case that the Z method is considered and (b) the circuit of transducer 1 in the case that the V method is considered. $Z_m^E=j\omega M_{eqv}+1/j\omega C_m^E+r_{mL}$.

Note several peculiarities of the impedances intended to be measured. The self-radiation impedances of the transducers in an array, Z_{acmm}, and the coupled impedances between any two transducers, Z_{acmn}, are supposed to be determined under the condition that all of the other transducers, with which effect of interaction cannot be neglected, are clamped ($U_l=0$ at $l\neq m,n$). This follows directly from expression (1). In accordance with the reciprocity principle $z_{acil}=z_{acli}$. Assuming that the mutual and coupled impedances are represented as $z_{acil}=r_{acil}+jx_{acil}$ and $Z_{acil}=R_{acil}+jX_{acil}$, where r_{acil}, R_{acil} and x_{acil}, X_{acil} are the active and reactive components of the impedances, according to expression (2) we obtain

$$R_{acil}=r_{acil}\mathrm{Re}\{U_l/U_i\}-x_{acil}\mathrm{Im}\{U_l/U_i\}, \tag{3}$$

$$X_{acil}=x_{acil}\mathrm{Re}\{U_l/U_i\}+r_{acil}\mathrm{Im}\{U_l/U_i\}. \tag{4}$$

Usually the active component of the total radiation impedance and therefore the active component of the coupled impedance are of a particular interest in the process of designing a transducer, due to their effect on the radiation of acoustic energy. But in the case that velocities of the vibration of transducers in the array differ in phase (and the differences may change because of steering), both the active and reactive components of the mutual radiation impedance have to be known in order to determine the total radiation resistance.

General properties of the mutual radiation impedances may be illustrated with examples of the acoustic interaction between two transducers, for which the theoretical solutions are available. Considering these examples is also worthwhile because the results obtained for such systems by the experimental methods can be compared with theoretical expectations and thus an accuracy of measurement can be roughly estimated. A well-known expression for the normalized mutual impedance between two simple sources of equal strength in a free space[2] is

$$z_{ac12}/R_{ac11}=e^{jkd}/kd=\sin kd/kd+j\cos kd/kd. \tag{5}$$

Here R_{ac11} is the self-radiation resistance, which for a simple source (practically for a transducer of a small wave size) may be expressed as $R_{ac11}=\pi\rho c(S_{av}/\lambda)^2$. Parameters ρ and c are the density and sound speed in water, λ is the wavelength, S_{av} is the average area[10] of the radiating surface of the transducer, $k=2\pi/\lambda$, and d is the separation between the acoustical centers of the transducers. Dependence (5) is depicted in Fig. 3.

A remarkable property of formula (5) is that it proves to be valid to a great accuracy for the transducers having radiating surfaces with comparable lateral dimensions in arrays of different configurations, in which the mutual impedance depends on separation between transducers only [arrays of the type shown in Figs. 1(a) and 1(b)], even if the dimensions of their radiating surfaces are not very small compared with a wavelength. In these cases only the self-radiation impedance of the actual transducer in the array has to be used

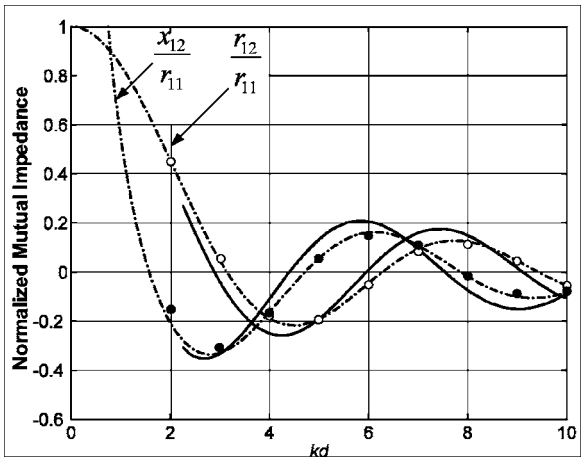


FIG. 3. A typical behavior of the normalized mutual impedances for the transducers flush with the array surfaces: simple sources [formula (5)] (dashed lines), an array of circular pistons with $2a/\lambda=0.3$ in the rigid plane[2] (circles), and a coaxial array of finite cylinders with $h/\lambda=0.36$ and $2a/\lambda=0.73$ in a rigid cylindrical baffle[4] (solid line).

as the normalizing resistance R_{ac11} in formula (5). Thus, the results of calculation of the normalized mutual impedances for the circular pistons vibrating in a rigid plane[2] and for the rings vibrating in a rigid cylindrical baffle,[4] which are plotted in Fig. 3, show practically complete agreement with those obtained by formula (5), although the dimensions of the circular pistons were $2a\cong 0.3\lambda$ and dimensions of the rings were $2a\cong 0.73\lambda$ and $h\cong 0.36\lambda$.

In the case that the length of the transducers is much larger than their lateral dimensions, the following formula,

$$z_{ac12}/R_{ac11} = J_0(kd) - jN_0(kd), \tag{6}$$

where $J_0(kd)$ and $N_0(kd)$ are the Bessel functions of the first and the second kind, may be used as the two-dimensional analog of formula (5) to approximate the mutual impedances of transducers. This formula at $R_{ac11}=0.5\ \rho c(2\pi a)^2/\lambda$ describes the normalized mutual impedance per unit length of the infinitely long cylindrical radiators of small radius ($ka \ll 1$).

From the above examples the conclusion can be made that the modulus of the mutual impedance drops with separation between acoustical centers of the transducers approximately as $|z_{ac12}/R_{ac11}|=1/kd=\lambda/2\pi d$, which follows from formula (5).

III. MEASUREMENT OF THE MUTUAL IMPEDANCE BETWEEN TWO TRANSDUCERS IN THE FREE FIELD

A. *Z* Method

Consider two simultaneously operating transducers. Their equivalent circuits in general can be represented as shown in Fig. 2(a). Assume that the transducers are electromechanically identical. All the electromechanical parameters of the transducers including internal mechanical impedance Z_m^E may be considered as known. Otherwise, they can be determined by common measurements performed on an unloaded transducer (in air).[11] The equivalent electromechanical circuit of each transducer may be reduced to the form

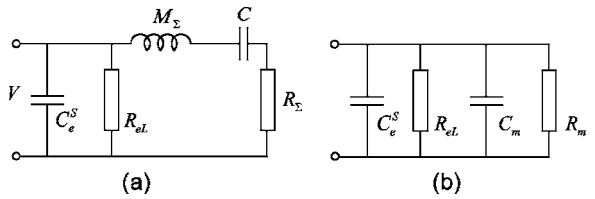


FIG. 4. Modifications of the equivalent circuit of interacting transducers with the mechanical branch transformed into electrical side: (a) with series motional impedance and (b) after converting the motional impedance into admittance.

shown in Fig. 4(a), where the mechanical branch is transformed onto the electrical side. In Fig. 4(a) $C=C_m^E n^2$,

$$R_{\Sigma 1} = R_{11} + R_{12} + R_{mL} \quad \text{and} \quad M_{\Sigma 1} = M + X_1/\omega \tag{7}$$

for transducer 1, and

$$R_{\Sigma 2} = R_{11} + R_{21} + R_{mL} \quad \text{and} \quad M_{\Sigma 2} = M + X_2/\omega \tag{8}$$

for transducer 2. In formulas (7) and (8)

$$X_1 = X_{11} + X_{12}, \quad X_2 = X_{11} + X_{21}. \tag{9}$$

We denote the electrical analogs of the acoustical impedances in the same way as their acoustical counterparts but without the subscript "*ac*." Thus, for example, $Z_{il}=R_{il}+jX_{il}$ is the electrical analog of the acoustical coupled impedance. The relation between acoustical impedances and their electrical analogs is $Z_{il}=Z_{acil}/n^2$. Note that the normalized coupled impedances, being determined on the electrical side of a transducer, are the same as the normalized coupled impedances on acoustical side, namely, $Z_{12}/R_{11}=Z_{ac12}/R_{ac11}$. Thus, for the purpose of estimating of the mutual impedances it is not necessary to know the absolute values of the impedances in acoustical units.

The self-radiation impedance, which is denoted as Z_{11} for both transducers, may not be considered as known. It differs from the radiation impedance of a single transducer measured in the free space because of the presence of an interacting transducer. Moreover, by definition the self-radiation impedance has to be determined under the condition that the interacting transducer is blocked, in which case in the equivalent circuits of Fig. 2 $U_2=0$ and therefore $Z_{12}=z_{12}U_2/U_1=0$.

The electrical analogs of acoustical quantities can be obtained by measuring the input impedances of the transducers. After converting the series motional impedance into admittance the electrical circuit of transducer input may be represented in the form shown in Fig. 4(b), which is convenient for interpreting the results of measuring the transducer parameters by an impedance analyzer. In Fig. 4(b) C_m is the motional "capacitance," which at some frequencies may become inductance. The circuit parameters in Fig. 4(a) can be expressed through measured admittance values as follows:

$$R_{\Sigma i} = G_m/[(\omega C_m)^2 + G_m^2], \quad X_{\Sigma i} = -\omega C_m/[(\omega C_m)^2 + G_m^2], \tag{10}$$

where $G_m=1/R_m$ and

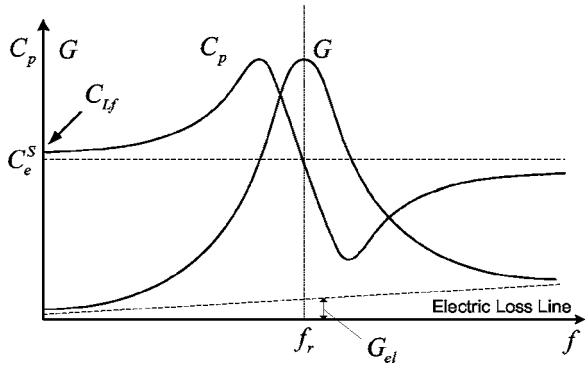


FIG. 5. Qualitative illustration of a typical plot for the input parameters of a transducer in the parallel circuit representation.

$$X_{\Sigma i} = \omega M_{\Sigma i}[1 - (f_r/f)^2]. \tag{11}$$

When measuring in air $M_{\Sigma i}=M=M_{eqv}/n^2$ and $f_r=f_a$, and when measuring in water $M_{\Sigma i}=M+X_i/\omega$ and $f_r=f_w$. The directly measured quantities are $G=G_{eL}+G_m$ and $C_p=C_e^S+C_m$. Typical plots for C_p and G, from which the motional capacitance, C_m, and conductivity, G_m, may be determined versus frequency, are depicted in Fig. 5. At a resonant frequency $C_m=0$, and therefore $C_p=C_e^S$. Thus, the clamped capacitance C_e^S can be considered as known. If a measurement transducer may be approximated as having one mechanical degree of freedom, this quantity also can be obtained as $C_e^S=C_{Lf}(1-k_{eff}^2)$, where C_{Lf} and k_{eff} are the capacitance measured at low frequency (typically at 1 kHz) and the effective coupling coefficient of the transducer, respectively. The motional conductivity, $G_m=G_p-G_{eL}$, in general can be found as it is shown in Fig. 5, although very often G_{eL} is much smaller than G_m and can be neglected. Given that $C_m(f)$ and $G_m(f)$ are measured in a frequency range around the resonant frequency, $R_{\Sigma i}$ and $X_{\Sigma i}$ can be calculated by formulas (10). After this the wanted self and mutual radiation impedances may be found from relations (11) and (7)–(9). The technique of determining the radiation impedances at a resonant frequency is outlined in Eqs. (12)–(16).

The reactive components of radiation impedance may be determined in two alternative ways. Formula (11) can be represented as

$$\begin{aligned} X_{\Sigma 1} &= \omega M(1 - f_a^2/f^2) + X_{11} + X_{12} \\ &= \omega(M + X_1/\omega)(1 - f_w^2/f^2). \end{aligned} \tag{12}$$

At the frequency $f=f_w$ it follows from Eq. (12) that

$$X_{11} + X_{12} = 2\pi f_w M(f_a^2/f_w^2 - 1) = 2\pi M F(w), \tag{13}$$

where it is denoted for brevity

$$F(f_w) = f_w(f_a^2/f_w^2 - 1). \tag{14}$$

At the frequency $f=f_a$, taking into account Eqs. (12) and (10), we obtain

$$X_{11} + X_{12} = -\omega_a C_m(f_a)/\{[\omega_a C_m(f_a)]^2 + G_m^2(f_a)\}. \tag{15}$$

Remembering that $C_m(f_w)=0$ at the frequency $f=f_w$, the active component of radiation impedance will be obtained from Eq. (10) in the form

$$R_{\Sigma 1} = R_{11} + R_{12} + r = 1/G_m(f_w). \tag{16}$$

In order to determine all the components of impedances Z_{11} and Z_{12}, two more equations in addition to the equations (13) or (15) and (16) have to be obtained. For this purpose the following experiments can be made.

Experiment I. Measuring the input impedance of one of the transducers, while equal voltages are applied to both of them in phase, $V_2=V_1$. Because of symmetry in this case $U_1=U_2$ and $Z_{ac12}=z_{ac12}$. At the resonant frequency f_{wI} of the transducer measured in water in experiment I according to formulas (13) and (16) we obtain

$$R_{11} + r_{12} + r = 1/G_{mI} \tag{17}$$

and

$$X_{ac11} + x_{ac12} = 2\pi M F_I. \tag{18}$$

In Eqs. (17) and (18) it is denoted

$$G_{mI} = G_m(f_{wI}), \quad F_I = F(f_{wI}). \tag{19}$$

(Subscripts made by roman numbers here and further correspond to the number of an experiment.)

Experiment II. Measuring the input impedance of one of the transducers, while applied voltages are equal by magnitude and opposite in phase, $V_2=-V_1$. Because of symmetry $U_2=-U_1$ and $Z_{ac12}=-z_{ac12}$. At the resonant frequency f_{wII} of the transducer in this experiment we have

$$R_{11} - r_{12} + r = 1/G_{mII} \tag{20}$$

and

$$X_{ac11} - x_{ac12} = 2\pi M F_{II}. \tag{21}$$

Combining the results of measurements expressed by formulas (17) and (20), we arrive at

$$R_{11} + r = (G_{mI} + G_{mII})/2G_{mI}G_{mII} \tag{22}$$

and

$$r_{12} = (G_{mI} - G_{mII})/2G_{mI}G_{mII}. \tag{23}$$

We note that in carefully designed transducers the resistance of mechanical loss is much smaller than the radiation resistance and therefore can be neglected (at least for investigation of acoustical interaction such transducers have to be chosen). Otherwise the resistance of loss can be determined separately by measuring the conductivity $G_m=1/r$ in air and introduced into formula (22). After this note the resistance of mechanical loss henceforth will be neglected for the sake of brevity.

Thus, from Eqs. (22) and (23) it follows that

$$r_{12}/R_{11} = (G_{mI} - G_{mII})/(G_{mI} + G_{mII}). \tag{24}$$

From formulas (18) and (21) it follows that

$$X_{11} = \pi M(F_I + F_{II}) \tag{25}$$

and

$$x_{12} = \pi M(F_I - F_{II}). \tag{26}$$

Thus,

$$x_{12}/X_{11} = (F_I - F_{II})/(F_I + F_{II}). \tag{27}$$

Calculating the absolute values of the reactances by formulas (25) and (26) requires knowing the equivalent mass of the transducer. Therefore it can be advantageous to use formula (15) for this purpose, although it looks more complicated. In this case all the quantities needed for calculation are available through experimentation.

Practically both experiments can be accomplished by measuring the input impedance of transducers when connected in parallel in phase and antiphase, accordingly.

The Z method of evaluating self- and mutual radiation impedances between transducers is based on their comparison with the internal impedances of the transducers. The results obtained are less accurate when the ratio of radiation impedance components to be measured to the corresponding parameters of comparison is smaller. The situation becomes especially critical for the acoustic reactances, which have to be compared with ωM_{eqv}, and in the case of deviation from the resonant frequency. Some relief can be achieved by a proper selection of the transducers intended for investigating the acoustical interaction. Thus, precautions should be taken to minimize the equivalent mass of the transducers (for example, in the case where a ring transducer is concerned, it is better to use thinner rings). At the same time, in order to increase the accuracy of measuring the mutual impedances, the above-described experimental technique can be modified. The modification is to measure the input impedance of one of the transducers while another transducer is operating under a larger applied voltage. Suppose that $V_2 \gg V_1$. Qualitatively it is clear that in this case $U_2 \gg U_1$, and therefore $Z_{ac12} \gg z_{ac12}$. In this way the increased value of Z_{12} can be measured with greater accuracy. But in order to calculate z_{12} from the measured results the exact value of ratio of velocities $K_U = U_2/U_1$ has to be known. A peculiarity in this case is that the coupled impedances in the mechanical branches of the transducers' equivalent circuits in Fig. 4(a) become different, namely, $Z_{ac21} = Z_{ac12}/K_U^2$. Therefore $U_2/U_1 \neq V_2/V_1$, and the value of K_U has to be determined separately. This can be done through the following experiment.

Experiment III. Measuring of the mutual impedance between the two transducers under the condition that voltages applied to them are in phase but have different magnitudes, and $V_2 \gg V_1$. The self-radiation impedance of the transducer can be considered as known, resulting from experiments I and II by formulas (22) and (25). Components of the coupled impedance, $Z_{12} = z_{12}K_U$, may be determined as a result of performing the same procedures as in experiment I. Namely, at the resonant frequency f_{wIII}

$$R_{11} + R_{12} = 1/G_{mIII} \tag{28}$$

and

$$X_{11} + X_{12} = 2\pi M F_{III}. \tag{29}$$

From relations (28) and (29), and by taking into consideration expressions (22) and (25) for R_{11} (remember that resistance r is neglected) and X_{11}, we arrive at

$$R_{12} = 1/G_{mIII} - (G_{mI} + G_{mII})/2G_{mI}G_{mII}, \tag{30}$$

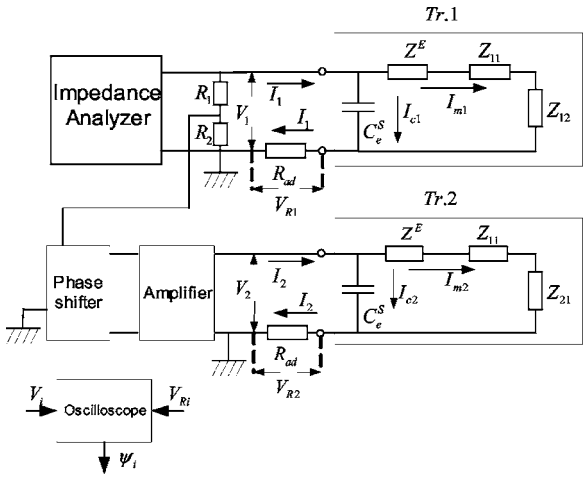


FIG. 6. Electrical circuit of an experimental setup for measuring the mutual impedances by the Z method in a general case that different voltages are applied to the interacting transducers, $Z^E = Z_m^E/n^2$.

$$X_{12} = \pi M(2F_{III} - F_I - F_{II}). \tag{31}$$

The mutual resistance and reactance will be found from relation

$$r_{12} + jx_{12} = (R_{12} + jX_{12})/K_U, \tag{32}$$

but for this purpose the ratio of velocities K_U remains to be determined. One of the possible setups for fulfillment experiment III and for determining K_U is represented in Fig. 6, where the equivalent circuits of interacting transducers with their mechanical branches transformed into the electrical sides are included. The condition $(R_1 + R_2) \gg |Z_{input}|$ has to be fulfilled for the voltage divider in order to not interfere with the transducer impedance measurement. Currents flowing through the transducers are denoted in Fig. 6 as follows: I_i is the total current through transducer i $(i = 1, 2)$, I_{Ci} is the current through the blocked transducer, and I_{mi} is the motional current through the electrical analog of the mechanical branch. The motional current is proportional to the vibration velocity of mechanical system $I_{mi} = nU_i$, therefore

$$K_U = U_2/U_1 = I_{m2}/I_{m1}, \tag{33}$$

and the ratio of velocities may be determined experimentally as the ratio of the motional currents I_{m2} and I_{m1}. The total current through a transducer

$$I_i = I_{Ci} + I_{mi} \tag{34}$$

can be measured as shown in Fig. 6. Namely,

$$I_i = V_{Ri}/R_{ad}, \tag{35}$$

where V_{Ri} is the voltage across the known additional resistance R_{ad} connected in series with the transducer (it has to be chosen $R_{ad} \ll |Z_{\text{input}}|$ in order not to change voltage applied to the transducer). Simultaneously the phase angle between the total current (or voltage V_{Ri}) and applied voltage V_i has to be measured. We denote this angle as

$$\arg(V_{Ri}/V_i) = \psi_i. \tag{36}$$

As the transducers are assumed to be electromechanically identical and with their parameters predetermined, the

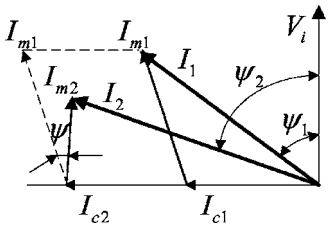


FIG. 7. The vector diagram for evaluation of the motional currents.

blocked capacitances can be considered as equal, $C^s_{e1}=C^s_{e2}$, and known. (If the capacitances were not exactly equal, they could be equalized by adding a capacitance in parallel to the transducer with smaller C^s_e.) Thus, the current through the capacitance can be calculated as

$$I_{Ci} = j\omega C^s_e V_i. \tag{37}$$

After the currents I_i and I_{Ci} are determined, the motional currents I_{mi} may be calculated with the procedure illustrated by the vector diagram in Fig. 7. (Note that in this experiment voltages applied to the transducers are in phase.) Knowing I_{m2} and I_{m1} we arrive at the required ratio for the velocities of the transducers,

$$K_U = I_{m2}/I_{m1} = |I_{m2}/I_{m1}|e^{j\psi}, \tag{38}$$

where ψ is the phase angle between vectors I_{m2} and I_{m1} shown in the vector diagram. In order to increase the accuracy of calculating the motional currents from results of measuring the total currents, the blocked capacitances can be tuned at the measurement frequency by inserting inductors in parallel. In this case, $I_{mi}=I_i$. Note that if the transducers are not identical but their parameters are known, this does not change the matter in principle, it just complicates the calculations.

After K_U is determined by formula (38), the components r_{12} and x_{12} of the mutual impedance may be calculated from Eq. (32) as follows:

$$r_{12} = (R_{12}\cos\psi + X_{12}\sin\psi)\left|\frac{I_{m1}}{I_{m2}}\right|, \tag{39}$$

$$x_{12} = (X_{12}\cos\psi - R_{12}\sin\psi)\left|\frac{I_{m1}}{I_{m2}}\right|. \tag{40}$$

In this case voltages V_2 and V_1 are in phase, and it is very likely that the phase shift ψ is small (i.e., $\cos\psi\cong 1$, $\sin\psi\cong 0$) and

$$r_{12} \cong R_{12}/K_U, \quad x_{ac12} \cong X_{ac12}/K_U, \tag{41}$$

where R_{12} and X_{ac12} are given by formulas (30) and (31).

As noted previously, the results of measuring the coupled reactances are less accurate than that of the coupled resistances. From relation (40) it follows that if the phase shift between U_2 and U_1 was $\psi=-\pi/2$, then $x_{12}=R_{12}/|K_U|$ (i.e., the mutual reactance could be measured as the coupled resistance), and the accuracy of determining this quantity could be greatly increased. This can be achieved by setting

$$V_2 \cong jV_1. \tag{42}$$

Some additional phase shift may occur between the velocities U_2 and U_1 due to asymmetric acoustical loading of the transducers by the coupled impedances resulting in a less accurate approximation. The phase shifter is included in the measurement circuit in Fig. 6 to provide the needed phase shift between the voltages applied to the transducers. After the phase shift of $\pi/2$ is insured between the motional currents I_{m2} and I_{m1}, the procedure of measuring the coupled resistance R_{12} and determining the mutual reactance $x_{12}=R_{12}/|K_U|$ is the same, as it is demonstrated in experiment III.

B. *V* Method

Returning back to Fig. 2(b), the following relations can be obtained between the output voltage of transducer 1 and velocity U_2 of vibration for transducer 2 depending on position "*oc*" or "*sc*" of the switch:

$$V_{1oc} = z_{ac12}U_2 n/j\omega C^s_e(Z^E_{moc} + Z_{ac11}), \tag{43}$$

$$V_{1sc} = I_{m1}R = z_{ac12}U_2 n/(Z^E_m + Z_{ac11}). \tag{44}$$

In these relations $Z^E_{moc}=Z^E_m+n^2/j2\pi fC^S_e$ is the mechanical impedance of the open-circuited transducer, I_{m1} is the motional current through transducer 1, and resistance R is assumed to be much smaller than $1/\omega C^S_e$. We note that $U_2 n=I_{m2}$, and therefore Eq. (44) can be rewritten as

$$z_{ac12}/(Z^E_m + Z_{ac11}) = I_{m1}R/I_{m2} = V_{sc}/I_{m2}. \tag{45}$$

The procedure of determining the motional current I_{m2} is described in relation to experiment III and therefore the mutual impedance z_{ac12} can be determined from Eq. (45) around the resonant frequency band. This holds so far as all of the parameters of the transducer are known including the self-radiation resistance R_{ac11}, which can be obtained from the above-described Z method. At the resonant frequency of transducer in air, f_a, $Z^E_m\cong 0$ and Eq. (45) becomes

$$(z_{ac12}/Z_{ac11})_{f_a} = (V_{sc}/I_{m2})_{f_a}. \tag{46}$$

For frequencies below the resonant frequency, at which the mechanical system of a transducer may be considered as stiffness controlled, $Z_{ac11}\ll Z^E_m$, $z_{ac12}\ll Z^E_m$, $Z^E_m\cong 1/j2\pi fC^E_m$, and from Eq. (43) we obtain

$$z_{ac12} = (V_{1oc}/V_2)[(1-k^2_{eff})/j2\pi fC^E_m k^2_{eff}]. \tag{47}$$

In Eq. (47) it is taken into account that $n^2C^E_m/C^s_e=k^2_{eff}/(1-k^2_{eff})$.

After the ratio $V_{1oc}/V_2=|V_{1oc}/V_2|e^{j\psi_{1oc,2}}$, where $\psi_{1oc,2}$ is the phase angle between the voltages, is measured, both the active and reactive components of the mutual impedance can be determined from Eq. (47). The block diagram of an experimental setup for implementing the *V* method is shown in Fig. 8.

The *V* method is especially advantageous for estimating of the relative change of mutual impedance versus separation between transducers. If the voltage applied to a radiating

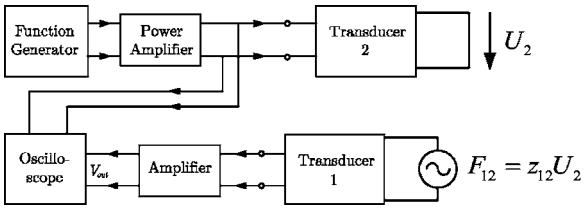


FIG. 8. The block diagram of an experimental setup for implementing the V method.

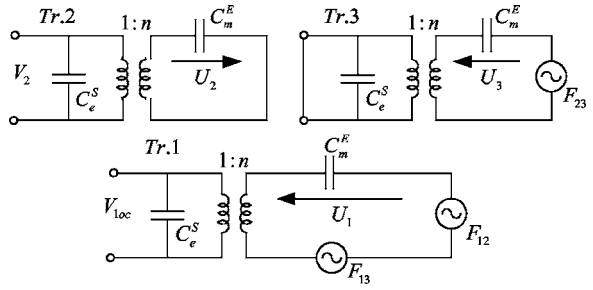


FIG. 9. The application of the V method to determine the mutual impedances in an array.

transducer is kept constant (V_2=const) and output voltages of transducer 1 at separation y and at separation d, corresponding to the position of transducer 1 adjacent to the radiating transducer, are measured, then by using expressions (46) and (47) we obtain

$$|z_{ac12}/Z_{ac11}|_y/|z_{ac12}/Z_{ac11}|_d = |V_{sc}/I_{m2}|_y/|V_{sc}/I_{m2}|_d \tag{48}$$

at the resonant frequency and

$$|z_{ac12}(y)|/|z_{ac12}(d)| = |V_{1oc}(y)|/|V_{1oc}(d)| \tag{49}$$

at frequencies below the resonance.

An important feature of the V method is that it is possible to use the measurement transducers operating far below their resonant frequency as a tool for investigating mutual impedances. The only condition that has to be met is that the configuration of the radiating surface of the measurement transducers has to be the same as the configuration of the actual transducers, for which the results of investigation are intended.

The Z and V methods may be considered as complementary to each other. The Z method is advantageous for the determining of the absolute values of self- and mutual radiation impedances at the frequencies close to resonance of the transducers. The V method can be used for measurements in a frequency range below the resonant frequency of the measurement transducers intended for an investigation of the mutual impedances. It has some shortcomings when it is used for absolute measurements of the mutual radiation impedances around the resonant frequencies of transducers. Moreover, implementing the V method at these frequencies requires the self-radiation impedance of a transducer to be known in advance.

IV. DETERMINING THE MUTUAL IMPEDANCES BETWEEN TRANSDUCERS IN AN ARRAY

The technique described above for measuring the mutual radiation impedances between the two transducers in the free space cannot be applied in a straightforward way to analogous measurements in an array. When measuring in an array, one has to deal with a number of simultaneously vibrating transducers, although the interaction between only two of them has to be measured. Due to the general property of the mutual impedances to decrease with increasing of separation between transducers, the contribution of sufficiently remote transducers in the measurement results of the two transducers under investigation can be neglected, and often only a part of an array needs to be tested. For example, in an array of transducers shown schematically in Fig. 1(c) the objective is to determine the mutual impedances between transducers z_{ac12}, z_{ac13}, and so on until the mutual impedance reaches a level that may be considered as negligible. That is, if the ratio $|z_{ac1i}|/|z_{ac12}|$ becomes small enough, then the number of transducers, which can be considered as representative for conducting the measurements, may be restricted by a group of i members of the array from each side of the central transducer 1.

It is most advantageous to fulfill an estimation of the mutual impedances by the V method in a frequency range below the resonant frequency of the measurement transducers. Certain limitations of this method arise from the fact that in practice we are interested in knowing the values of mutual impedances in the frequency range around the resonant frequency of the actual transducers in the array. Therefore, the measurement transducers having a radiating surface of the same shape as the actual transducers must have a higher resonant frequency. So far as arrays populated by transducers with flat radiating surfaces are concerned (such as, for example, Tonpilz or flexural-type transducers), the size of the surface can be independent of the resonant frequency of a transducer. In the case that the ring transducers populate an array, the diameter of the radiating surface is inversely proportional to the resonant frequency. Therefore the results obtained below the resonant frequency of a measurement transducer can be applied to an actual transducer in the frequency range around its resonant frequency only in the case that the resonant frequency of the measurement transducer is higher. Thus, the results obtained below the resonance of measurement transducers made of PZT ceramic may be applicable for the ring transducers made of a material with smaller sound speed.

Since the frequency range of measurements is below the resonant frequency of the measurement transducers, their mechanical systems may be considered as being stiffness controlled, i.e., $Z_m^E \cong 1/j\omega C_m^E$, and all of the radiation impedances can be neglected. The effects of interaction between transducer 1 and transducers 2 and 3 [see Fig. 1(a)] can be represented by the forces $F_{ac12}=2z_{ac12}U_2$ and $F_{ac13}=2z_{ac13}U_3$, as this is illustrated with the equivalent circuit in Fig. 9. An outline of application for the V method is as follows. Assume that voltage V_2 is applied to transducers 2, inputs of transducers 3 are short-circuited, and the output voltage V_{1oc} of transducer 1 is measured. From the equivalent circuits in Fig. 9 we obtain $U_2=V_2 nj\omega C_m^E$ and $U_3=z_{ac12}U_2 j\omega C_m^E$. Upon substituting these values of velocities

U_2 and U_3 into expressions for forces F_{ac12} and F_{ac13} we obtain for the total acoustic force acting on transducer 1

$$F_{ac1} = F_{ac12} + F_{ac13} = 2z_{ac12}[1 + z_{ac13} j\omega C_m^E] j\omega C_m^E n V_2. \tag{50}$$

The second term in the brackets can be neglected because the mutual radiation impedance is much smaller than $1/j\omega C_m^E$. From this point the analysis becomes analogous to that for the two transducers in the free space, and finally we arrive at the following expressions for acoustomotive force F_{ac1}, for the output voltage of transducer 1 under this force and for the mutual impedance

$$F_{ac1} = 2z_{ac12} j\omega C_m^E n V_2, \tag{51}$$

$$(V_{1oc})_2 = 2z_{ac12} j\omega C_m^E V_2 k_{eff}^2/(1 - k_{eff}^2), \tag{52}$$

$$z_{ac12} = [(V_{1oc})_2/V_2](1 - k_{eff}^2)/j2k_{eff}^2 \omega C_m^E. \tag{53}$$

In order to determine the mutual impedance z_{13}, voltage V_3 has to be applied to transducers 3, outputs of transducers 2 have to be short-circuited, and the output voltage of transducer 1 has to be measured. After using exactly the same procedure as in the previous case for the mutual impedance z_{ac13} we will obtain an expression analogous to expression (46), namely,

$$z_{ac13} = [(V_{1oc})_3/V_3](1 - k_{eff}^2)/j2k_{eff}^2 \omega C_m^E. \tag{54}$$

Based on formulas (53) and (54) a general conclusion can be drawn that if the equal voltages are applied to the consecutive transducers i, while measuring the output voltage of transducer 1, $(V_{1oc})_i$, then

$$|z_{ac1i}|/|z_{ac12}| = |(V_{1oc})_i|/|(V_{1oc})_2|. \tag{55}$$

This relation is convenient for estimating a comparative contribution of the mutual impedances between transducers in an array to the total radiation impedance of a single transducer, while formulas of the type (53) and (54) can be used for evaluating the components of the mutual impedances.

In order to determine the self- and mutual radiation impedances of transducers in an array in the frequency band around the resonant frequency, the Z method can be used. We will assume that all of the transducers under consideration are located inside the array and therefore have equal self-radiation impedances. The transducers located at the edges of the array may generally have different self-radiation impedances, which may cause an additional error in determining the mutual impedances. In order to avoid an "edge effect" and to keep the self-impedances of the measurement transducers approximately equal, at least one more transducer must be added from each side to a group of transducers chosen for measurements, although these transducers are not intended to actively participate in measurements. Thus, only a restricted group of transducers may be used to model a real situation in an array in terms of interaction between neighboring transducers. This is shown in Fig. 1(c) as an example of a particular array, in which only z_{ac12} and z_{ac13} are of significance and have to be measured together with the self-impedance Z_{ac11} of the transducers. The assumption that the transducers are identical by their electromechanical properties remains in place.

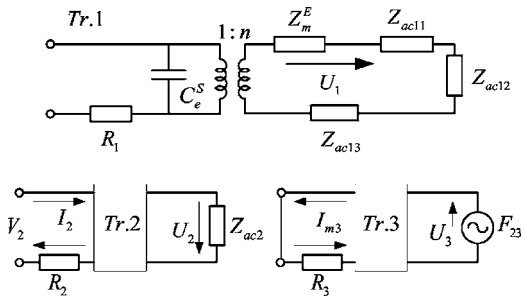


FIG. 10. The application of the Z method to measure the mutual impedances in an array.

The quantities Z_{ac11}, z_{ac12}, and z_{ac13} may be determined by measuring the input impedance of transducer 1. The equivalent circuit of this transducer, in which effects of interaction with transducers 2 and 3 are accounted for by the coupled impedances $Z_{ac12} = 2z_{ac12}U_2/U_1$ and $Z_{ac13} = 2z_{ac13}U_3/U_1$, is shown in Fig. 10. In order to determine the wanted radiation impedances consider the following experiments.

Experiment I a: Voltages V_1 and V_2 such as $|V|_2 \gg |V_1|$ are applied in phase to transducers 1 and 2. Outputs of transducers 3 are closed through a small resistance R_3. Under action of acoustomotive force $F_{32} = z_{ac12}U_2$ the motional current I_{m3} through the electrical side of transducer 3 will be $I_{m3} = U_3 n$. As it follows from an equivalent circuit in Fig. 10 at $R_3 \ll 1/\omega C_e^S$

$$I_{m3} = V_{R_3}/R_3, \tag{56}$$

where V_{R_3} is the voltage across resistance R_3. The radiation impedance of transducer 1 may be represented as

$$Z_{acIa} = Z_{ac11} + 2z_{ac12}(U_2/U_1)_{Ia} + 2z_{ac13}(U_3/U_1). \tag{57}$$

The ratios of velocities can be replaced by ratios of the motional currents, I_{mi}, on the electrical side of the transducers. Thus, $U_2/U_1 = I_{m2}/I_{m1}$ and $U_3/U_1 = I_{m3}/I_{m1}$. We represent a ratio of the motional currents as

$$I_{mi}/I_{m1} = |I_{mi}/I_{m1}| e^{j\psi_{1i}}, \tag{58}$$

where ψ_{1i} is a phase shift between the currents.

In the same way as in experiment I we obtain the following equations analogous to Eqs. (17) and (18) at the resonant frequency of transducer 1 in water, which in this experiment we denote as f_{wIa}:

$$R_{\Sigma Ia} = R_{11} + 2\,\mathrm{Re}\{z_{12}|I_{m2}/I_{m1}|_{Ia} e^{j(\psi_{12})_{Ia}} + z_{13}|I_{m3}/I_{m1}|_{Ia} e^{j(\psi_{13})_{Ia}}\} = 1/G_{mIa}, \tag{59}$$

$$X_{\Sigma Ia} = X_{11} + 2\,\mathrm{Im}\{z_{12}|I_{m2}/I_{m1}|_{Ia} e^{j(\psi_{12})_{Ia}} + z_{13}|I_{m3}/I_{m1}|_{Ia} e^{j(\psi_{21})_{Ia}}\} = 2\pi M F_{Ia}. \tag{60}$$

The total input resistance $R_{\Sigma Ia}$, resonant frequency f_{wIa}, and motional currents I_{m1}, I_{m2} can be measured by using the experimental setup illustrated in Fig. 6 and procedures described in experiment III. Current I_{m3} can be calculated from

formula (56) after voltage V_{R_3} is measured. In order to get a number of equations necessary for calculating all of the radiation impedances, several more experiments have to be fulfilled.

Experiment II a. Voltage V_2 of the same magnitude as in experiment Ia is applied to transducers 2 in antiphase while transducers 3 remains short-circuited. As $V_1/V_2 \ll 1$ and therefore $U_1/U_2 \ll 1$, the coupled impedance $Z_{ac21} = z_{ac12}U_1/U_2$ and $Z_{ac23} = z_{ac12}U_3/U_2$ in the equivalent circuit of transducer 2 can be neglected in comparison with its self-radiation impedance. Therefore it can be concluded that $(U_2)_{IIa} \cong -(U_2)_{Ia}$. Since $U_3 \sim z_{ac12}U_2$, it can be also concluded that $(U_3)_{IIa} \cong -(U_3)_{Ia}$. This can be summarized as

$$(I_{m2})_{IIa} \cong -(I_{m2})_{Ia}, \quad (I_{m3})_{IIa} \cong -(I_{m3})_{Ia}. \tag{61}$$

The currents $(I_{m1})_{Ia}$ and $(I_{m1})_{IIa}$ have different magnitudes, but are in phase with voltage V_1 at the resonant frequency. Therefore the conclusion can be made that

$$(\psi_{12})_{Ia} \cong (\psi_{12})_{IIa}, \quad (\psi_{13})_{Ia} \cong (\psi_{13})_{IIa}. \tag{62}$$

Taking into account relations (61) and (62) the input resistance and reactance of transducer 1 at resonant frequency f_{wIIa} may be represented as

$$R_{\Sigma IIa} = R_{11} - 2\,\mathrm{Re}\{z_{12}|I_{m2}|_{Ia}e^{j(\psi_{12})_{Ia}}/|I_{m1}|_{IIa} + z_{13}|I_{m3}|_{Ia}e^{j(\psi_{13})_{Ia}}/|I_{m1}|_{IIa}\} = 1/G_{mIIa}, \tag{63}$$

$$X_{\Sigma IIa} = X_{11} - 2Im\{z_{12}|I_{m2}|_{Ia}e^{j(\psi_{12})_{Ia}}/|I_{m1}|_{IIa} + z_{13}|I_{m3}|_{Ia}e^{j(\psi_{13})}_{Ia}/|I_{m1}|_{IIa}\} = 2\pi M F_{IIa}. \tag{64}$$

After multiplying Eqs. (59) and (63) by $|I_{m1}|_{Ia}$ and $|I_{m1}|_{IIa}$, respectively, and summing and subtracting them we obtain

$$R_{11} = (R_{\Sigma Ia}|I_{m1}|_{Ia} + R_{\Sigma IIa}|I_{m1}|_{IIa})/(|I_{m1}|_{Ia} + |I_{m1}|_{IIa}), \tag{65}$$

$$4\,\mathrm{Re}\{z_{12}|I_{m2}|_{Ia}e^{j(\psi_{12})_{Ia}} + z_{13}|I_{m3}|_{Ia}e^{j(\psi_{13})_{Ia}}\} = R_{\Sigma Ia}|I_{m1}|_{Ia} - R_{\Sigma IIa}|I_{m1}|_{IIa} - R_{11}(|I_{m1}|_{Ia} - |I_{m1}|_{IIa}). \tag{66}$$

Similarly from Eqs. (60) and (64) we obtain

$$X_{11} = 2\pi M(F_{Ia}|I_{m1}|_{Ia} + F_{IIa}|I_{m1}|_{IIa})/(|I_{m1}|_{Ia} + |I_{m1}|_{IIa}), \tag{67}$$

$$4Im\{z_{12}|I_{m2}|_{Ia}e^{j(\psi_{12})_{Ia}} + z_{13}|I_{m3}|_{Ia}e^{j(\psi_{13})_{Ia}}\} = 2\pi M[F_{Ia}|I_{m1}|_{Ia} - F_{IIa}|I_{m1}|_{IIa}] - X_{11}(|I_{m1}|_{Ia} - |I_{m1}|_{IIa}). \tag{68}$$

Thus, the self-radiation impedance of the transducers can be calculated from Eqs. (65) and (67), and in order to determine the four components of the mutual impedances two more equations are needed in addition to Eqs. (66) and (68). To obtain them the following experiments may be made; Experiment IIIa, in which case voltage $V_3 \gg V_1$ is applied to transducers 3 and transducers 2 are short-circuited, and experiment IVa, in which case the voltage of the same magnitude is applied to transducers 3 in antiphase with voltage V_1. By repeating the procedure used in experiments Ia and IIa, two missing equations analogous to Eqs. (66) and (68) will be obtained. It may be well expected that the angles ψ_{12} and ψ_{13} are small at the resonant frequency of transducer 1, at which measurements take place. In this case $\sin\psi_{1i} \cong 0$, $\cos\psi_{1i} \cong 1$, and the equations of the type (66) and (68) will be greatly simplified.

V. CONCLUSION

The Z and V methods of experimentally investigating acoustical interaction between transducers may be considered as complementary to each other. An application of the V method is more advantageous for a comparative estimation of mutual impedances in a broad frequency range depending on a separation between transducers. The Z method has the advantage of more accurate measurement of the absolute values of the mutual impedances in a frequency range around the resonant frequency of the measurement transducers.

ACKNOWLEDGMENTS

The author wishes to acknowledge his late friend and colleague L. D. Lubavin, with whom this work was started back in Russia. Thanks go to Dr. David Brown for discussing the results and to Corey Bachand for his help in revising the paper for publication. The work was sponsored in part by ONR 321 MS.

[1] R. T. Richards, J. B. Blottman, and B. McTaggart, "Physics of Array Element Interaction Phenomena," in *Power Transducers for Sonics and Ultrasonics*, edited by B. F. Harmonic, O. B. Wilson, and J.-N. Decarpigny Proceedings of the International Workshop, Toulon, France, 1990.

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

[3] E. M. Arase, "Mutual radiation impedance of square and rectangular pistons in a rigid plane," J. Acoust. Soc. Am. **36**, 1521–1525 (1964).

[4] D. H. Robey, "On the radiation impedance of an array of finite cylinders," J. Acoust. Soc. Am. **27**, 706–710 (1955).

[5] J. E. Greenspon and C. H. Sherman, "Mutual radiation impedance and nearfield pressure for pistons on a cylinder," J. Acoust. Soc. Am. **36**, 149–153 (1964).

[6] C. H. Sherman, "Mutual radiation impedance of sources on a sphere," J. Acoust. Soc. Am. **31**, 947–952 (1959).

[7] F. Porges and C. H. Sherman, "Measurement of Variation of Radiation Resistance with Separation of Pairs of Underwater Transducers," *Proceedings of the Third International Congress on Acoustics*, Stuttgart, 1959, Vol. II, pp. 657–677.

[8] F. B. Stumpf, "Interaction radiation resistance for a line array of two and three magnetostrictive-stack transducers at an air-water surface," J. Acoust. Soc. Am. **36**, 174–176 (1964).

[9] B. Aronov, T. Oishi, and D. A. Brown, "Methods of experimental investigation of acoustic interaction between electroacoustical transducers in array," J. Acoust. Soc. Am. **112**(5), 2407(A) (2002).

[10] B. S. Aronov, "The energy method for analyzing the piezoelectric electroacoustic transducers," J. Acoust. Soc. Am. **117**(1), 210–220 (2005).

[11] R. G. Bobber, *Underwater Electroacoustic Measurements* (Naval Research Laboratory, Washington, DC, 1970).

Characterization of impedance boundary as damped harmonic oscillators via impulse reflection

K.-Y. Fung[a] and Xiaodong Jing[b]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, People's Republic of China

(Received 17 August 2005; revised 10 March 2006; accepted 26 March 2006)

This paper demonstrates that various sound-absorbing/reflective materials or structures can be effectively characterized over a wide band of frequency by the sum of only a few damped harmonic oscillators. This method is physically sound in the sense that each damped harmonic oscillator corresponds to the natural resonance of the impedance boundary as it is numerically advantageous to afford a simple, recursive, efficient, and unconditionally stable boundary algorithm for computational aero-acoustics applications. The time-domain damped-harmonic-oscillator impedance characterization method is finally validated by accurate predictions of reflected pulses from impedance boundaries physically formed of Helmholtz resonators or various exits of a circular pipe. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2198185]

PACS number(s): 43.58.Bh, 43.20.El, 43.20.Ye, 43.20.Bi [JBS] Pages: 3831–3838

I. INTRODUCTION

The problem of sound reflection is of fundamental importance to acoustics, and the concept of acoustic impedance $Z=\hat{p}/\hat{u}_n$, defined in the frequency domain by the ratio Fourier-transformed acoustic pressure $\hat{p}$ and velocity $\hat{u}_n$ normal into a surface, has led to numerous successful treatments and applications for noise control and abatement. The recent advances and interest in computational aero-acoustics (CAA) necessitate the extension of acoustic impedance to its time-domain equivalent for transient or broadband analysis. Various methods have been proposed for the construction of the time-domain equivalent of acoustic impedance and its efficient implementation as a general boundary condition.[1,2] These methods, however, have been indirectly inferred from data taken under the harmonic assumption. Moreover, with regard to the proper mathematical treatment of a boundary condition, little attention has been given to the physical implications of the equivalent time-domain reflection process of an impedance boundary. In this paper, a direct time-domain method of characterizing an impedance boundary is proposed which finds its basis on the resonance nature of practical sound-absorbing materials or structures.

When the classical concept of impedance is extended to its time-domain equivalent, the concern of causality arises. The direct inverse Fourier transform of a frequency-domain impedance model, such as the two- or three-parameter impedance models for porous materials, generally fails to give a causal time-domain impedance boundary condition (TDIBC), as discussed by Berthelot.[3] The reason is that an analytical, numerical, or empirical impedance model in the frequency domain is likely to have been constructed from a relation or data that are frequency band limited, and therefore is incapable of the provision of asymptotically correct behaviors.

Recently, Fung and Ju[4,5] proposed the construction of causal, stable TDIBC through the ratio of domain-exiting and entering characteristics and in the form of a system of damped harmonic oscillators, plausibly related to the physical resonance of a sound-absorbing material. Their method renders simple, recursive, efficient, and unconditionally stable TDIBC as a generalized closure for the linearized Euler's equation. Actually, in the field of electromagnetism, damped sinusoids identified from the transient response of a conducting object have long been treated as the "fingerprints" of the object and used for target identification.[6] This lends credence and support for finding its counterpart in acoustics.

Here, we systematically examine several well-used sound-absorbing structures or materials and present the results for micro-perforate panels (MPP), porous panels, Helmholtz-resonator structures, and various pipe exits. It is demonstrated that the physically realized, time-domain damped-harmonic-oscillator behavior of an impedance boundary is inherently related to the natural resonances of the sound-absorbing structure or material. From a physical point of view, there are three basic features within a physical impedance boundary: damping, mass, and elasticity. A simple combination of these features forms physically a damped harmonic oscillator (DHO) or mathematically a damped sinusoid. Complex impedance behaviors are simply the result of a combination of a series of damped harmonic oscillators. It is further shown that just a few of such oscillators can render a reflection coefficient model valid over a broad bandwidth of frequency for the commonly used sound-absorbing materials and structures. It is physically sound as it is numerically advantageous for their immediate applicability as TDIBC.

An impulse acoustic experiment is designed to validate the proposed time-domain impedance characterization

[a]Author to whom correspondence should be addressed. Electronic mail: mmkyfung@polyu.edu.hk
[b]Present address: School of Jet Propulsion, No. 407, Beijing University of Aeronautics and Astronautics, Beijing 100083, P. R. China. Electronic mail: jingxd@buaa.edu.cn

method. In the experiment, a short-duration acoustic pulse instead of a harmonic wave is used to examine the reflection property of a physical impedance boundary such as a Helmholtz-resonator terminus or exit in a one-dimensional waveguide. It is demonstrated that the measured reflected pulse can be accurately predicted by an acoustic propagation model based on the DHO characterization of the impedance boundary.

II. DHO AS ACOUSTIC BOUNDARY DESCRIPTOR

The reflectivity of an impedance boundary is conventionally characterized by the reflection coefficient $\hat{W}(\omega)$, which is generally a complex-valued function of real angular frequency ω defined by the ratio of the reflected sound pressure $\hat{p}^-(\omega)$ to the incident $\hat{p}^+(\omega)$,

$$\hat{W}(\omega) = \frac{\hat{p}^-(\omega)}{\hat{p}^+(\omega)}, \tag{1}$$

and is related to the acoustic impedance $Z(\omega)$ as

$$\hat{W}(\omega) = \frac{Z(\omega) - 1}{Z(\omega) + 1}. \tag{2}$$

On the assumption that $\hat{W}(\omega)$ can be analytically continued to the entire complex ω plane and that only causal reflection processes are physically realizable, i.e., $\int_{-\infty}^{0} p^+(t-\tau)W(\tau)d\tau = 0$, the poles of $\hat{W}(\omega)$ must lie in the upper half complex plane so the inverse Fourier transform of $\hat{W}(\omega)$ is

$$W(t) = jH(t)\sum_k \text{residues}[\hat{W}(\omega)e^{j\omega t}, \lambda_k], \tag{3}$$

where λ_k are the poles of $\hat{W}(\omega)$, $j=\sqrt{-1}$, and $H(t)$ is the Heaviside function. Thus, the time-domain equivalent of the reflection coefficient is the reflection impulse,

$$W(t) = A_0\delta(t) + H(t)\sum_k A_k \exp(j\lambda_k t) + A_k^* \exp(j\lambda_k^* t). \tag{4}$$

Here, λ_k are also referred as the complex natural frequencies, whose physical meaning will be discussed next, and A_k are correspondingly the complex amplitudes or residues. The δ function in Eq. (4) results from the limiting value of $\lim_{\omega\to\infty}|\hat{W}(\omega)|$. Since $W(t)$ must be real, both the natural frequencies and the complex amplitudes appear in pairs as $\lambda_k, \lambda_k^* = \pm\alpha_k + j\beta_k$ and $A_k, A_k^* = a_k \pm jb_k$. So, mathematically the time-domain equivalent of the reflection coefficient can be expressed as the sum of damped sinusoids.

There are methods to identify the natural frequencies and the complex amplitudes, λ_k and A_k, for which the frequency-domain counterpart of Eq. (4) in partial fraction form,

$$\hat{W}(\omega) = A_0 + \sum_k \frac{A_k}{s - j\lambda_k} + \frac{A_k^*}{s - j\lambda_k^*}, \tag{5}$$

or in rational form,

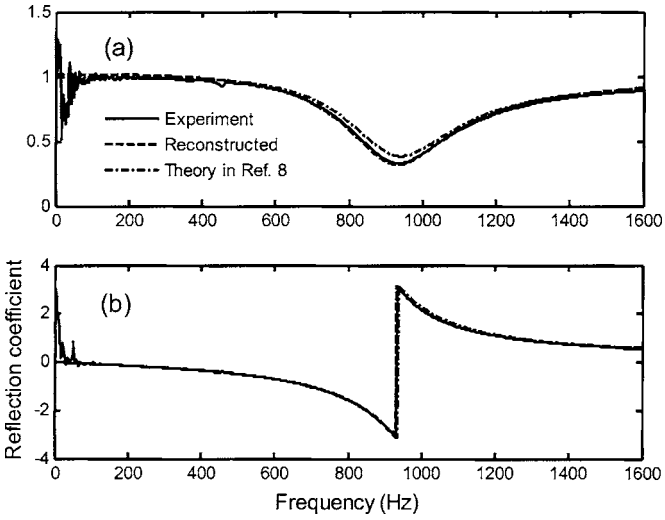


FIG. 1. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the MPP with hole diameter 1.0 mm, plate thickness 1.0 mm, porosity 1.2% and cavity depth 20 mm.

$$\hat{W}(\omega) = \frac{\sum_{n=0}^{M} p_n s^n}{1 + \sum_{m=1}^{M} q_m s^m}, \tag{6}$$

where $s=j\omega$ and p_n and q_m are real coefficients, would be useful. Given a set of N $(N>2M+1)$ values of $\hat{W}(\omega)$ at the discrete frequencies ω_l, an iterative least-squares approach can be used to compute these coefficients.[7] Then, it is straightforward to factorize Eq. (6) for λ_k and A_k.

III. VALIDATION OF DHO IMPEDANCE CHARACTERIZATION

A. Impedance tube measurement

We first examined several real impedance boundaries formed by practical sound-absorbing materials or structures in a B&K 4206 impedance tube. A single-layer MPP is measured and its reflection coefficient presented in Fig. 1. It is shown from the analysis of the above frequency-domain identification method that the reflection coefficient of the MPP possesses only one pole over the frequency range from 0 to 1600 Hz, so the measured curve can be fairly well reconstructed by the damped-sinusoid pair given in Table I. The theoretical results computed from Maa's[8] formula are also given in Fig. 1 for comparison. It is further shown that

TABLE I. Identified DHOs for the MPP and the porous panel.

Types of sound-absorbing structures	Natural frequencies (Hz), $\lambda_k/2\pi$	Complex amplitudes[a] (Hz), A_0 or $A_k/2\pi$ $(k\geqslant 1)$
Microperforated panel	–	1.0
	$\pm 892.6 + j244.0$	$-321.1 \mp j96.3$
Porous panel	–	−0.2827
	$0.0 + j86.2$	−25.0
	$0.0 + j283.8$	520.0
	$\pm 1265.3 + j741.0$	$500.2 \pm j582.6$

[a]The values of A_0 correspond to the null natural frequency indicted by "-".

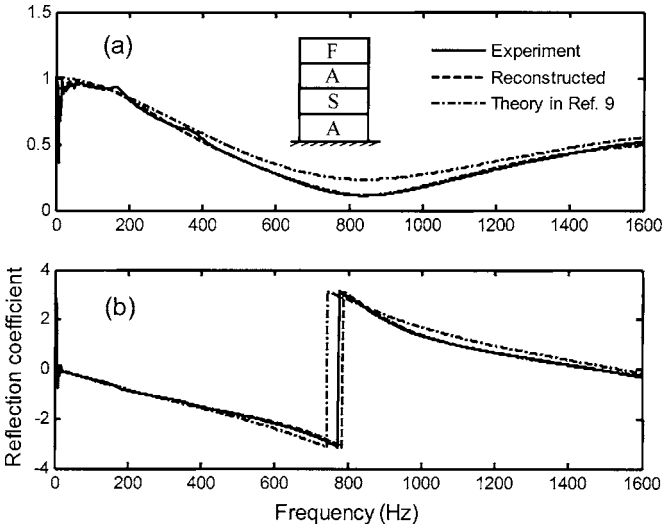


FIG. 2. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the porous panel of composite layers of FASA (F-fiber, S-sponge, A-air) and respective thicknesses of 28, 27, 25, and 20 mm.

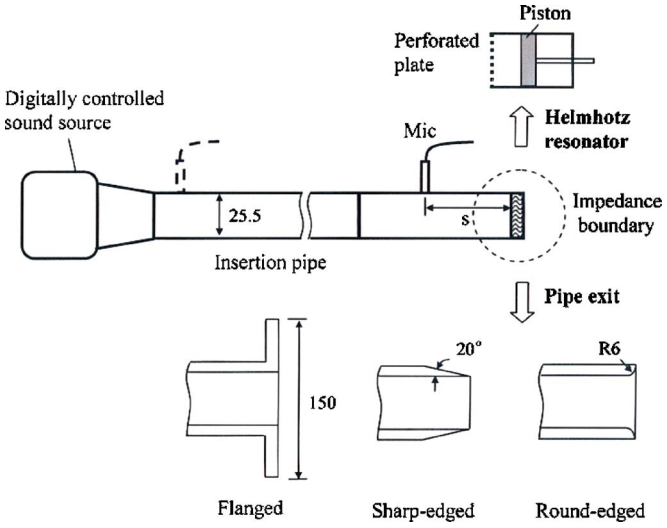


FIG. 3. Schematic of impulse experiment for impedance boundaries of (a) Helmholtz resonator and (b) pipe exit; distance from microphone to impedance sample $s=391$ mm for the Helmholtz resonator and $s=417$ mm for the pipe exit.

the real part of the identified $\lambda_k/2\pi$ is nearly equal to the resonant frequency of the single-layer MPP, 894 Hz, while its imaginary part is related to the damping factor.

Porous panels are also common sound-absorbing structures. A porous panel made of common sound-absorbing materials, a fiber of flow resistivity 7526.8 rayls/m and sponge of 748.0 rayls/m, was examined. The identified natural frequencies λ_k and the complex amplitudes A_k of the porous panel are given in Table I. The measured and the reconstructed reflection coefficients of the porous panel are shown in Fig. 2, and the theoretical results based on Delany and Bazley's[9] one-parameter empirical model are also given for comparison purpose. It is shown that the porous panel behaves in a way similar to the MPP, but more pairs of DHO are needed to reconstruct the measured reflection coefficient. It is evident that the resonant characteristics of the porous panel are more complex than that of the single-layer microperforated panel.

B. Impulse measurement

Figure 3 shows the experimental setup, which mainly consists of a one-dimensional waveguide made of a pipe 25.5 mm in diameter, an impedance terminus at one end of the waveguide and an impulse sound source at the opposite end. A B&K $\frac{1}{4}$-in. microphone, flush mounted into the pipe wall, is used to measure the sound pressure in the duct; the measured signal is digitally sampled by an NI DAQ Board 6062E. The impulse sound source, which is made of a cabinet loudspeaker, can be controlled to emit sound pulses of arbitrary waveform and duration as short as 0.5 ms with high fidelity, as demonstrated in Ref. 10. In order to control the sound emission, it first needs to obtain the transfer function of the sound source connected to an infinite long pipe. For this purpose, a 1.4-m-long insertion pipe is used to increase the time delay between the initial sound pulse and the subsequent reflections from the pipe terminus and thereby edit out the unwanted reflections. The microphone is first positioned near the sound source, indicated by the dashed line in Fig. 3, to obtain the transfer function of the source-pipe system, and then moved to the side of the impedance boundary to measure the sound pulse reflection.

In the experiment, the terminus is physically either a Helmholtz resonator or a pipe exit. As shown in Fig. 3, the Helmholtz resonator is made of a micro-perforated plate backed by a cavity, whose depth can be adjusted by moving the piston. Three different types of pipe exit, flanged, sharp-edged, and round-edged, are investigated in the experiment, among which the sharp-edged one is used to reduce the influence of the pipe wall thickness. The experiment of the Helmholtz resonator surface was carried out at 22 °C ambient temperature and that of the pipe exit at 17 °C.

Acoustic pulses as short as 1 ms are used in the experiment, and they are of three different desired waveforms referred as to sine, three-pole, and Butterworth (Appendix A). Since the incident sound pulse generated at one end of the pipe is much shorter than the time delay corresponding to twice the distance from the microphone to the test sample at the other end, the incident sound pulse can be clearly separated from the reflected pulse. It is thus straightforward to obtain the reflection coefficient of a test sample from the Fourier-transformed time signals per Eq. (1) and its natural frequencies and complex amplitudes by the identification method in Sec. II. In a test, the three-pole sound pulse is used for DHO identification and characterization. For subsequent analyses of various incident pulses $p^+(t)$, the reflected pulses $p^-(t)$ are predicted using the efficient recursive formula of Ref. 4,

$$p_k^-(t) = A_k[p^+(t) + z_k p^+(t-\Delta t)]\frac{\Delta t}{2} + z_k p_k^-(t-\Delta t)$$

$$p^-(t) = A_0 p^+(t) + \sum_k p_k^-(t) \tag{7}$$

where Δt is the time step, and $z_k = \exp(j\lambda_k \Delta t)$ ensures Eq. (7)'s stability for causal reflection, i.e., $|z_k| < 1$ when $\mathrm{Im}\lambda_k$

TABLE II. Identified DHOs for the Helmholtz resonators.

Cavity depth (mm)	Natural frequencies (Hz), $\lambda_k/2\pi$	Complex amplitudes[a] (Hz), A_0 or $A_k/2\pi$ $(k \geqslant 1)$
20	–	0.849
	$\pm 1259.6 + j718.1$	$-623.8 \mp j471.9$
60	–	0.836
	$\pm 502.1 + j569.5$	$-575.5 \mp j707.0$
	$\pm 3098.9 + j149.7$	$-65.3 \mp j110.4$
	$\pm 5904.4 + j127.6$	$-51.6 \mp j60.2$

[a]The values of A_0 correspond to the null natural frequency indicted by "-".

>0. Finally, the time-domain impedance characterization method is validated by the well-matched predicted and measured reflections.

1. Helmholtz resonators

For the Helmholtz resonator with a cavity of 20 mm, the identified natural frequencies and complex amplitudes are given in Table II. Figure 4 shows that the reflection coefficient of the Helmholtz resonator can be well reconstructed with only one damped harmonic oscillator. In Fig. 5, the predicted reflection pulse computed from Eq. (7) is compared with the measured reflection. It is not surprising to find the well-matched pulses, Fig. 5(a), for the three-pole pulse were used to identify the characterizing parameters. The DHO descriptor of an impedance boundary is rather validated by the excellent agreement between the predicted and measured reflection pulses for the sine and Butterworth incident, respectively shown in Figs. 5(b) and 5(c). When the depth of the Helmholtz resonator is increased to 60 mm, three pairs of natural frequencies and complex amplitudes, listed in Table II, are identified and the well-reconstructed reflection coefficient is shown in Fig. 6. It can be seen in Fig. 7 that the predicted and measured reflected pulses agree well not only for the three-pole incident, but also for the sine and the Butterworth pulses. In Figs. 4 and 6, the theoretical reflection coefficients from Ref. 8 are also plotted for comparison.

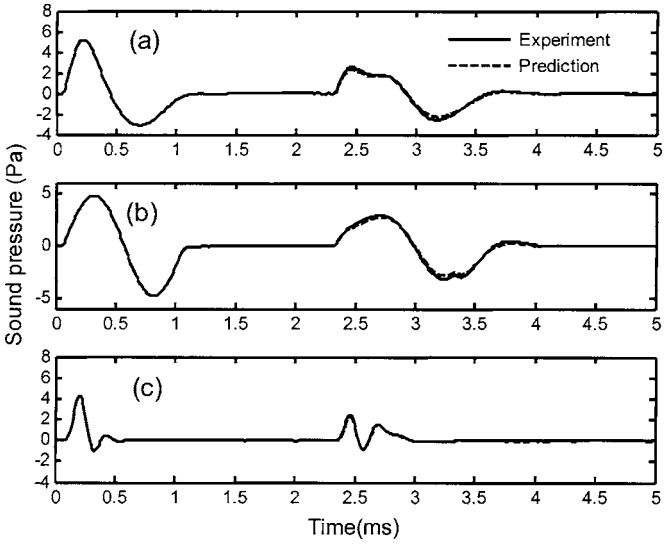


FIG. 5. Comparison of predicted and measured sound pulses reflected from the Helmholtz resonator of Fig. 4 on the (a) three-pole, (b) sine, and (c) Butterworth incident pulse.

As we know, a sound-absorbing structure in resonance has the locally lowest acoustic reflection or highest absorption coefficient. For the 20-mm cavity Helmholtz resonator, the one identified DHO is clearly associated with the minimum of the reflection coefficient and so are the three identified DHOs for the 60-mm cavity Helmholtz resonator. In fact, the real parts of the $\lambda_k/2\pi$ of these DHOs approximate respectively the frequencies where the minima of the reflection coefficient are found with minor differences attributable to the imaginary parts. Therefore, in these and the MPP case in Sec. III A, the association between the identified DHOs and the intrinsic resonances of an impedance boundary is clear.

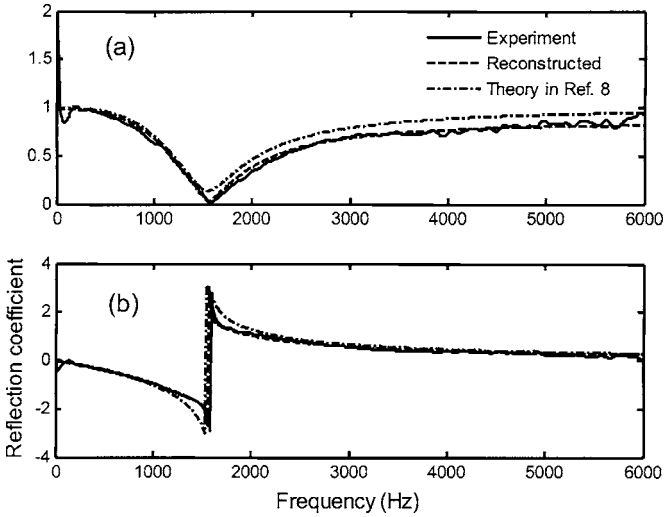


FIG. 4. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the Helmholtz resonator with hole diameter 0.3 mm, plate thickness 0.15 mm, porosity 0.785%, and cavity depth 20 mm.

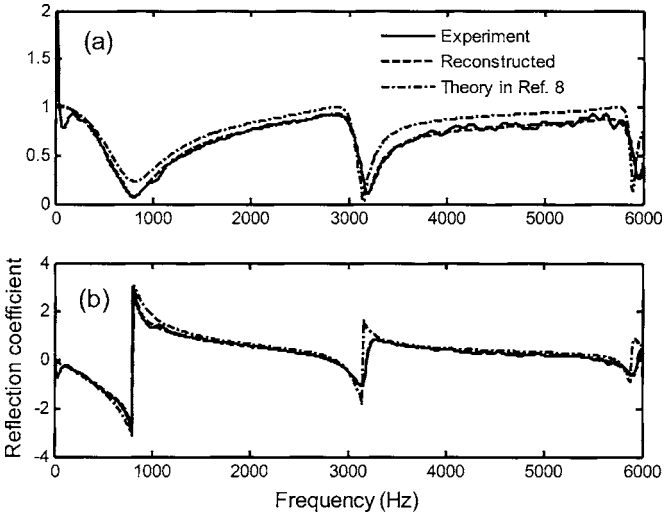


FIG. 6. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the Helmholtz resonator with hole diameter 0.3 mm, plate thickness 0.15 mm, porosity 0.785%, and cavity depth 60 mm.

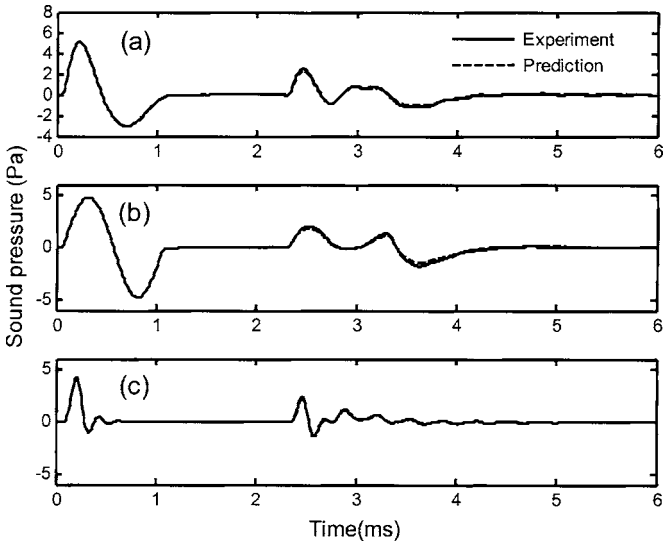


FIG. 7. Comparison of predicted and measured sound pulses reflected from the Helmholtz resonator of Fig. 6 on the (a) three-pole, (b) sine, and (c) Butterworth incident pulse.

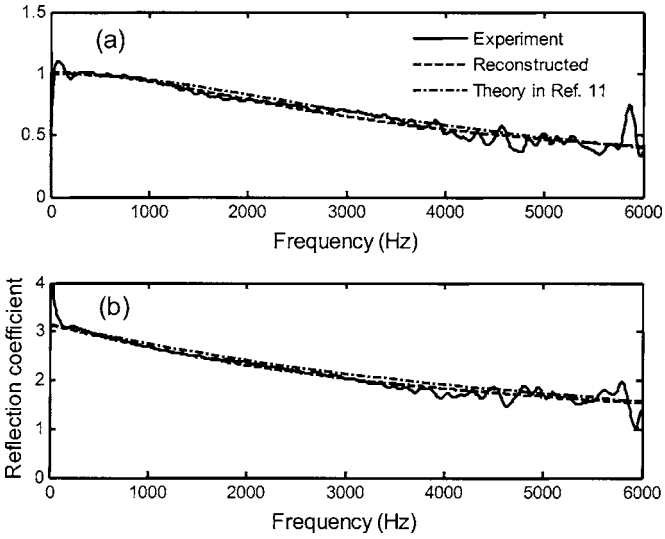


FIG. 8. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the flanged pipe exit.

2. Pipe exits

As a common acoustic element the pipe exit has long attracted much research attention. The radiation impedance of a circular pipe exit with an infinite flange has been derived by Morse and Ingard[11] and that of the unflanged one by Levine and Schwinger.[12] To the authors' best knowledge, these analytical works have so far no direct experimental confirmation in the open literature. For such open-end measurements the common impedance tube would be inapplicable due to its sensitivity to the external variances particularly at low frequency. The majority of experiments[13–15] are shock tube based, in which the involved flow nonlinearity prevents a direct comparison with the linear theories. Since the present time-domain impulse method has a finite extent in space and time and is not susceptible to external variances as frequency-domain methods, it is hoped that the following results on the pipe exit would complement these classical works.

According to Morse and Ingard,[11] the acoustic impedance of a pipe exit with an infinite flange is given as below,

$$Z(\bar{\omega}) = 1 - \frac{2J_1(\bar{\omega})}{\bar{\omega}} + j\frac{4}{\pi}\int_0^{\pi/2} \sin(\bar{\omega}\cos x)\sin^2 x\, dx, \quad (8)$$

where $\bar{\omega}$ is the angular frequency normalized by the sound speed and the pipe diameter. The reflection coefficient can be computed from Eq. (2). It has been shown in Ref. 5 that the low-frequency approximation of Eq. (8), $Z(\bar{\omega})=\frac{1}{8}\bar{\omega}^2 + j(4/3\pi)\bar{\omega}$, or the high-frequency approximation, $Z(\bar{\omega})=1 + j\pi/4\bar{\omega}$, renders a noncausal reflection coefficient. Nonetheless, it is shown in Appendix B that the exact reflection coefficient given by Eq. (8) is causal and can be cast as an infinite sum of DHOs.

Figure 8 not only shows that the measured reflection coefficient via the three-pole acoustic pulse agrees well with the theory of Morse and Ingard,[11] but also that the measured reflection coefficient of the flanged pipe exit can be well reconstructed with only one damped harmonic oscillator for the frequency range of 0–6000 Hz, or the normalized frequency range of 0–0.45. This result is consistent with the analysis in Appendix B, which gives a purely imaginary predominant pole at $\lambda_1=1.456j$ and residue of $A_1=-1.848$ in comparison with the corresponding measured values of $1.128j$ and -1.359 given in Table III, respectively. The limited frequency range of measurement prevents the identification of the higher-order poles whose normalized natural frequencies are above 7.906, much higher than the cut-on frequency of the higher-order modes in the pipe. The small deviations between the experimentally identified and exact values can be simply due to dropping all higher order poles in the approximation. Figure 9 further demonstrates that the reflected pulses can be very well predicted based on either the DHO of the three-pole incident pulse or when cross-checked with the sine and Butterworth incident pulses. Different from those Helmholtz-resonator impedance surfaces, the DHOs of the pipe exit are purely damped, i.e., no real-valued resonance frequency.

It is shown in Fig. 10 that the sharp-edged pipe exit has similar acoustic properties as those of the flanged one, but its damping factor, i.e., the imaginary part of the natural frequency, is larger than that of the flanged pipe exit (Table III), thus allowing more acoustic energy to reflect back. Also shown in the figure is that the measured and the recon-

TABLE III. Identified DHOs for the pipe exits.

Type of pipe exit	Natural frequencies (Hz), $\lambda_k/2\pi$	Complex amplitudes[a] (Hz), A_0 or $A_k/2\pi$ $(k\geq 1)$
Flanged	–	0.181
	$0.0+j2453.8$	-2918.7
Sharp-edged	–	0.232
	$0.0+j3209.7$	-4001.7
Round-edged	–	0.10
	$0.0+j2576.4$	-2946.6

[a]The values of A_0 correspond to the null natural frequency indicted by "-".

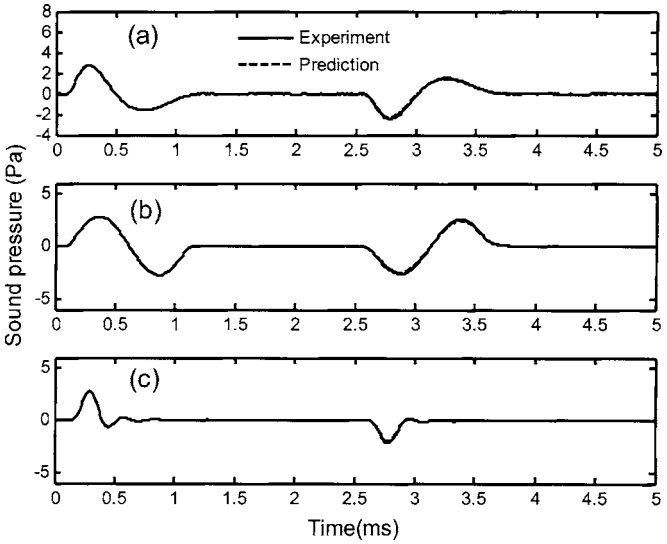


FIG. 9. Comparison of predicted and measured sound pulses reflected from the pipe exit of Fig. 8 on the (a) three-pole pulse, (b) sine pulse, and (c) Butterworth incident pulse.

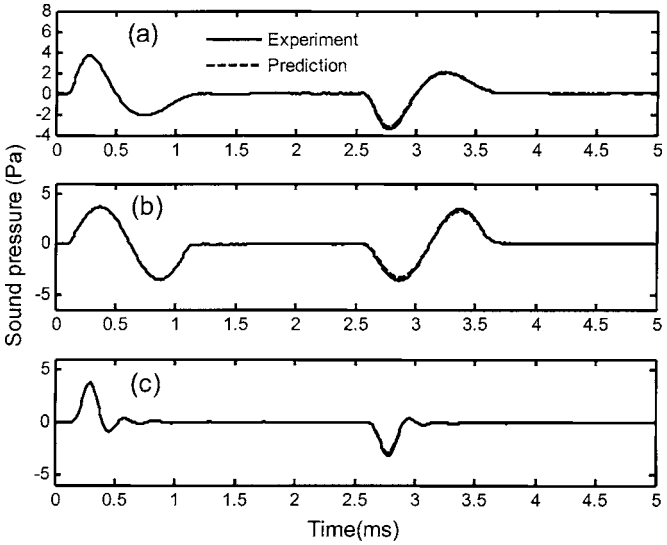


FIG. 11. Comparison of predicted and measured sound pulses reflected from the pipe exit of Fig. 10 on the (a) three-pole pulse, (b) sine pulse, and (c) Butterworth incident pulse.

structed reflection coefficient agree well with the theory of Levine and Schwinger[12] for the circular pipe exit of zero wall thickness. Figure 11 shows the result for the pulse reflection at the sharp-edged pipe exit. In Figs. 12 and 13 it is demonstrated that the acoustic properties of a round-edged pipe exit can also be well represented by a purely damped harmonic oscillator, for which no analytical solution is known to the authors. From the identified DHO descriptor given in Table III, we can see that the acoustic properties of the round-edged pipe exit are quite close to those of the flanged one.

In terms of the DHO description of its acoustic properties, we can conclude that a pipe exit, unlike a Helmholtz resonator, is basically a nonresonant acoustic element at least in the case below the cutoff frequency.

IV. CONCLUSIONS

We have shown here that common sound-absorbing materials and structures possess the characteristics of composite DHO, which is a simpler and physically more meaningful descriptor than the conventional impedance in the following ways:

(1) Within the frequency range of practical interest, the reflection coefficient of an impedance boundary can be well represented by the sum of only a few DHOs, bearing the intrinsic resonant characteristics as the natural frequencies of the Helmoltz resonator structures or the nonresonant pure dampings of pipe exits. Unlike the conventional frequency-domain characterization and measurement of impedance, such as in an impedance tube, the DHO descriptor gives a reliable extended broadband description to the lowest frequencies where the classically measured impedance data are often unreliable due to sensitivity to error.
(2) This has been validated by the good agreement between

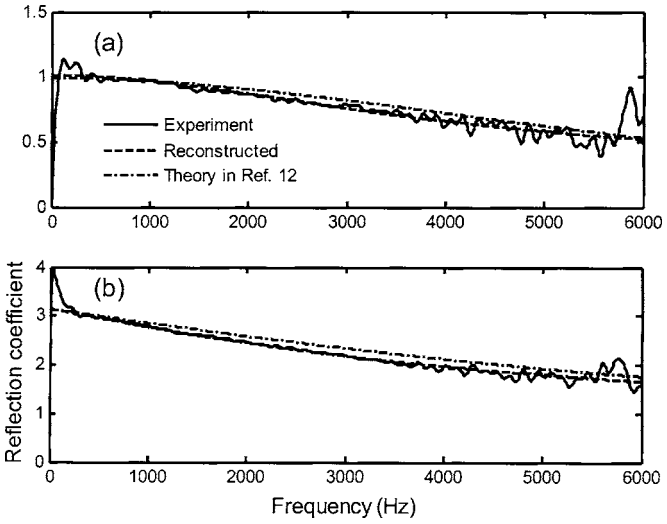


FIG. 10. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the sharp-edged pipe exit.

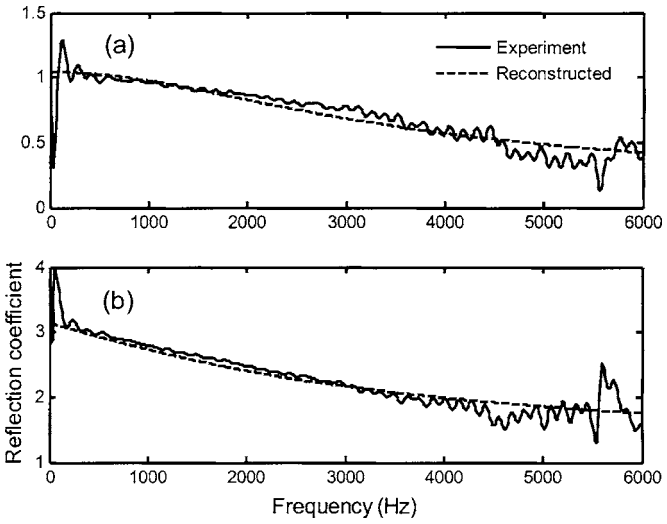


FIG. 12. (a) Magnitude. (b) Phase. Comparison of various presentations of the reflection coefficient for the round-edged pipe exit.

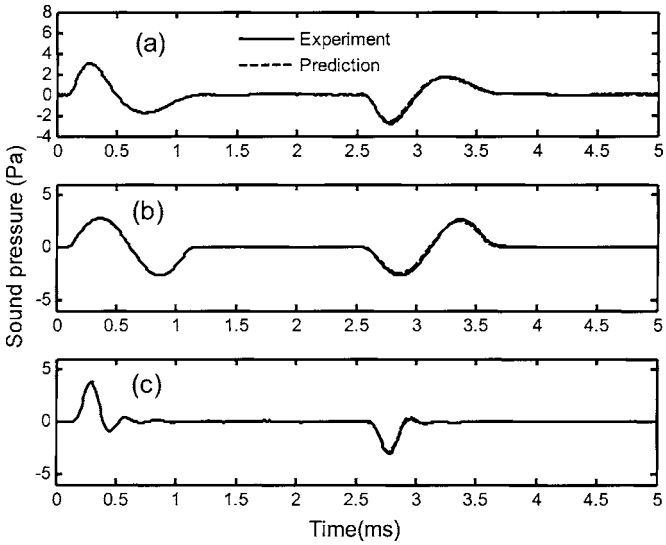


FIG. 13. Comparison of predicted and measured sound pulses reflected from the pipe exit of Fig. 12 on the (a) three-pole pulse, (b) sine pulse, and (c) Butterworth incident pulse.

measured and predicted pulses in the time- and frequency-domain and by various theories on a one-dimensional waveguide.

(3) We have yet encountered, so far, in our identification of DHOs of impedance boundaries, from either the measured pulses or the exact analytical solution of the reflection coefficient, a non-causal result.

(4) A set of causal DHOs is readily employed as a stable TDIBC for CAA.

Therefore, it is physically sound and numerically advantageous to characterize an impedance surface as a sum of DHOs.

ACKNOWLEDGMENTS

Financial support for the second author on the Hong Kong Research Grants Council's Competitive Earmarked Research Grant PolyU UGC earmarked Grant 5158/01E is gratefully acknowledged.

APPENDIX A: DESIRED WAVEFORMS OF THE SOUND PULSE IN EXPERIMENT

The three desired waveforms are described in detail as follows. The sine pulse is the simplest one and consists of a single-period sine wave. The three-pole pulse is synthesized from three pairs of damped sinusoids as below,

$$p(t) = H(t)\sum_{k=1}^{3} A_k \exp(j\lambda_k t) + A_k^* \exp(j\lambda_k^* t), \tag{A1}$$

where λ_k, $\lambda_k^* = \pm 2598.7 + j4394.7$, $\pm 7462.2 + j4223.7$, $\pm 13456.2 + j2635.4$ and correspondingly A_k, $A_k^* = 1.2002 \pm j2.6113$, $-1.2277 \mp j1.2012$, $0.027468 \mp j0.102788$. The Butterworth pulse is the impulse response of a sixth-order Butterworth low-pass filter whose cutoff frequency is 5 kHz. The duration of all the pulses is approximately 1 ms.

APPENDIX B: CAUSALITY OF THE REFLECTION COEFFICIENT OF FLANGED PIPE EXIT

Following Rudinger's[13] analysis, the first-order Bessel function is written in the following integral form:

$$J_1(\bar{\omega}) = \frac{2\bar{\omega}}{\pi}\int_0^{\pi/2} \cos(\bar{\omega}\cos x)\sin^2 x\, dx. \tag{B1}$$

The poles of $\bar{W}(\omega)$, $\lambda_k, \lambda_k^* = \pm\alpha_k + j\beta_k$, satisfy the relation

$$1 + Z(\bar{\omega}) = 0. \tag{B2}$$

Substituting Eqs. (8) and (B1) into the above equation, we get

$$\int_0^{\pi/2} e^{\beta_k \cos x} \cos(\alpha_k \cos x)\sin^2 x\, dx = \frac{\pi}{2} \tag{B3}$$

and

$$\int_0^{\pi/2} e^{\beta_k \cos x} \sin(\alpha_k \cos x)\sin^2 x\, dx = 0.$$

Assuming $\beta_k \leqslant 0$, we have $e^{\beta_k \cos x} \leqslant 1$ since $0 \leqslant x \leqslant \pi/2$. Thus, the following inequalities hold:

$$\left|\int_0^{\pi/2} e^{\beta_k \cos x} \cos(\alpha_k \cos x)\sin^2 x\, dx\right| \leqslant \int_0^{\pi/2} |\cos(\alpha_k \cos x)|\sin^2 x\, dx \leqslant \int_0^{\pi/2} \sin^2 x\, dx = \frac{\pi}{4}. \tag{B4}$$

The contradiction between Eqs. (B3) and (B4) indicates that all the poles of $W(\omega)$ must lie in the upper half of the complex ω plane, thus confirming the causality of Morse and Ingard's[11] analytical solution for the flanged pipe opening.

Equation (8) provides a direct means for solving Eq. (B2) for the poles of the reflection coefficient. Numerically, a Newton-Raphson search gives the first three pairs of normalized poles and residues at $\bar{\lambda}_k, \bar{\lambda}_k^* = 1.456j$, $\pm 7.906 + 3.510j$, $\pm 14.42 + 4.312j$ and $\bar{A}_k, \bar{A}_k^* = -1.848, -1.0 \pm 0j$, $-1.0 \pm 0j$.

[1] C. K. W. Tam and L. Auriault, "Time-domain impedance boundary conditions for computational acoustics," AIAA J. **34**(5), 917–923 (1996).
[2] Y. Ozyoruk, L. N. Long, and M. G. Jones, "Time-domain numerical simulation of a flow-impedance tube," J. Comput. Phys. **146**(1), 29–57 (1998).
[3] Y. H. Berthelot, "Surface acoustic impedance and causality," J. Acoust. Soc. Am. **109**(4), 1736–1739 (2001).
[4] K.-Y. Fung and H. Ju, "Broadband time-domain impedance models," AIAA J. **39**(8), 1449–1454 (2001).
[5] H. Ju and K.-Y. Fung, "Time-domain impedance boundary conditions with mean flow effects," AIAA J. **39**(9), 1683–1690 (2001).
[6] L. B. Felsen, *Transient Electromagnetic Fields* (Springer-Verlag, Berlin, 1976).
[7] K. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," IEEE Trans. Autom. Control **8**, 56–58 (1963).
[8] D. Y. Maa, "Microperforated-panel wideband absorbers," Noise Control Eng. J. **29**(3), 77–84 (1987).
[9] M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[10] X. Jing and K.-Y. Fung, "Generation of desired sound impulses," J. Sound Vib. (to be publsihed).

[11]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[12]H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," Phys. Rev. **73**(4), 383–406 (1948).
[13]G. Rudinger, "On the reflection of shock waves from an open end of a duct," J. Appl. Phys. **26**(8), 981–993 (1955).
[14]J. Brown and A. Vardy, "Reflection of pressure waves at tunnel portal," J. Sound Vib. **173**(1), 95–111 (1994).
[15]H. D. Kim and T. Setoguchi, "Study of the discharge of weak shocks from an open end of a duct," J. Sound Vib. **226**(5), 1011–1028 (1999).

Steerable frequency-invariant beamforming for arbitrary arrays

Lucas C. Parra[a)]
Biomedical Engineering Department, City College of New York, New York 10031

(Received 9 September 2005; revised 27 March 2006; accepted 28 March 2006)

Frequency-invariant beamforming aims to parameterize array filter coefficients such that the spectral and spatial response profiles of the array can be adjusted independently. Solutions to this problem have been presented for specific sensor configurations often requiring a larger number of sensors. However, in practical applications, the number and location of sensors are often restricted. This paper proposes to find an optimal linear basis transformation that decouples the frequency response from the spatial response. A least-squares optimal basis transform can be computed numerically for arbitrary sensor configurations, for which typically no exact analytical solutions are available. This transform can be further combined with a spherical harmonics basis resulting in readily steerable broadband beams. This solution to broadband beamforming effectively decouples the array geometry from the steering geometry. Furthermore, for frequency-invariant beams, this approach results in a significant reduction in the number of beam-design parameters. Here, the method is demonstrated for an optimal design of far-field response for an irregular linear array with as few as three sensors. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197606]

PACS number(s): 43.60.Fg [EJS] Pages: 3839–3847

I. INTRODUCTION

An array of spatially distributed sensors can be made selective in space and frequency by filtering and summing the output of multiple sensors. For a fixed geometry, the spectro-spatial response profile is determined by the filter coefficients. Changing a given coefficient will typically affect both the frequency as well as the spatial response profile. This spectro-spatial coupling complicates broadband filter design as well as adaptive beamforming algorithms. The goal of this work is to find a parameterization of the filter coefficients that decouples the spatial selectivity from the frequency selectivity for arbitrary array configurations. Once decoupled, a frequency-invariant response is obtained by choosing the same coefficients for multiple frequencies. This simplifies broadband beamforming as the frequency response and the spatial profile can be adjusted independently.

Broadband and frequency-invariant beamforming has been addressed by Ward *et al.*,[1] covering far-field problems, and Abhayapala *et al.*[2] and Kennedy *et al.*[3] for near-field problems. Their approach is based on the spatial Fourier transform of a continuous aperture. In practical implementations, the aperture needs to be sampled with a discrete number of sensors. This leads to specific array configurations typically with a large number of sensors on a linear or rectangular lattice.[2,4,5] More recent work presents analytic inversion methods for linear,[6] cylindrical,[7] and spherical arrays.[8] Although these array configurations may be optimal in terms of frequency invariance, reduced aliasing, or spatial resolution, they may not be practical in some applications. In particular, speech acquisition with embedded microphones requires broadband arrays often with a very small number of microphones (two to five) in a constrained spatial arrangement. This paper presents a numerical approach to construct an optimal frequency-invariant response for an arbitrarily chosen array configuration.

Following previous work,[8] the resulting frequency-invariant response is made steerable by combining it with the spherical harmonics decomposition of the beam pattern.[2,9] The required coefficients for rotating spherical harmonics are given by the Wigner rotation matrix.[10] A numerical approach related to the present method has been considered in the context of near-field design but has not been fully developed.[3] A related technique has been presented for arbitrarily placed sensors on a sphere but without addressing frequency invariance.[11]

The paper is organized as follows: Secs. II and III introduce the notation and define the goal of frequency-invariant beamforming. Sections IV and V present existing solutions based on analytic expansions of plane waves and the corresponding inversion formulas. The numerical least-squares approach proposed in this work is presented in Sec. VI. Beam steering is discussed in Sec. VII giving some special considerations to the linear array. Section VIII discusses sensor noise and the resulting regularization of the least-squares solution. Section IX explains how the proposed basis transformation can be used to make existing adaptive beamforming algorithms frequency invariant. Finally, examples are presented using data-independent beam design for a linear array. The paper closes with a discussion on how the proposed method can be applied to other array configurations, directional sensors, and near-field beam design. To better follow the main argument, the reader may skip Secs. IV, V, VII A, and VIII in a first reading.

II. ARRAY RESPONSE

This section defines some terminology commonly used with sensor arrays and beam forming. Denote the signal sampled by the nth sensor at time t as $x_n(t)$. A beamformer

[a)]Electronic mail: parra@ccny.cuny.edu

convolves the N sensor signals with the corresponding filter coefficients $c_n(t)$ and sums the result to generate an array output signal:

$$y(t) = \sum_{n=1}^{N} c_n(t) * x_n(t). \tag{1}$$

The convolution is denoted here with the symbol $*$.

For far-field design, it is customary to discuss the effect of this processing by considering the signals produced by a planar wave impinging on the sensors. Assume a steady-state plane wave with radial frequency ω traveling in direction $\Omega=(\vartheta,\varphi)$, where spherical angles ϑ and φ represent elevation and azimuth, respectively. Orientation and frequency can also be specified in Cartesian coordinates as a three-vector $\mathbf{k}$ pointing in direction Ω with $k=\|\mathbf{k}\|$ characterizing the wavenumber $k=\omega/c$. Because of the simple linear relationship, the wavenumber k will also be referred to as "frequency" in this paper. The plane wave elicits, at location $\mathbf{r}_n$, a pressure signal:[9]

$$x_n(t) = e^{i(\mathbf{k}\cdot\mathbf{r}_n-\omega t)}, \tag{2}$$

with $i^2=-1$. To obtain the angle-dependent frequency response, consider the temporal Fourier transform $\mathcal{F}$ of this pressure signal:

$$x_n(\nu) = \mathcal{F}\{x_n(t)\} = 2\pi\delta(\nu-\omega)e^{i\mathbf{k}\cdot\mathbf{r}_n}. \tag{3}$$

This is the response of the sensors to the plane wave which defines the *sensor response* $g_n(\mathbf{k})$:

$$g_n(\mathbf{k}) = e^{i\mathbf{k}\cdot\mathbf{r}_n}. \tag{4}$$

The δ-function expresses the monochromatic nature of the plane wave, and is typically omitted to facilitate a more compact notation.[9] Similarly, the Fourier transform of the resulting response of the filter and sum:

$$y(\nu) = \mathcal{F}\{y(t)\} = 2\pi\delta(\nu-\omega)f(\mathbf{k}), \tag{5}$$

leads to the definition of the *filter-array response* $f(\mathbf{k})$:

$$f(\mathbf{k}) = \sum_{n=1}^{N} c_n(k)g_n(\mathbf{k}). \tag{6}$$

III. FREQUENCY-INVARIANT BEAMFORMING

Now consider Eq. (6) with $\mathbf{k}$ rewritten in terms of the *arrival direction* Ω and frequency k,

$$f(k,\Omega) = \sum_{n=1}^{N} c_n(k)g_n(k,\Omega). \tag{7}$$

This equation can be seen as a parameterization of the filter-array response for each frequency k with coefficients $c_n(k)$, and basis functions $g_n(k,\Omega)$. Modifying coefficients $c_n(k)$ will affect the frequency and spatial response simultaneously because $g_n(k,\Omega)$ depends on both the frequency k and arrival direction Ω. The goal of *frequency-invariant beamforming* is to find a new parameterization for $c_n(k)$:

$$c_n(k) = \sum_{l=1}^{L} b_{nl}(k)\tilde{c}_l(k), \tag{8}$$

such that the basis transform $b_{nl}(k)$ converts the array response into a frequency-invariant array response,

$$\sum_{n=1}^{N} g_n(k,\Omega)b_{nl}(k) = \tilde{g}_l(\Omega). \tag{9}$$

The basis transform replaces the N sensors indexed by n by a new set of L *virtual* sensors indexed by l. These virtual sensors are now frequency invariant. This basis transform, in turn, factorizes the filter-array response, which is seen by combining Eqs. (7)–(9):

$$f(k,\Omega) = \sum_{l=1}^{L} \tilde{c}_l(k)\tilde{g}_l(\Omega). \tag{10}$$

By comparing Eq. (10) with Eq. (7), it becomes clear that modifying the new parameters $\tilde{c}_l(k)$ will affect the spatial response profile $f(k,\Omega)$ only at frequency k. In fact, the spatial response profile of the filter array is fully determined by coefficients $\tilde{c}_l(k)$. In particular, a frequency-invariant beamformer is obtained by choosing the same coefficients for all frequencies, $\tilde{c}_l(k)=\tilde{c}_l$.

The challenge of uncoupling the spatial from the spectral response lies in finding a basis transform $b_{nl}(k)$ that satisfies Eq. (9)—even if only approximately.

IV. ARRAY RESPONSE FOR VARIOUS GEOMETRIES

To understand existing analytic solutions to this problem, it will be useful to express the sensor response [Eq. (4)] in terms of the arrival direction Ω. The plane wave response of a sensor located at $\mathbf{r}_n=(r'_n,\Omega'_n)$ can be expanded[10,9] as

$$g_n(k,\Omega) = 4\pi\sum_{l=0}^{\infty} i^l j_l(kr'_n) \sum_{m=-l}^{l} Y_l^m(\Omega)Y_l^{m*}(\Omega'_n), \tag{11}$$

where $j_l(kr)$ are spherical Bessel functions of the first kind, and $Y_l^m(\Omega)$ are spherical harmonics. The response of a spherical array is given by Eq. (11) with $r'_n=r'$. For a horizontally placed circular array with $\vartheta'_n=\pi/2$, and restricting to the horizon as arrival direction $\vartheta=\pi/2$, Eq. (11) becomes[7]

$$g_n(k,\varphi) = \sum_{l=-\infty}^{\infty} i^l J_l(kr')e^{il(\varphi-\varphi'_n)}, \tag{12}$$

where $J_l(kr)$ are Bessel functions of the first kind. In a vertically aligned linear array with $\vartheta'_n=0$, Eq. (11) becomes independent of φ, and is therefore symmetric about the array axis:[6]

$$g_n(k,\Omega) = e^{-ikr'_n\cos\vartheta}. \tag{13}$$

V. ANALYTIC INVERSION APPROACHES

Let us now consider existing analytic solutions to the factorization problem (9). A general theory for continuous apertures has been proposed[1] with corresponding approximations for discrete arrays.[1,5] Some specific array configura-

tions make use of elegant analytic inversion formulas, as in the case of spherical,[8] hemispherical,[12] circular,[13,7] linear,[6] and rectangular[4] arrays. Equations (11)–(13) were presented to give an indication of possible approaches. To give an idea of these methods, consider the case of a circular array. For a circular array, one may choose:

$$b_{nl}(k) = N^{-1} i^{-l} J_l^{-1}(kr) e^{il\varphi_n}. \tag{14}$$

Inserting Eqs. (14) and (12) into Eq. (9) will indeed give a frequency-invariant function of space, $\tilde{g}_l(\varphi) = e^{-il\varphi}$, assuming that the following orthogonality condition holds for any l, m:

$$\frac{1}{N}\sum_{n=1}^{N} e^{i(m-l)\varphi_n} = \delta_{ml}. \tag{15}$$

Unfortunately, this is only correct for $|l| < N$, and only if the N sensors are placed on an equidistant lattice along the circular array. As a result, Eq. (9) is only approximately correct. For circular, as well as spherical, arrays the approximation can only be improved with a larger number of sensors, while the lattice must be carefully arranged to match the analytic inversion formulas.[13,8] The same is true for linear and rectangular arrays.[1,5] In addition, for those configurations, the angular response profile at any given frequency only partially determines the required spatial Fourier basis coefficients.[6] Arbitrary Fourier coefficients have to be chosen outside of the determined range, further compromising the accuracy of the approximations.

VI. LEAST-SQUARES SOLUTION

The goal of this work is to directly minimize the approximation error resulting from the restricted number of sensors, and to overcome the restrictions on sensor locations imposed by analytic inversions. The proposed solution is to invert Eq. (9) numerically. To this end, discretize the arrival directions with angles Ω_q, $q = 1, \ldots, Q$, and write Eq. (9) in matrix notation

$$\mathbf{G}(k)\mathbf{B}(k) = \tilde{\mathbf{G}}, \tag{16}$$

where $[\mathbf{G}(k)]_{qn} = g_n(k, \Omega_q)$, $[\mathbf{B}(k)]_{nl} = b_{nl}(k)$, and $[\tilde{\mathbf{G}}]_{ql} = \tilde{g}_l(\Omega_q)$. The new desired spatial basis vectors can be defined in the columns of $\tilde{\mathbf{G}}(k)$, while $\mathbf{G}(k)$ is determined by the array configuration. For each frequency, k, Eq. (16) specifies LQ conditions with LN unknowns. Typically, only a few sensors are available; yet one would like to parameterize the response for many different arrival directions. Therefore, with $Q > N$, the problem is overdetermined. The *least-squares* solution to this problem (i.e., the $\mathbf{B}$ that will reproduce $\mathbf{G}$ with the smallest square error) is computed with the pseudo-inverse, $\mathbf{G}^\dagger = (\mathbf{G}^H\mathbf{G})^{-1}\mathbf{G}^H$,

$$\mathbf{B}(k) = \mathbf{G}^\dagger(k)\tilde{\mathbf{G}}. \tag{17}$$

This requires that $\mathbf{G}$ is of full rank—a condition that has to be verified in practice. Note that the discretization of the arrival directions is only used to compute the basis transform $b_{nl}(k)$. When applying the new basis, the arrival direction can take on any arbitrary values. For a given array geometry, the basis transform is computed once for each frequency, k, and remains unaltered by subsequent beam design.

The accuracy of the proposed method depends on how well $\mathbf{G}$ can be inverted, that is, how well can $\tilde{\mathbf{G}}$ be represented by $\mathbf{B}$: $\varepsilon = \|\mathbf{G}\mathbf{B} - \tilde{\mathbf{G}}\|^2 = \|\mathbf{G}\mathbf{G}^\dagger\tilde{\mathbf{G}} - \tilde{\mathbf{G}}\|^2$. This is the minimum attainable square error in satisfying the factorization condition (9) for a specific set of angles Ω_q. Any other basis transform is suboptimal (in the least-squares sense).

Uniform sampling of angles Ω gives equal weight to the optimality criteria ε. Nonuniform sampling will place a bigger weight on the areas in angular space that are sampled more densely. To avoid large errors for intermediate angles that have not been sampled, one should choose band-limited basis functions $\tilde{\mathbf{G}}$. Spherical harmonics represent such a basis and will be presented next.

VII. BEAM STEERING

Ideally, the new spatial basis, $\tilde{g}_l(\Omega)$, should be easily steerable. This means that the overall orientation of a beam can be rotated without changing its spatial profile, i.e., without changing coefficient $\tilde{c}_l(k)$. This implies that after rotation of the frame of reference, the basis can be expressed in terms of the same basis. The basis of spherical harmonics satisfies this property;

$$Y_l^{m'}(\Omega') = \sum_{m=-l}^{l} D^l_{mm'}(\alpha, \beta, \gamma) Y_l^m(\Omega). \tag{18}$$

Ω and Ω' are the spherical angles before and after the rotation of the frame of reference. The rotation can be specified by the Euler angles α, β, γ, where γ is an initial spin about the original z axis, β changes its elevation, and α is a subsequent change in longitude. Explicit expressions for coefficients $D^l_{mm'}(\alpha, \beta, \gamma)$ were first given by Wigner[10,14] and are variably referred to as Wigner D-functions or Wigner rotation matrix. Efficient algorithms are available to compute these coefficients from a conventional 3×3 rotation matrix given in Cartesian coordinates.[15]

An additional advantage of the spherical harmonics is their uniform resolution.[16] Equation (18) states that any rotated version of the spherical harmonics of order L can be represented exactly by harmonics of, at most, order L. This implies that band-limited beam patterns, which do not oscillate faster than π/L, can be represented exactly with harmonics of at most order L.[17,18]

Thanks to identity (18) the spherical harmonics are readily steerable and therefore are a natural choice for the new basis:[8]

$$\tilde{g}_{lm}(\Omega) = Y_l^m(\Omega). \tag{19}$$

The virtual sensors now require double index lm instead of just l. With the matrix notation adopted for Eq. (16), a pair lm indexes a column of matrix $\tilde{\mathbf{G}}$. Arranging the Wigner rotation coefficients as matrix $[\mathbf{D}(\alpha, \beta, \gamma)]_{lm,l'm'} = D^l_{mm'}(\alpha, \beta, \gamma)$, the basis is then rotated with

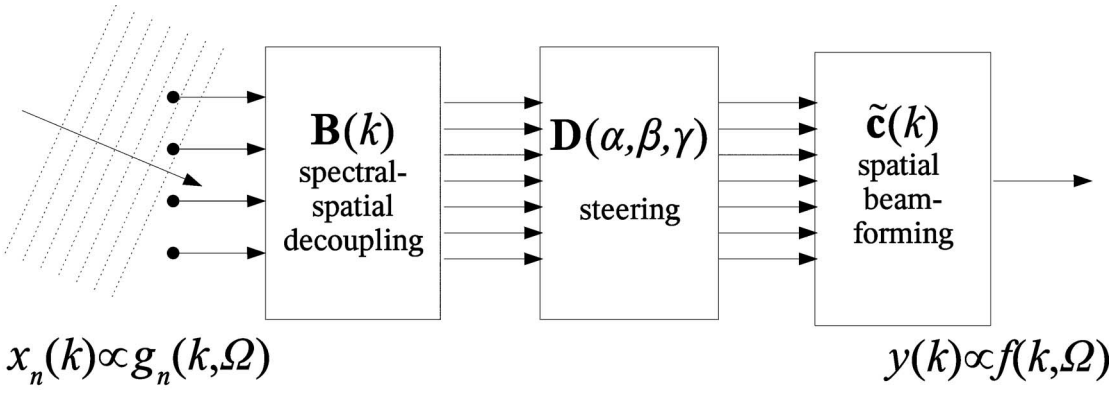


FIG. 1. Signal flow diagram.

$$\tilde{\mathbf{G}}' = \tilde{\mathbf{G}}\mathbf{D}(\alpha,\beta,\gamma), \tag{20}$$

and the basis transform that includes an arbitrary rotation is given by

$$\mathbf{B}'(k,\alpha,\beta,\gamma) = \mathbf{B}(k)\mathbf{D}(\alpha,\beta,\gamma). \tag{21}$$

To summarize the overall design structure, let us combine the previous equations in a compact form. The response of the array is now given by

$$f(k,\Omega) = \mathbf{g}(k,\Omega)\mathbf{B}(k)\mathbf{D}(\alpha,\beta,\gamma)\tilde{\mathbf{c}}(k). \tag{22}$$

The corresponding signal-flow diagram is shown in Fig. 1. Vector $\mathbf{g}(k,\Omega)$ is the response of the N sensors to a plane wave. $\mathbf{B}(k)$ uncouples the spatial response from the spectral response. $\mathbf{D}(\alpha,\beta,\gamma)$ steers the spatial response into an arbitrary direction; and $\tilde{\mathbf{c}}(k)$ defines its spatial profile for each frequency separately. Choosing the same coefficients $\tilde{\mathbf{c}}(k)$ for all frequencies yields an array response that is frequency invariant.

This structure uncouples the steering geometry from that of the array geometry. In previous approaches, the ability to steer seemed inevitably linked to the choice of array architecture and the corresponding analytic inversion formulas. The problem of steering the array has also been uncoupled from that of choosing its spatial profile, and finally, the design of the frequency response has been separated from that of the spatial response.

A. Steering in a linear array

For the special case of a linear array, the situation is complicated by the lack of full control over the two-dimensional beam pattern. In a vertical linear array, one can specify the response profile only in elevation ϑ. Rotating an axis-symmetric shape in elevation by β violates the axial symmetry around the z axis. Therefore, one cannot rotate a beam pattern by angle β and at the same time preserve symmetry in φ. As a result, there is no basis in ϑ that is isomorphic with respect to a shift. However, one can aim to find a transformation that approximately preserves a spatial profile defined in $\vartheta \in [0,\pi]$. One may choose, for instance, the Legendre basis and expand its shifted version in the same basis function set;

$$P_{l'}(\cos\vartheta') \approx \sum_{l=0}^{L} D_{ll'}(\beta)P_l(\cos\vartheta), \tag{23}$$

$$D_{ll'}(\beta) = \frac{2l+1}{2}\int_0^{\pi} P_{l'}(\cos(\vartheta-\beta))P_l(\cos\vartheta)\,d\cos\vartheta, \tag{24}$$

which defines a basis rotation or shift analogous to Eq. (18). The approximation results from the truncation of the sum at $L<\infty$. An optimal choice of basis would minimize this approximation error, implying that a shift of the basis can be expressed accurately within the same order ($l \leqslant L$). Therefore, in an ideal basis the shift matrix, $[\mathbf{D}(\beta)]_{ll'} = D_{ll'}(\beta)$, will have triangular structure such as in the case of spherical harmonics.[16] Figure 2 shows that the shift matrix $\mathbf{D}(\beta)$ for the Legendre basis is, in fact, approximately triangular resulting in small truncation errors. Therefore, for the linear array we suggest to use,

$$\tilde{g}_l(\vartheta) = P_l(\cos\vartheta). \tag{25}$$

Using the appropriate definitions for matrix $\mathbf{D}(\beta)$ based on Eq. (24), one can write the corresponding equations for this *approximate shift* of the beam pattern as in Eqs. (20) and (21).

In principle, there should be a closed-form solution to integral (24). However, in practice, it may be more efficient to evaluate the integral once (perhaps numerically) for each desired shift angle β, and store the L^2 coefficients for later use. To evaluate the Legendre coefficients of a general function, $f(z)$, numerically one may discretize the angle with Q samples, $z_q = \cos\vartheta_q$, and convert the integral into a discrete sum. This can be expressed efficiently in matrix-vector notation by defining matrix $[\mathbf{P}]_{ql} = P_l(z_q)$, and vector $[\mathbf{f}]_q = f(z_q)$. The Legendre coefficient for function $f(z)$ arranged as vector $[\tilde{\mathbf{c}}]_l = \tilde{c}_l$ are computed with

$$\tilde{\mathbf{c}} = \mathbf{P}^{\dagger}\mathbf{f}, \tag{26}$$

where the pseudo-inverse $\mathbf{P}^{\dagger}$ implements the integration sum and proper quadrature weights for any sampling of ϑ.[19] Transformation (26) is inverted in approximation with

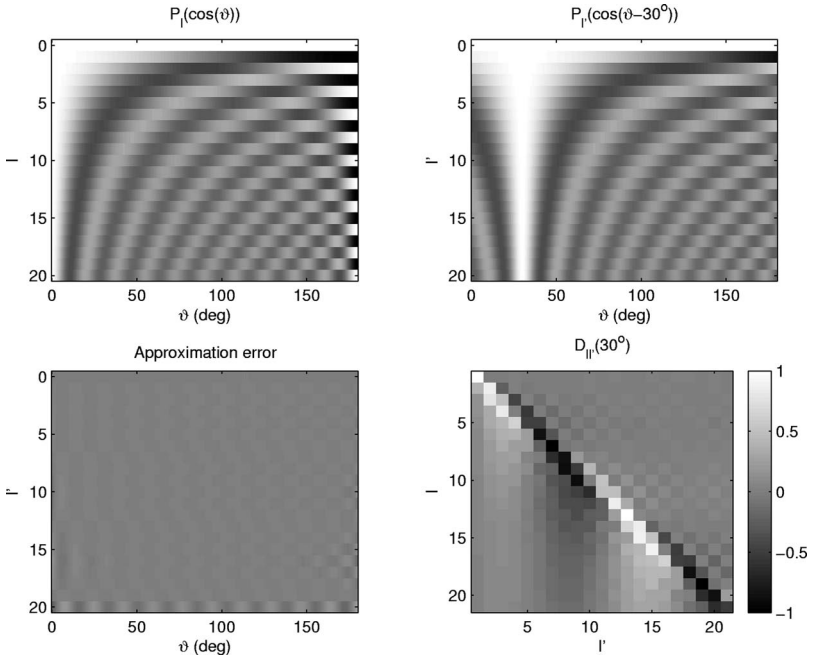


FIG. 2. Legendre basis functions and their rotations. Top two panels show Legendre basis functions as intensity images up to order $L=20$. Left and right panels represent $P_l(\vartheta)$ and $P_l(\vartheta-\beta)$. This example is for a rotation angle $\beta=30°$. Bottom left panel shows the approximation error in Eq. (23) due to truncation at $L=20$. Bottom right panel shows the corresponding matrix $D_{ll'}$ as defined in Eq. (24). All four panels use the same intensity gray scale.

$$\mathbf{f} \approx \mathbf{P}\tilde{\mathbf{c}}. \tag{27}$$

As with any pseudo-inverse, the approximation is exact if $\dim(\mathbf{f})=\dim(\tilde{\mathbf{c}})$, while the square error is minimized for $\dim(\mathbf{f})>\dim(\tilde{\mathbf{c}})$. With this notation, the shift matrix $\mathbf{D}(\beta)$ in Eq. (24) is,

$$\mathbf{D}(\beta)=\mathbf{P}^{\dagger}\mathbf{P}'(\beta), \tag{28}$$

where matrix $\mathbf{P}'$ captures the L Legendre polynomials evaluated at Q cosines shifted by β: $[\mathbf{P}'(\beta)]_{ql'}=P_{l'}(\cos(\vartheta_q-\beta))$.

VIII. EFFECTIVE BEAM CONTROL, MEASUREMENT NOISE, AND REGULARIZATION

When computing the basis $\mathbf{B}$ with Eq. (17), it was required that matrix $\mathbf{G}$ be of full rank. However, in practice, the matrix $\mathbf{G}^H(k)\mathbf{G}(k)$ is ill conditioned for the lowest frequencies and, therefore, cannot be accurately inverted when computing the pseudo-inverse $\mathbf{G}^{\dagger}(k)$. This is to be expected for very low frequencies with a large wavelength since a limited aperture prevents effective spatial resolution. Similarly, for high frequencies, the finite spacing of sensors generates aliasing side lobes, once again leading to a noninvertible $\mathbf{G}^H(k)\mathbf{G}(k)$ (recall the goal of using only a small number of sensors). This limitation is inherent to *any* beamforming design and is typically resolved by restricting the frequency band of operation or carefully choosing sensor spacing and array aperture. In the present numerical inversion approach, the instability leads to rather large gains which may arbitrarily magnify sensor noise. This section shows how this problem can be resolved by considering the effect of measurement noise.

Assume zero-mean wide-sense stationary additive sensor noise $w_n(t)$ with power spectrum $\sigma^2(k)$. The sensor signal in response to a plane wave is now (in the frequency domain)

$$x_n(k)=g_n(k,\Omega)+w_n(k). \tag{29}$$

With the appropriate definition of matrix $[\mathbf{W}(k)]_{qn}=w_n(k)$, Eq. (16) now becomes

$$(\mathbf{G}(k)+\mathbf{W}(k))\mathbf{B}(k)=\tilde{\mathbf{G}}. \tag{30}$$

Note that the signal of the steady-state plane wave $\mathbf{G}(k)$ is deterministic, while the noise $\mathbf{W}(k)$ is a random variable. To find the optimal basis, one now has to minimize the expected value over the random noise:

$$E[\|(\mathbf{G}(k)+\mathbf{W}(k))\mathbf{B}(k)-\tilde{\mathbf{G}}\|^2]. \tag{31}$$

For noise that is uncorrelated with the signal, this evaluates (omitting dependency on k) to the following trace:

$$\mathrm{Tr}(\mathbf{B}^H(\mathbf{G}^H\mathbf{G}+E[\mathbf{W}^H\mathbf{W}])\mathbf{B}-\mathbf{B}^H\mathbf{G}^H\tilde{\mathbf{G}}-\tilde{\mathbf{G}}^H\mathbf{G}\mathbf{B}). \tag{32}$$

This criterion is minimized by

$$\mathbf{B}=(\mathbf{G}^H\mathbf{G}+\mathbf{\Sigma})^{-1}\mathbf{G}^H\tilde{\mathbf{G}}, \tag{33}$$

which, as usual, is obtained by setting the derivative with respect to $\mathbf{B}^H$ equal to zero and solving for $\mathbf{B}$. If the noise is spatially uncorrelated and homogeneous in space, then the $N\times N$ matrix $\mathbf{\Sigma}(k)=E[\mathbf{W}(k)^H\mathbf{W}(k)]$ is diagonal with powers $\sigma_n^2(k)$ on the diagonal. This result is the conventional regularization of the pseudo-inverse. The examples presented in Sec. XI will use this regularization.

IX. APPLICATION TO ADAPTIVE BEAMFORMING DESIGN

The time domain output $y(t)$ of the filter array in response to sensor readings $x_n(t)$ is given by the convolution and sum:

$$y(t)=\sum_{n=1}^{N} c_n(t)*x_n(t)=\sum_{l=1}^{L}\tilde{c}_l(t)*\tilde{x}_l(t). \tag{34}$$

The second equality here results from Eq. (8) and the following definition of the *frequency-invariant virtual sensor* readings:

$$\tilde{x}_l(t) = \sum_{n=1}^{N} b_{nl}(t) * x_n(t). \tag{35}$$

This suggests that adaptive algorithms, that are driven by the sensor observations, can be applied to the signals of the newly defined frequency-invariant virtual sensors. Instead of optimizing filter parameters, $c_n(k)$, based on sensor readings, $x_n(t)$, the adaptive algorithm now adapts parameters $\tilde{c}_l(k)$ based on virtual sensor readings, $\tilde{x}_l(t)$. All known adaptive beamforming algorithms, such as generalized sidelobe canceling, blind source separation, and others are, therefore, immediately applicable without further modification.

To find an optimal frequency-invariant response, the adaptive algorithm will now optimize the parameters $\tilde{c}_l$. If the algorithm is based on a gradient of some cost function, $J(\{\tilde{c}_l(k)\})$, with a set of frequency dependent parameters, $\{\tilde{c}_l(k)\}$, the frequency-invariant gradient is then simply the original gradient summed over all frequencies

$$\frac{\partial J}{\partial \tilde{c}_l} = \sum_k \frac{\partial J}{\partial \tilde{c}_l(k)}. \tag{36}$$

Note that this has the potential to significantly reduce the number of free parameters as the same coefficients are used for all frequencies. For most adaptive algorithms, this will result in significant improvements in convergence speed as well as estimation accuracy. This advantage is in addition to the potential advantage of a frequency-invariant response.

X. APPLICATION TO IRREGULAR LINEAR ARRAY WITH LEAST-SQUARES BEAM DESIGN

The proposed method was implemented for a linear array of omnidirectional sensors and Legendre polynomials as the beamforming basis, i.e., Eqs. (13) and (25). Notice that Eq. (13) does not require equidistant sensor placement. The least-squares solution given by Eq. (17) is used to compute the basis transform. Choosing Legendre polynomials, as in Eq. (25), as the frequency-invariant virtual array response in Eq. (17) means that

$$\tilde{\mathbf{G}} = \mathbf{P}. \tag{37}$$

The effectiveness of the resulting basis is demonstrated by estimating parameters $\tilde{\mathbf{c}}$ that optimally reproduce a desired beam pattern. To this end, one could use a variety of beam design methods.[20] A simple data-independent method is the least-squares beamformer, which will be used here. Assume the prescribed response is specified as a vector $\mathbf{f}$ with coefficients f_d, each of which represents the desired response for angles θ_d, $d=1,\ldots,D$. The response of the array at those angles can be written in matrix notation as

$$\mathbf{f} = \mathbf{G}(\boldsymbol{\theta})\mathbf{B}\tilde{\mathbf{c}}, \tag{38}$$

where the coefficients of matrix $\mathbf{G}(\boldsymbol{\theta})$ are given by $[\mathbf{G}]_{dn} = g_n(\theta_d)$ specifying the response of the nth sensor for the desired angle θ_d. The goal is to find the coefficient $\tilde{c}$ that reproduces the desired response $\mathbf{f}$. Equation (38) does not include rotation. Instead, rotation is introduced after determining the optimal coefficients for a desired response. With N sensors, one can satisfy, at most, N conditions on the response. A larger number of conditions can only be satisfied approximately. The coefficients that reproduce $\mathbf{f}$ with the least-squares error are given by

$$\tilde{\mathbf{c}} = (\mathbf{G}(\boldsymbol{\theta})\mathbf{B})^{\dagger}\mathbf{f}. \tag{39}$$

Without going into detail, one should note that it is also useful to regularize this inverse, as in Eq. (33).

The Legendre basis uses L basis coefficients and can be rotated without having to recompute coefficients. One can compare this with the coefficient obtained in a *naive* frequency-invariant basis with $[\tilde{\mathbf{G}}]_{qm} = \delta_{qm}$ in Eq. (17). This basis requires Q coefficients, which is typically significantly larger than L, and the coefficients have to be recomputed when the response is to be rotated. For this "naive" basis, Eq. (39), simplifies to

$$\tilde{\mathbf{c}} = (\mathbf{G}(\boldsymbol{\theta})\mathbf{G}^{\dagger})^{\dagger}\mathbf{f}. \tag{40}$$

In Eqs. (38)–(40), the dependence on frequency was omitted for simplicity. The optimal parameters for each frequency are computed with Eq. (39) or (40). For a frequency invariant basis, one should use the same coefficients for all frequency bands. One option is to use the coefficients computed with Eq. (40) averaged across frequencies, $\tilde{\mathbf{c}} = 1/T\Sigma_k\tilde{\mathbf{c}}(k)$. Though suboptimal, this approach is not only simpler but in practice shows also better error behavior compared to the "optimal" solution, which would require combining all frequencies prior to computing the pseudo-inverses in Eq. (39) or (40). Simulations show that this averaging of coefficients across frequencies results in more evenly distributed deviations from the desired solutions as compared to the globally optimal solution.

Note that even after regularization, effective beam design is not possible at frequencies for which $\mathbf{G}^H\mathbf{G}$ is not invertible. One measure for the instability of the inverse (or rank deficiency of $\mathbf{G}$) is the condition number. One can use the condition number as a criterion to exclude frequency bands when computing the average $\tilde{\mathbf{c}}$. The examples described below assume acoustic sensors with a sound propagation speed of 342 ms^{-1}. For an aperture of 10 cm, one finds a useful frequency range (with condition number <150 dB) of at least 100–5000 Hz.

XI. EXAMPLES

Figure 3 shows the results obtained for a linear array of omnidirectional sensors with irregular spacing and an aperture of 10 cm. Arbitrary spacing of a small number of sensors ($N=3$ and $N=5$) was used to highlight the advantage of the present technique as compared to existing analytic methods, which typically require a larger number of sensors in a regular arrangement.

The figure shows that frequency invariance is reasonably well maintained up to the Nyquist frequency of 5000 Hz despite using only one set of coefficients (instead of separate coefficients for each of the $T/2+1$ frequency bins). Time

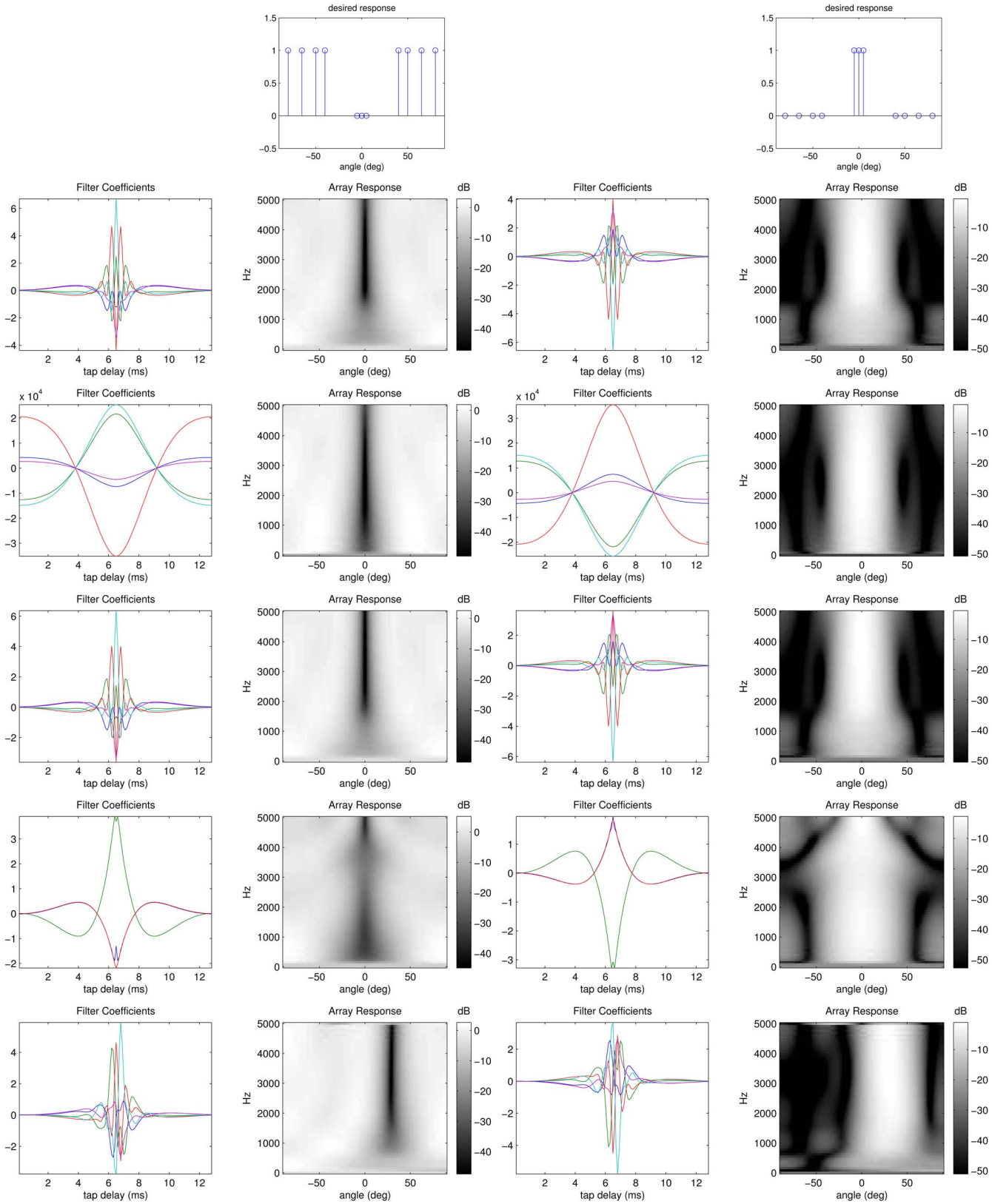


FIG. 3. (Color online) Frequency-invariant design of null beam (left) and main beam (right). Top panels show desired responses. Each following row compares results for different design parameters. First and third columns show time-domain filter coefficients $c_n(t)$. Second and third columns show magnitude response as function of frequency k and angle ϑ. Gray-scale represents response in dB. First row is obtained with a Legendre basis with five sensors assuming −30 dB noise. The same parameters are used for the following rows except that: Second row assumes zero noise; Third row uses naive basis; Fourth row uses only three microphones; Fifth row shifts the Legendre basis by 30° degrees.

domain coefficients were computed with $T=128$, which results in a reduction of the number of free parameters by a factor of 65.

In these examples, the Legendre basis included orders up to $L=20$, and the basis transform was computed with Eq. (33) using $Q=200$ equidistant samples ϑ_n. Compare this to

the third row showing the result with the naive basis. There is no significant difference to the Legendre basis (shown in the first row) despite the reduction of the number of parameters by a factor of $Q/L=10$. The Legendre basis therefore makes the array steerable and reduces the number of parameters without compromising accuracy. In practice, the order of the Legendre basis should be chosen depending on the desired angular bandwidth.

White noise with a power of −30 dB was assumed, except in the second row where no regularization was used, i.e., the basis was computed with Eq. (17). However, note that when the condition number exceeded 150 dB, those frequency bands were excluded regardless of the noise assumptions. Notice the significant scale increase in the filter coefficients indicating excessive low-frequency gain. Therefore, noise-based regularization should always be used in practice.

In these examples, $N=5$ sensors are located at $r_n=0$, 2, 5, 7, and 10 cm. Compare this with the fourth row that has only $N=3$ sensors located at $r_n=0$, 5, and 10 cm. This comparison indicates that the deviation from frequency invariance is primarily due to the limited number of sensors. Not surprisingly, as the number of sensors increases, frequency invariance is improved.

Finally, the last row demonstrates the effect of shifting the Legendre basis by an angle of $\beta=30°$ (as in Fig. 2) leading to a corresponding shift of the beam pattern.

Simulations with different sensor locations (not shown) indicate that the performance of the proposed method does not depend significantly on the specific arrangement of sensors. This is expected as the proposed method is guaranteed to make the best use of a fixed and known sensor configuration. However, as with conventional beam-design methods, some sensor arrangements are better suited to minimize aliasing and increase resolution. The present work did not aim to determine such optimal sensor locations. See, for instance, Ref. 21 for a modern technique to optimize locations.

Similarly, the sensitivity to errors in sensor position is comparable to conventional beam design. Simulations on the examples above (not shown) indicate only minor effects for location errors of about 1 mm (≈5% of the microphone spacing) but a significant deterioration for larger position errors. As with conventional beam design, it is preferable to use adaptive rather than data-independent methods as they can adapt to position errors. Section IX outlined how to implement adaptive design methods using the frequency-invariant basis.

XII. CONCLUSION

The previous section demonstrated the proposed method on a linear array. The implementation for a volumetric array is straightforward using definition (4) or expansion (11) for $\mathbf{G}$, and using spherical harmonics (19) to define matrix $\tilde{\mathbf{G}}$ analogous to matrix $\mathbf{P}$. The linear array was chosen here because the resulting frequency-invariant beam patterns are easier to visualize on paper, and because steering required special consideration as a result of the restricted linear geometry. No such complication should arise for a volumetric array since the Wigner rotations in Eq. (18) are exact. For a planar array, analogous considerations to those of the linear array may be necessary to cope with the symmetry across the array plane.

The presentation in this paper was restricted to omnidirectional sensors. However, the method applies equally well to sensors with directional response. If the sensor response is specified by $r(k,\Omega)$, then $g(k,\Omega)$ has to be replaced by $r(k,\Omega)g(k,\Omega)$ everywhere.

Finally, the notation here considers only the far-field response. One can generalize the argument to the near field by replacing plane waves of orientation Ω with point sources located at position $\mathbf{r}$. In that case, the array response to a planar wave, $g_n(k,\Omega)$, is to be replaced by the array response to a spherical wave,[9] $g_n(k,\mathbf{r})=e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_n)}/\|(\mathbf{r}-\mathbf{r}_n)\|$, and $\mathbf{r}$ has to be discretized over the desired range of point source positions when computing $\mathbf{B}$. The expansion for the spherical rather than a planar wave is given by Eq. (11) whereby the spherical Bessel function $j_l(kr_n')$ is replaced by $j_l(kr_n')/j_l(kr)$.[9] The rationale leading to a steerable beam design remains fully applicable.

ACKNOWLEDGMENTS

This work was initially motivated by the work of Meyer and Elko on frequency-invariant spherical arrays,[8] and the need to develop adaptive frequency-invariant beams for acoustic source separation.[22] I would also like to thank Walter Kellermann for useful comments, Chris Alvino for his careful review of multiple versions of this manuscript, and the anonymous reviewers of an earlier conference paper[23] who highlighted the need to address the issue of noise and stability.

[1] D. B. Ward, R. A. Kennedy, and R. C. Williamson, "Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns," J. Acoust. Soc. Am. **97**, 1023–1034 (1995).

[2] T. Abhayapala, R. Kennedy, and R. Williamson, "Nearfield broadband array design using a radially invariant modal expansion," J. Acoust. Soc. Am. **107**, 392–403 (2000).

[3] R. Kennedy, T. Abhayapala, and D. Ward, "Broadband near-field beamforming using a radial beampattern transformation," IEEE Trans. Signal Process. **46**, 2147–2156 (1998).

[4] W. Liu and S. Weiss, "A new class of broadband arrays with frequency invariant beam patterns," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 2004), Vol. **2**, pp. 185–188.

[5] D. B. Ward, R. A. Kennedy, and R. C. Williamson, "FIR filter design for frequency invariant beamformers," IEEE Signal Process. Lett. **3**, 69–71 (1996).

[6] T. Sekiguchi and Y. Karasawa, "Wideband beamspace adaptive array utilizing FIR fan filters for multibeam forming," IEEE Trans. Signal Process. **48**, 277–284 (2000).

[7] H. Teutsch and W. Kellermann, "EB-ESPRIT: 2D Localization of multiple wideband acoustic sources using eigenbeams," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 2005), Vol. **3**, pp. 89–92.

[8] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 2002), Vol. **2**, pp. 1781–1784.

[9] E. G. Williams, *Fourier Acoustics* (Academic, New York, 1999).

[10] A. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University Press, Princeton, N.J., 1957).

[11] Z. Li and R. Duraiswami, "A robust and self-reconfigurable design of spherical microphone array for multi-resolution beamforming," in *Proceedings of the International Conference on Acoustics, Speech, and Signal*

Processing (IEEE, New York, 2005), Vol. **4**, pp. 1137–1140.
[12]Z. Li and R. Duraiswami, "Hemispherical microphone arrays for sound capture and beamforming," in *Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE, New York, 2005), pp. 106–109.
[13]S. Chan and C. Pun, "On the design of digital broadband beamformer for uniform circular array with frequency invariant characteristics," in *Proceedings of the International Symposium on Circuits and Systems* (IEEE, New York, 2002), Vol. **1**, pp. 693–696.
[14]E. Wigner, *Gruppentheorie und ihre Anwendungen auf die Quantenmechanik der Atomspektren (Group theory and its applications to the quantum mechanics of atomic spectra)* (Friedr. Vieweg, and Sohn, Berlin, 1931).
[15]C. H. Choi, J. Ivanic, M. S. Gordon, and K. Ruedenberg, "Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion," J. Chem. Phys. **111**, 8825–8831 (1999).
[16]R. Jakob-Chien and B. K. Alpert, "A fast spherical filter with uniform resolution," J. Comput. Phys. **136**, 213–230 (1997).
[17]R. Duraiswami, Z. Li, D. Zotkin, E. Grassi, and N. Gumerov, "Plane-wave decomposition analysis for spherical microphone arrays," in *Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics* (IEEE, New York, 2005), pp. 150–153.
[18]P. N. Swarztrauber, "The spectral approximation of discrete scalar and vector functions on the sphere," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **16**, 934–949 (1979).
[19]P. N. Swarztrauber and W. F. Spotz, "Spherical harmonics projectors," Math. Comput. **73**, 753–760 (2003).
[20]B. Van Veen and K. Buckley, "Beamforming techniques for spatial filtering," *Digital Signal Processing Handbook* (CRC Press, New York, 1997), pp. 61.1–61.20.
[21]S. Blank and M. Hutt, "On the empirical optimization of antenna arrays," IEEE Antennas Propag. Mag. **47**, 58–67 (2005).
[22]W. Liu and D. Mandic, "Semi-blind source separation for convolutive mixtures based on frequency-invariant transformation," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (IEEE, New York, 2005).
[23]L. C. Parra, "Least squares frequency invariant beamforming," in *Proceedings of the Workshop on Applications of Signal Processing to Acoustics and Audio* (IEEE, New York, 2005).

Time reversal operator decomposition with focused transmission and robustness to speckle noise: Application to microcalcification detection

Jean-Luc Robert, Michael Burcher, and Claude Cohen-Bacrie
Philips Research USA, 345 Scarborough Road, Briarcliff Manor, New York 10510

Mathias Fink
Laboratoire Ondes et Acoustique, Université Paris 7, ESPCI, 10 rue Vauquelin, 75005 Paris, France

(Received 28 September 2005; revised 3 March 2006; accepted 3 March 2006)

The decomposition of the time reversal operator (DORT) is a detection and focusing technique using an array of transmit-receive transducers. In the absence of noise and under certain conditions, the eigenvectors of the time reversal operator contain the focal laws to focus ideally on well-resolved scatterers even in the presence of strong aberration. This paper describes a new algorithm, FDORT, which uses focused transmission schemes to acquire the operator. It can be performed from medical scanner data. A mathematical derivation of this algorithm is given and it is compared with the conventional algorithm, both theoretically and with numerical experiments. In the presence of strong speckle signals, the DORT method usually fails. The influence of the speckle noise is explained and a solution based on FDORT is presented, that enables detection of targets in complex media. Finally, an algorithm for microcalcification detection is proposed. In-vivo results show the potential of these techniques. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2190163]

PACS number(s): 43.60.Fg, 43.80.Qf, 43.20.Fn [TDM] Pages: 3848–3859

I. INTRODUCTION

Conventional medical ultrasound systems transmit pulses of sound into the body and map the envelope of returned echoes to form images. The resolution and the signal-to-noise ratio of these images rely on the ability to focus the transmitted pulse in order to maximize the transmitted field around the focal point. In conventional systems, the resolution is limited by diffraction.

In breast ultrasound imaging, as in many medical ultrasound applications, the heterogeneity of the breast degrades the quality of focusing. In particular, the speed of sound in the layer of subcutaneous fat is different from the speed of sound in other tissues. The result of propagation through a medium with variable speed of sound is a distortion of the wavefront and, therefore, a widening of the focal zone. Most of the time, the wavefront is assumed to be only time-delayed, and cross-correlation-based methods are used to find the delays.[1–3] Those studies show that the distortion can be significant. Then a time delay filter is used to correct for the aberration. The drawback of those methods is the assumption that the propagating medium only induces time-delay distortion. This is only the case if a thin aberrating layer on the transducer surface leads to the distortion. Experimental studies[2] show that the mean nearest-neighbor correlation coefficient in the breast is 0.8; thus, cross-correlation based methods are not optimal. Moreover, when there are multiple close scatterers, their wavefronts will interfere and a wavefront-following scheme will not lead to an accurate estimation of the aberrator.[4,5]

Another approach to automatically focus on a well resolved scatterer through heterogeneous media is the time reversal mirror.[6] However, in a set of well resolved targets, the time reversal mirror can focus only on the most reflective one.[7] This led to the development of a detection technique called the DORT (the French acronym for decomposition of the time reversal operator) method. This method is based on the decomposition of the transfer matrix that was introduced to describe the iterative time reversal process.[8,9] It requires a per-channel transmission and reception scheme (a pulse is sent with one transducer at a time and the echo is recorded on each transducer); this is known as a full data set. In the case of well separated point scatterers and with an acceptable level of noise, the number of significant eigenvalues has been shown[9] to be equal to the number of scatterers and the corresponding eigenvectors are the frequency Green functions of each scatterer, corresponding to a focusing on the scatterer. The eigenvalue is linked to the target reflectivity. However, DORT use is not limited to point scatterers. Studies[10] have also demonstrated the focusing properties of the time reversal operator's first eigenvector on deterministic extended objects. The advantage of DORT is that its focusing property is independent of any assumption on the aberrator, unlike the usual cross-correlation-based method.

The DORT method has shown great robustness as a detection and focusing method in media with aberration, but other interest arises. The method was not primarily designed to be an imaging method. However, the fact that the method is a frequency-domain process and the analogy between the time reversal matrix and the covariance matrix[11] used in passive detection in ocean acoustics or in radar opens the door to array processing methods developed for the latter. There is a large body of literature discussing matched field processing for ocean acoustics[12] and this analogy gives hope that it could be adapted to active array imaging modes like medical ultrasound. Among numerous algorithms, it is worth men-

 0001-4966/2006/119(6)/3848/12/$22.50

tioning MUSIC[11,13,14] (Multiple Signal Classification) which leads to a significant improvement in the resolution: Two wires whose distance was less than one-third of the point spread function, have been separated.[13] The classical resolution limit is half the point spread function. Image of objects have been reconstructed using the decomposition of the scattering operator, which is the far field limit of the Time Reversal Operator.[15]

The DORT method, then, is attractive for use in medical ultrasound imaging. The two main reasons that prevent using the conventional DORT method with medical ultrasound systems are the following: First, most of the ultrasound scanners used for the acquisitions are not very flexible and cannot collect a full data set. They usually transmit a focused beam using several transducers with the appropriate time delays. Second, the importance of speckle signals (sub-resolution scatterers) and the complexity of the scattering medium yield very noisy eigenvectors if the conventional DORT method is used.

Having in scope future medical applications such as aberration correction or array processing, this paper presents a modified DORT method that preserves all the interesting properties of the original method but is designed for data resulting from focused transmits and is efficient in media where speckle is important. We term this method Focused DORT (FDORT). The aim of this paper is not to focus only on medical applications but rather to show that FDORT offer similar results as DORT in theory, simulations, and experiments. As studies on DORT have mainly been conducted on point scatterer distributions,[8,9] it seems pertinent to use similar setups for comparison purposes and, therefore, most of this paper deals with point scatterers, although they are not prevalent in medical imaging. The interest of a DORT method with focused beams is not limited to medical ultrasound. Experiments have been done in underwater acoustics,[16] and other applications exist in nondestructive evaluation.

Section II addresses the problem of the transmit schemes. It begins with a short review of the principles of the conventional DORT method, with a full data set, mainly to set the formalism, and then establishes the theory of the FDORT method, based on focused transmit acquisitions. More generally the derivations are valid for any shape of the transmit. We address focused pulses as a particular case. Interesting results are summarized in Sec. II C, so that a reader not interested in the mathematical justification can begin at that point. Finally Sec. II D presents simulation results.

Section III discusses the influence of speckle noise on the DORT results. It begins with an explanation of the speckle influence, and proposes a solution based on FDORT to reduce its effect. Experimental results on a medical phantom illustrate the efficiency of the implementation.

Section IV presents an implementation of FDORT restricted to a small area that optimizes the sensitivity of detection of point scatterers in speckle and gives local information on the scattering medium. A method to differentiate point scatterers from extended scatterers or speckle, and thus improve point scatterer detection is proposed and illustrated by an application to *in vivo* microcalcification detection.

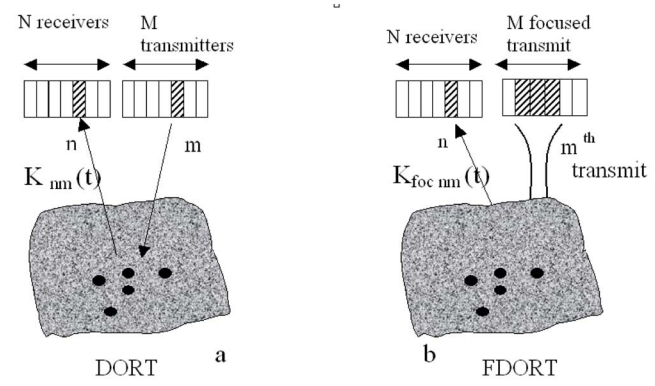


FIG. 1. (a) DORT: Each element of the transmit array is fired individually: Insonification of the medium by element m of the transmit array and reception of the corresponding echo by element n of the receive array gives $k_{nm}(t)$. (b) FDORT: Groups of elements of the transmit array are used to transmit a focused pulse; reception of the echo by element n of the receive array gives $k_{foc\,nm}(t)$.

II. EFFECT OF THE TRANSMIT SCHEME

A. The DORT method: Background theory

1. The transfer matrix and the time reversal operator

The theory of the DORT method has been thoroughly covered in the literature.[8,9,17] It is introduced here to set the formalism. The method is based on a matrix description of a transmit-receive process performed by an array of transducers,[8,9] or between two different arrays (an array of M transmitters and an array of N receivers).[17] Most of the DORT experiments are conducted using the same array in transmission and reception, but in order to introduce the modified method in Sec. II B, which uses focused pulses in transmission and per channel reception, we will consider in the following the general case of two different arrays.

If the system (propagation medium and electro-acoustic response) is linear and time-invariant, the process of transmitting and receiving can be described by a collection of filters: Each transmitting element m and each receiving element n are linked by an inter-element impulse response $k_{nm}(t)$, so that:

$$r_n(t) = e_m(t) \otimes k_{nm}(t), \tag{1}$$

where $r_n(t)$ is the signal received on the nth transducer when $e_m(t)$ is transmitted on the mth transducer, as seen in Fig. 1. Thus when the input to element m is a delta impulse, the output of element n is $r_n(t)=k_{nm}(t)$. The Fourier transform yields: $R_n(\omega)=K_{nm}(\omega)$. The repetition of the process for each pair (n,m) of transmitting and receiving elements leads, at a given frequency, to the transfer matrix $\mathbf{K}(\omega)$:

$$\mathbf{K}(\omega) = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1M} \\ K_{21} & K_{22} & & \\ \vdots & & \ddots & \\ K_{N1} & & & K_{NM} \end{bmatrix}$$

The matrix K describes the propagation between the transmit and the receive arrays and is therefore dependent on the scattering medium. If a signal $\mathbf{E}(\omega) = [E_1(\omega), E_2(\omega), \ldots, E_M(\omega)]^T$, where $E_m(\omega)$ is the input of

element m, at the frequency ω and T is the transpose (that transforms a $1^{\times}M$ row vectors into a $M^{\times}1$ column vector) is transmitted into the medium, the received echo is given by the following matrix formulation:

$$\mathbf{R}(\omega) = \mathbf{K}(\omega)\mathbf{E}(\omega), \tag{2}$$

where $\mathbf{R}(\omega)=[R_1(\omega),R_2(\omega),\ldots,R_N(\omega)]^T$, $R_n(\omega)$ being the signal received by the nth element of the receive array. $\mathbf{E}(\omega)$ and $\mathbf{R}(\omega)$ are vectors expressed in the transmit and receive basis, respectively, formed by the elements of the arrays. Reference to the frequency ω is omitted in the following. To perform time reversal, the receive (*Rx*) array has to transmit back the time reversed signal (corresponding to a phase conjugation in the frequency domain) $\overline{\mathbf{KE}}$ and the transmit (*Tx*) array has to receive it (during the time reversal the role of both arrays are exchanged). Because of the reciprocity theorem, the propagation from the receive to the transmit array is given by $\mathbf{K^T}$. The signal received by the *Tx* array at the end of the *time reversal cycle* is then $\mathbf{K}^T\overline{\mathbf{KE}}$, which is simply the conjugate of $(\mathbf{K^H K})\mathbf{E}$, where $\mathbf{H}$ stands for the Hermitian, or conjugate, transpose (transpose followed by conjugation), and $\mathbf{T_{Tx}}=\mathbf{K^H K}$ is defined as the time reversal operator expressed in the *Tx* array (*Tx* basis). We can similarly define the time reversal operator from the *Rx* array point of view: $\mathbf{T_{Rx}}=\mathbf{K K^H}$.

For all the results recalled here, there is no need for $\mathbf{K}$ to be either symmetric or square (which is obviously not the case if N is not equal to M) as it was in the earliest papers.[8] Then in general $\mathbf{K^H K}$ and $\mathbf{K K^H}$ are different, but both have the same rank, equal to the rank of $\mathbf{K}$.

As $(\mathbf{K^H K})^{\mathbf{H}}=\mathbf{K^H K}$ the time reversal operator is Hermitian positive in an orthogonal basis and thus can be diagonalized. Moreover, the eigenvalues are real and positive, and the eigenvectors are orthogonal.

The DORT method is based on the diagonalization of this time reversal operator. Section II A 2 describes the information given by the method in a simple case.

Practically, the diagonalization of the time reversal operator is not used. Indeed, it is mathematically equivalent to the singular value decomposition (SVD) of $\mathbf{K}$:

$$\mathbf{K} = \mathbf{U S V^H}, \tag{3}$$

where $\mathbf{S}$ is a $N\times M$ diagonal matrix completed by lines of zeros, containing the singular values of $\mathbf{K}$; $\mathbf{U}$ is a $N\times N$ unitary matrix whose columns are the eigenvectors of $\mathbf{K K^H}$ (the time reversal operator expressed in the receive basis); and $\mathbf{V}$ is the $M\times M$ matrix whose columns are the eigenvectors of $\mathbf{K^H K}$.

2. Case of isotropic, pointlike scatterers, and single scattering

a. Expression of the transfer matrix. In the case of isotropic point scatterers and under the Born approximation, the eigenvectors and eigenvalues of the time reversal operator can be theoretically calculated. For clarity, the number of point scatterers is assumed to be two. Let $\mathbf{P}$ and $\mathbf{Q}$ be the scatterers' positions. We denote by $\mathbf{H_{Rx}}(\mathbf{P})$ and $\mathbf{H_{Tx}}(\mathbf{P})$ the monochromatic Green functions of $\mathbf{P}$ expressed in the *Rx* and *Tx* bases, respectively. For example, $\mathbf{H_{Tx}}(\mathbf{P})$ is a $1^{\times}M$ vector and $H_{Tx}(\mathbf{P})_m$ describes the propagation between the ith element and $\mathbf{P}$. Let $D(\mathbf{P})$ and $D(\mathbf{Q})$ be the reflectivity of each scatterer. We also assume the absence of noise and we omit the acousto-electrical responses of the transducers, as they have no influence on the results.

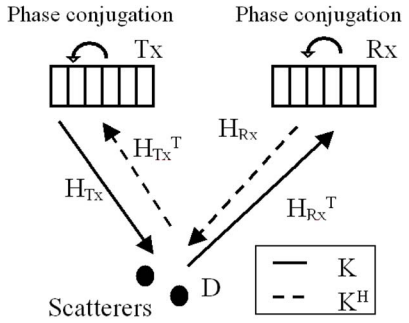


FIG. 2. The time reversal process can be seen as a full cycle between 3 *actors*: The *Tx* array, the *Rx* array and the scatterers. Propagation between the *actors* is described by $\mathbf{H_{Tx}}$ and $\mathbf{H_{Rx}}$. Reflection from the scatterers is equivalent to a multiplication by $\mathbf{D}$, and the backpropagation by the arrays is equivalent to a phase conjugation, included in the hermitian transpose $\mathbf{H}$. $\mathbf{K}$ describes the one-way propagation between *Tx* and *Rx*, represented by the solid line. The time reversal operator can be expressed mathematically from the point of view of any of these 3 *actors*. Then one has to start from the desired *actor* and make a full cycle.

A transmit-receive process can be divided into three stages, as seen in Fig. 2: Propagation from the *Tx* array to the scatterers, reflection on the scatterers, and finally propagation to the *Rx* array. The transmit-receive process between *Tx* element m and *Rx* element n is then: $\mathbf{K_{nm}}=H_{Tx}(\mathbf{P})_m D(\mathbf{P}) H_{Rx}(\mathbf{P})_{\mathbf{n}}+H_{Tx}(\mathbf{Q})_{\mathbf{m}} D(\mathbf{Q}) H_{Rx}(\mathbf{Q})_{\mathbf{n}}$ and finally, one can write the transfer matrix as the product of 3 terms:[8]

$$\mathbf{K} = (\mathbf{H_{Rx}})^{\mathbf{T}}\mathbf{D}\mathbf{H_{Tx}}, \tag{4}$$

where

$$\mathbf{H_{Rx}} = \begin{pmatrix} H_{\mathbf{Rx}}(\mathbf{P})_1 & H_{\mathbf{Rx}}(\mathbf{P})_2 & \cdots & H_{\mathbf{Rx}}(\mathbf{P})_N \\ H_{\mathbf{Rx}}(\mathbf{Q})_1 & H_{\mathbf{Rx}}(\mathbf{Q})_2 & \cdots & H_{\mathbf{Rx}}(\mathbf{Q})_N \end{pmatrix},$$

and

$$\mathbf{D} = \begin{pmatrix} D(\mathbf{P}) & 0 \\ 0 & D(\mathbf{Q}) \end{pmatrix}.$$

$\mathbf{H_{Tx}}$ has the same structure as $\mathbf{H_{Rx}}$; its rows are the Green functions expressed in the *Tx* array. It follows from Eq. (4) that the rank of $\mathbf{K}$ is equal to the number of scatterers.

b. Time reversal operator from the scatterers' point of view. In the receive basis, the time reversal operator becomes:

$$\mathbf{K K^H} = (\mathbf{H_{Rx}})^{\mathbf{T}}\mathbf{D}\mathbf{H_{Tx}}\mathbf{H_{Tx}^H}\overline{\mathbf{D}\mathbf{H_{Rx}}}. \tag{5}$$

Although experimentally we only have access to the time reversal operator expressed from the point of view of one of the arrays, in order to understand the properties of its eigenvectors, it is more convenient to express it from the point of view of the scatterers, in other words in the scatterers' basis. As depicted in Fig. 2, the time reversal process can be seen as a cycle. From the scatterers' point of view, the cycle begins at the scatterers location and is seen as follows: The scatterers emit an echo toward the *Rx* array, which back-

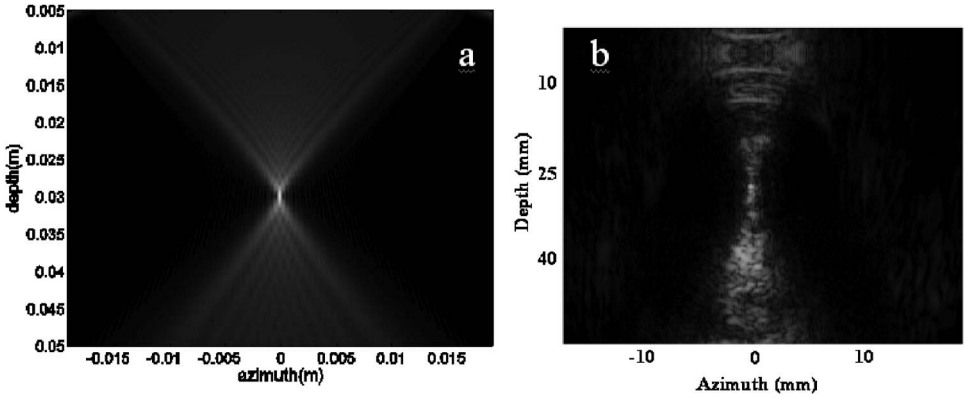


FIG. 3. (a) Magnitude of the field resulting from focusing on a scatterer ("the target") located at depth $z=30$ mm. The shadow of the transducer, in gray, is the coupling area. Any scatterer located in that area is coupled with the first one: The eigenvectors of the time reversal operator are then linear combination of the scatterers' Green functions. If the second scatterer is located outside of the shadow in the black area, the scatterers are well resolved and the Green functions are separated, each corresponding to one eigenvector. (b) Map of scatterers coupled to a given target ($x=0$, $z=25$ mm) in a medical image. The backscatter signal has been projected onto the target Green function before the beamforming and then only the scatterers whose Green functions are not orthogonal to the target's are imaged with an intensity proportional to the coupling.

propagates the echo toward the scatterers. The signal is reflected by the scatterers, received by the *Tx* array and backpropagated one more time toward the scatterers. Let $\mathbf{T_{scat}}$ be the time reversal operator in the scatterer basis; as there are two scatterers, $\mathbf{T_{scat}}$ is a 2^*2 matrix and $\mathbf{T_{scat},}_{ij}$ is the signal received by scatterer i after a full time reversal process when the initial echo was sent by scatterer j. $\mathbf{T_{scat}}$ is expressed as

$$\mathbf{T_{scat}} = \mathbf{D}\mathbf{H_{Tx}}\mathbf{H_{Tx}^H}\mathbf{D}\overline{\mathbf{H_{Rx}}}\mathbf{H_{Rx}^T}. \tag{6}$$

Furthermore,

$$\overline{\mathbf{H_{Rx}}}\mathbf{H_{Rx}^T} = \begin{pmatrix} \|\mathbf{H_{Rx}}(\mathbf{P})\|^2 & \langle \mathbf{H_{Rx}}(\mathbf{Q})|\mathbf{H_{Rx}}(\mathbf{P})\rangle \\ \langle \mathbf{H_{Rx}}(\mathbf{P})|\mathbf{H_{Rx}}(\mathbf{Q})\rangle & \|\mathbf{H_{Rx}}(\mathbf{Q})\|^2 \end{pmatrix}. \tag{7}$$

And a similar expression holds for $\mathbf{H_{Tx}}\mathbf{H_{Tx}^H}$. Because $\mathbf{D}$ is already diagonal, a condition for the expression of Eq. (6) to be diagonal is:

$$\langle \mathbf{H_{Rx}}(\mathbf{P})|\mathbf{H_{Rx}}(\mathbf{Q})\rangle = 0, \tag{8}$$

which means that $\overline{\mathbf{H_{Rx}}}\mathbf{H_{Rx}^T}$ is diagonal, and

$$\langle \mathbf{H_{Tx}}(\mathbf{P})|\mathbf{H_{Tx}}(\mathbf{Q})\rangle = 0, \tag{9}$$

which means that $\mathbf{H_{Tx}}\mathbf{H_{Tx}^H}$ is diagonal. In other words, the Green functions of the scatterers are orthogonal in both the transmit [Eq. (9)] and the receive [Eq. (8)] bases.

c. Physical interpretation. There is a physical interpretation of the scalar products from Eqs. (8) and (9). Using time-reversal arguments, transmitting $\bar{\mathbf{H}}_{\mathbf{Rx}}(\mathbf{Q})$ with the receive array results in focusing on the point $\mathbf{Q}$. The field received at the point $\mathbf{P}$ when transmitting such a signal is expressed as

$$\sum_{m=1}^{N} \bar{\mathbf{H}}_{\mathbf{Rx}}(\mathbf{Q})_m \mathbf{H_{Rx}}(\mathbf{P})_\mathbf{m} = \langle \mathbf{H_{Rx}}(\mathbf{P})|\mathbf{H_{Rx}}(\mathbf{Q})\rangle. \tag{10}$$

Thus the scalar product is equal to zero if it is possible to focus on one scatterer without sending energy to the other one, as illustrated in Fig. 3. The scatterers are then said to be well resolved or well separated.

d. Eigenvectors. The conditions of Eqs. (8) and (9) are satisfied if the scatterers are well resolved from the point of view of both arrays. In this case, Eq. (8) shows that the time reversal operator is diagonal in the scatterers' basis. The two eigenvectors of the time reversal operator expressed in the *Rx* basis [Eq. (5)] associated with nonzero eigenvalues are then the scatterers' Green functions, $\mathbf{H_{Rx}}(\mathbf{P})$ and $\mathbf{H_{Rx}}(\mathbf{Q})$ expressed in the *Rx* array. Identically, eigenvectors of the time reversal operator expressed in the *Tx* basis are the Green function in the *Tx* array. The SVD of K [Eq. (3)] gives the eigenvectors in both the *Rx* and *Tx* array. Moreover, the eigenvalue corresponding to the scatterer $\mathbf{P}$ is

$$\lambda = \|\mathbf{H_{Tx}}(\mathbf{P})\|^2 \|\mathbf{H_{Rx}}(\mathbf{P})\|^2 D(\mathbf{P})^2.$$

If the targets are not well resolved, there is a coupling between them through the nonzero diagonal terms of Eq. (7) and the eigenvectors are expressed as a linear combination of the Green functions.[9] The transmission of such an eigenvector does not lead to point focusing but DORT still provides useful information in this case. In particular, matched field methods like MUSIC can still be used.[13]

The assumption of isotropic scattering, used for the calculations, is true only for a point-like discontinuity in compressibility. Dipole scattering that accounts for discontinuities in density can be incorporated with additional complexity.[18]

B. The DORT method with focused pulse acquisitions (FDORT): Theory

1. The generalized transfer matrix

The transfer matrix for DORT is built on a full data set. In medical ultrasound, insonifications are usually obtained using a different process. Many elements of the array are fired with appropriate time delays (and apodization) to focus the energy at a specific location. The reflected signal, which largely originates from this region, is then recorded on the elements of the same array. The next pulse is sent using a different delay law, to focus the energy at a nearby location, and so on. At a given frequency, each transmitted pulse is expressed as a complex vector in the array basis. Assuming that M different focused pulses are transmitted during the acquisition, using a N-element array, let $\mathbf{B_m}$ be the $N^\times 1$ vector describing the mth transmitted pulse focusing at a position $\mathbf{X_m}$. The coefficients of $\mathbf{B_m}$ are the amplitude and phase

of the signal transmitted by each element of the array and are typically given by $B_{mn}=A_{mn}e^{j(\omega d_{mn}/c)}$ where d_{mn} is the distance between the array element n and the focal point $\mathbf{X}_m$, and $\mathbf{A}_{ml}$ is an apodization term.

We define $\mathbf{B}$ as the $M\times N$ matrix whose mth row is the vector $\mathbf{B_m}$.

Building a second transfer matrix using the signals received on each element with focused transmit pulses gives the generalized transfer matrix $\mathbf{K_{foc}}$. $(K_{\mathrm{foc}})_{mm}$ is the signal received on element n for the mth focused transmit, as seen in Fig. 1. $\mathbf{K_{foc}}$ can be expressed in terms of $\mathbf{K}$, the *true* transfer matrix described earlier:

$$\mathbf{K_{foc}} = \mathbf{K}\mathbf{B}^{\mathbf{T}}. \tag{11}$$

If one chooses a family of vectors forming a basis, then $\mathbf{B}$ is known and invertible and it is possible to get back to the transfer matrix $\mathbf{K}$ using $\mathbf{K}=\mathbf{K}_{\mathrm{foc}}(\mathbf{B}^{\mathbf{T}})^{-1}$. However, in the general case, as in our experiments, the $\mathbf{B_m}$ are not necessarily linearly independent, therefore, they do not form a basis. In the following $\mathbf{B}$ is considered to be noninvertible and unknown.

2. Case of isotropic, pointlike scatterers and single scattering

Substituting Eq. (4) into Eq. (11), it is still possible to express $\mathbf{K_{foc}}$ as

$$\mathbf{K_{foc}} = (\mathbf{H_{Rx}})^{\mathbf{T}}\mathbf{D}\mathbf{H_{Rx}}\mathbf{B}^{\mathbf{T}}. \tag{12}$$

It is worth noting that the same array is used in transmission and in reception in this section. $\mathbf{H_{Rx}}$ refers here to the propagation matrix between this one array and the scatterers. The rows of $\mathbf{H_{Rx}}$ are then the scatterers's Green function expressed in the array basis.

Defining $\mathbf{H_{Tx}}$ to be

$$\mathbf{H_{Tx}} = \mathbf{H_{Rx}}\mathbf{B}^{\mathbf{T}}, \tag{13}$$

$\mathbf{K_{foc}}$ can be expressed as

$$\mathbf{K_{foc}} = (\mathbf{H_{Rx}})^{\mathrm{T}}\mathbf{D}\mathbf{H_{Tx}},$$

as in the case of two different arrays introduced in Sec. I A 2. Therefore, from a mathematical standpoint, the process can be described like a classical DORT method between two different arrays. The receive array is a physical array, but the transmit array is here purely virtual. Coordinates of a vector in this array are given by the projection of the vector onto each $\mathbf{B_i}$. In the case of the focused beams, everything happens as if there were virtual transmit elements located at the foci of the beams and having a certain directivity angle.[19,20] The following derivations are not restricted to focused transmits.

This analogy yields the possibility to use the results of Sec. II A 2. A *pseudo* time reversal operator can then be built in the receive basis by choosing $\mathbf{T_{Rx}}=\mathbf{K_{foc}}(\mathbf{K_{foc}})^{\mathbf{H}}$. A condition for this *pseudo* time reversal operator to be diagonal in the scatterers' basis is again given by Eqs. (8) and (9). The physical interpretation of Eq. (8) is that the scatterers must be well-resolved by the array.

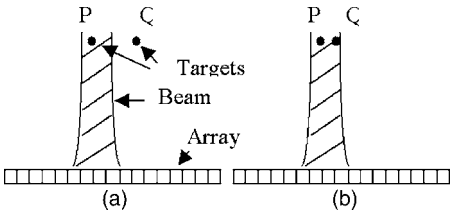


FIG. 4. Coupling condition with focused transmit. (a) The distance between the scatterers is greater than the width of the PSF; each beam insonifies only one scatterer, and they are not coupled. (b) the scatterers have a separation smaller than the point spread function; consequently, a beam may insonify both scatterers. In this case they are coupled.

Interpretation of Eq. (9) is not so straightforward as $\mathbf{H_{Tx}}$ does not correspond to a real array: One has to notice that, in the case of two scatterers $\mathbf{P}$ and $\mathbf{Q}$

$$\mathbf{H_{Tx}} = \begin{pmatrix} \langle \mathbf{B_1}|\mathbf{H_{Rx}}(\mathbf{Q})\rangle & \langle \mathbf{B_2}|\mathbf{H_{Rx}}(\mathbf{Q})\rangle & \cdots & \langle \mathbf{B_M}|\mathbf{H_{Rx}}(\mathbf{Q})\rangle \\ \langle \mathbf{B_1}|\mathbf{H_{Rx}}(\mathbf{P})\rangle & \langle \mathbf{B_2}|\mathbf{H_{Rx}}(\mathbf{P})\rangle & \cdots & \langle \mathbf{B_M}|\mathbf{H_{Rx}}(\mathbf{P})\rangle \end{pmatrix}. \tag{14}$$

Following Eq. (11)

$$\mathbf{H_{Tx}} = \begin{pmatrix} F_1(\mathbf{Q}) & F_2(\mathbf{Q}) & \cdots & F_M(\mathbf{Q}) \\ F_1(\mathbf{P}) & F_2(\mathbf{P}) & \cdots & F_M(\mathbf{P}) \end{pmatrix}, \tag{15}$$

where $F_m(\mathbf{Q})$ is the field obtained at position $\mathbf{Q}$ when the mth focused pulse is transmitted. Now the condition (9) is

$$\sum_{m=1}^{M} F_m(\mathbf{Q})\bar{F}_m(\mathbf{P}) = 0. \tag{16}$$

This condition is valid for any shape of the transmit beams and means that the series $F_m(\mathbf{Q})$ and $F_m(\mathbf{P})$ are uncorrelated. In other words, the scatterers have to be insonified in an uncorrelated way. For focused beams and supposing that the scatterers are at the same depth and separated by a lateral distance Δx, Eq. (16) can be rewritten as

$$C_0 F(x-\Delta x,z)\otimes_x F(-x,z) = C_0 R_{FF}(\Delta x) = 0, \tag{17}$$

with $F(x,z)$ being the beam pattern at the scatterers' depth and R_{FF} its autocorrelation function, as shown in the Appendix. If no apodization is used (rectangular transmit aperture), this simplifies to

$$F(\Delta x, z=f) = 0, \tag{18}$$

where f is the focal depth. The two scatterers have to be resolved by the transmit focused beam, or in other words Δx must be greater than the resolution of the beam. For two scatterers located in the focal plane of the array, this can be understood easily; indeed, each focused pulse insonifies only a limited part of the medium. In the focal plane, this area is the point spread function. The sum of Eq. (16) can only be nonzero if at least one of the products is nonzero; In other words, at least one pulse has to insonify both location $\mathbf{P}$ and $\mathbf{Q}$ as depicted in Fig. 4. If the targets are located so that no pulse insonifies more than one target, then the condition of Eq. (16) is realized. It is important to note that the coupling condition is independent of the focal depth of the array and

the presence of aberration, as demonstrated in the Appendix. FDORT, like DORT, is then robust in the presence of an aberrator. The derivation in the Appendix makes apparent a condition on the beam spacing: It must be smaller than the beam resolution, which is a classic condition in imaging and can be related to the sampling theorem.

Again, if both conditions of Eqs. (8) and (16) are realized, the *pseudo* time reversal operator is diagonal in the scatterers basis, and we deduce that the eigenvectors of the time reversal operator expressed in the receive basis are the Green functions of the point scatterers as seen by the array. Note that the eigenvectors in the transmit basis (also provided by the SVD of $\mathbf{K_{foc}}$) show how much each scatterer is insonified by each focused beam. The eigenvalue associated with the scatterer P is then given by Eqs. (6), (7), and (13)

$$\lambda(\mathbf{P}) = \|\mathbf{H_{Rx}}(\mathbf{P})\mathbf{B}^{\mathbf{T}}\|^2 \|\mathbf{H_{Rx}}(\mathbf{P})\|^2 D(\mathbf{P})^2. \quad (19)$$

C. Main results and comparison between DORT and FDORT

The first step of the FDORT method is to build a generalized transfer matrix. At a given frequency, the coefficients of this matrix are the Fourier coefficients of the receive signals on each of the N array elements for each of the M transmit focused pulses. Each column of the matrix corresponds to one transmit focused pulse, and each row corresponds to the signal received by one element. Then, a singular value decomposition of the transfer matrix is computed as shown in Eq. (3). If the point scatterers are well separated so that there is no coupling between them, the number of nonzero singular values in S gives the number of point scatterers in the medium and their reflectivities. The corresponding singular vectors in U are the Green functions of each scatterer as seen by the array. Backpropagation of these singular vectors in the medium leads to a focusing at the position of the scatterers, even in the presence of aberration.

This implementation and the results of FDORT are very similar to the DORT method. We now want to compare DORT, using the same array of N elements in both transmit and receive as it is usually implemented, and FDORT, using the same array of N elements. Usually, in a scanner, only a fraction of the total aperture is used for each transmission. Here, each focused pulse for FDORT is assumed to be transmitted with half the aperture ($N/2$ elements), which is the case in the simulations. The main difference between DORT and FDORT is the ability to separate two scatterers, so that each eigenvector is the Green function of one scatterer. This is here referred to as the resolution. For DORT, the ability to separate two scatterers depends only on the total aperture. For FDORT, the dependence is on both the *Rx* process [Eq. (8)] depending on the whole aperture, and the *Tx* process [Eq. (18)], determined here by the half aperture. This is in fact the same resolution condition that one has in conventional imaging. The resolution is then slightly lower with FDORT. An important result, *a priori* not obvious, is that if all the elements were used for each transmit with FDORT, FDORT would yield results comparable to DORT (using same array in *Rx* and *Tx*). Results are summarized in Table I.

TABLE I. Comparison of DORT and FDORT.

	DORT	FDORT
Data	Full data set	Focused transmit pulses
Separation of scatterers	Given by the total aperture: N elements	Limited by the transmit aperture: $N/2$ elements here
Singular values	1 nonzero eigen value per scatterer, depend on the reflectivity	1 nonzero eigenvalue per scatterer, depend on the reflectivity
Singular vectors in U(eigenvector of the time reversal operator in thereceive basis)	Green function of each scatterer as seen by the array	Green function of each scatterer as seen by the array
Singular vectors in V(eigenvector of the time reversal operator in the transmit basis)	Green function of each scatterer as seen by the array	No information on the Green function, but shows the signal transmitted by each focused pulse on the scatterer

However, one of the main advantage of FDORT is the possibility to use focused transmit and also to implement the algorithm presented in Secs. III and IV.

D. Simulations

A Philips one-dimensional (1D) linear array at 7.3 MHz center frequency is simulated using Field2 [J. A. Jensen, DTU] along with two point scatterers separated by 2 mm, both in the focal plane of the array. A phase-screen aberrator with parameters rms (root mean square) =30 nm and FWHM (full width half max) =4 mm is placed at the array surface. $M=100$ focused transmissions separated by 0.1 mm are used. For each transmission, the backscatter signal is recorded on each of the elements. A Fast Fourier Transform of the $N^{\times}M$ signals is taken to build the generalized transfer matrix at several frequencies. A singular value decomposition algorithm is used to compute eigenvalues and eigenvectors of the time reversal operator. Results are shown in Fig. 5. Two significant eigenvalues are observed. The corresponding eigenvectors are found to focus very accurately on each of the scatterers through the aberrator.

To compare the coupling condition of both algorithms, another simulation is ran where the distance between the 2 scatterers is reduced progressively from 1 mm to 0. The eigenvalues, which are function of the coupling are computed for both FDORT and DORT and displayed in Fig. 6. The same experiment is repeated with an aberrator and for scatterers out of the focal plan to show that this does not influence FDORT. As it is known[9] that coupling is responsible for an increase of the first eigenvalue and a decrease of the second one, the evolution of eigenvalues show the appearance

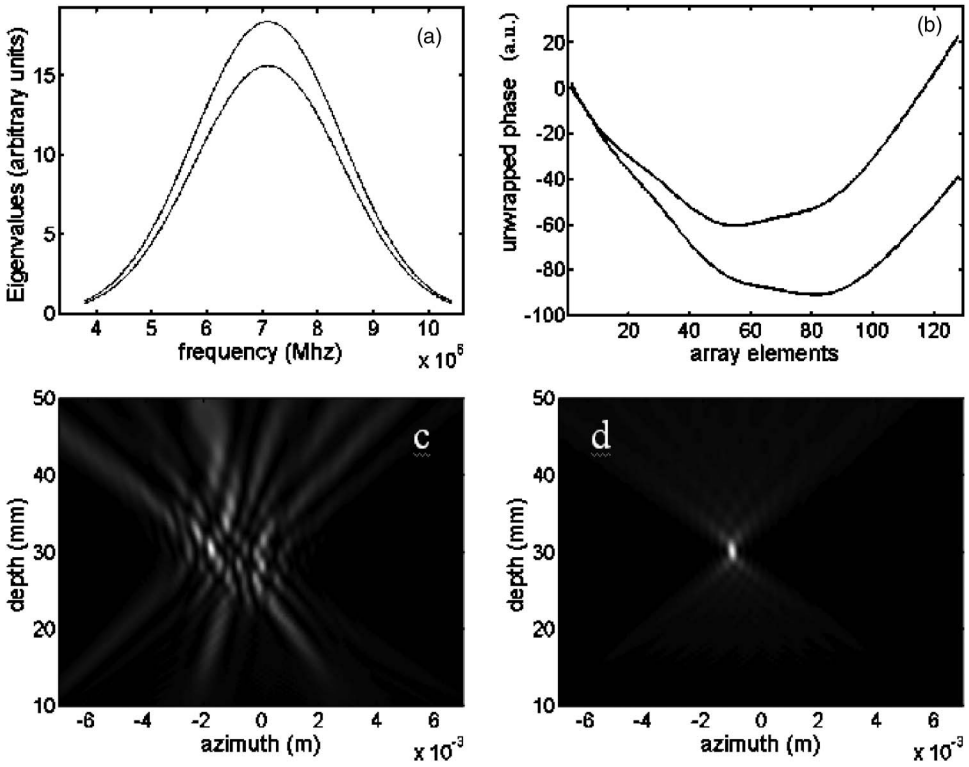


FIG. 5. Results of FDORT performed on simulated data with 2 scatterers and an aberrator. (a) Eigenvalues as a function of frequency: Two eigenvalues have a significant magnitude, corresponding to the two scatterers. (b) Phase of the corresponding eigenvectors at the central frequency. It is proportional to the focal delay law to focus on each of the targets. In particular, it contains the delay law introduced by the aberration. (c) Intensity of the field transmitted in the medium when a simple geometric delay law is used to focus on one of the targets. The aberration results in poor focusing. (d) Intensity of the transmitted field when one of the eigenvectors is used to focus on the target.

and increase of coupling when the distance decreases.

III. ADAPTATION OF FDORT FOR A NOISY ENVIRONMENT

The DORT and FDORT methods are able to detect and focus on well separated point scatterers, like wires (which can be considered point-like in the 2D geometry of the experiments), in water, even in the presence of a strong aberrator.[9] In medical applications, however, the scatterers are embedded in tissue that generates a speckle signal. The FDORT process was performed on a tissue-mimicking phantom with wire-targets, represented in Fig. 7. Eigenvalues and numerical backpropagation of the second eigenvectors are shown in Fig. 8 (bottom). Due to speckle, the eigenvectors become too noisy and FDORT, like DORT, fails: Focusing on the wires' locations is very poor. The next step, then, is to

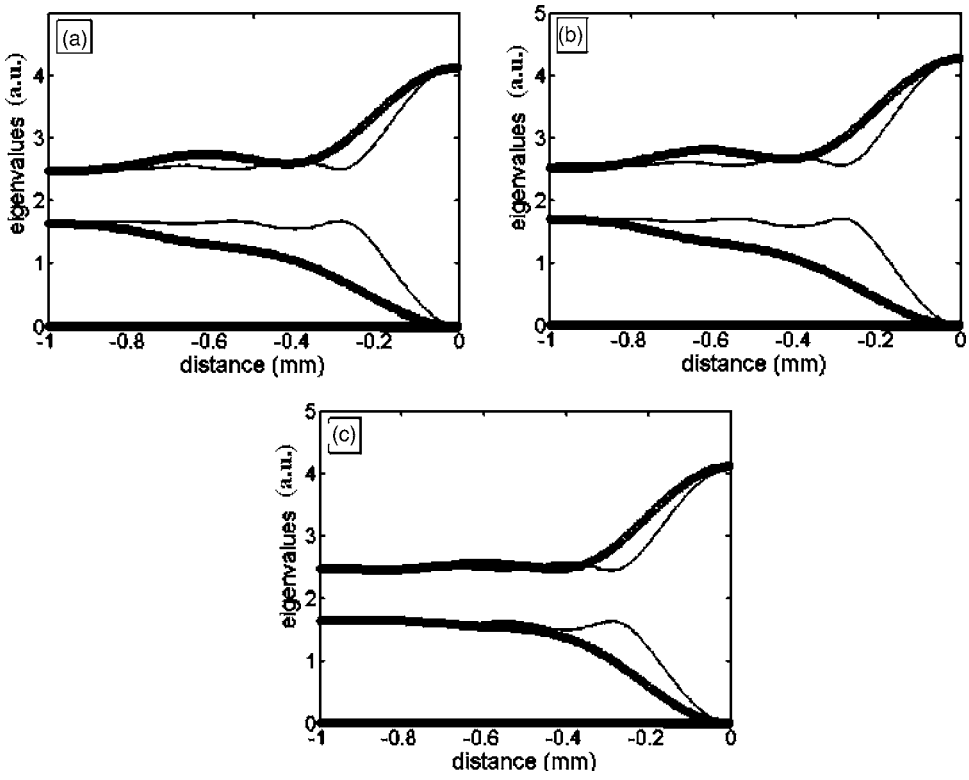


FIG. 6. Eigenvalues of DORT (thin line) and FDORT (thick line) vs distance for two scatterers located at the same depth (a) in the focal plane of the array ($f=20$ mm), (b) in the focal plane but in presence of aberration, and (c) out of the focal plane, at depth 30 mm. When the scatterers are too close, coupling occur and results in a separation of the eigenvalues. The importance of coupling depends on how well the scatterers are resolved by both transmit and receive array. It occurs over a larger range with FDORT because only half the aperture is used in transmission. FDORT's ability to separate two targets response is not degraded by aberration or for scatterers out of the focal plane. DORT and FDORT eigenvalues have been normalized to appear on the same scale.

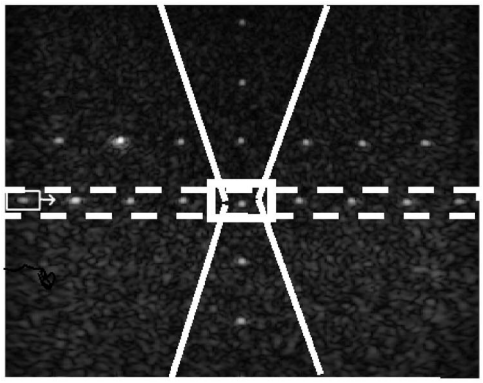


FIG. 7. Tissue-mimicking phantom used for the experiments. The zone of coupling for one scatterer is indicated by the white conical shape. Scatterers inside this area are coupled with the scatterer. The box drawn with dashed lines indicates the area used in Sec. III, obtained by time gating the signals. This reduces the influence of the coupled scatterers. The box in solid lines indicates the area used for Local FDORT in Sec. IV.

adapt the method to work in such an environment. For simplificity, the case of point scatterers in a speckle-generating medium will be considered first. The aim is to be able to detect the point scatterers and obtain a good focusing on each of them, despite the strong speckle signal. A good focus is indeed required for target detection, or if one's aim is to image a medium, for aberration correction.

A. Influence of noise on eigenvectors and eigenvalues

In this part is considered only the problem of acoustic noise, in particular speckle noise. Indeed, the influence of the electronic noise is already known. It is spatially uncorrelated and thus it can be shown[13] that the time reversal operator at a given frequency can be approximated by

where $(\mathbf{H_{Rx}})^{\mathbf{T}}\mathbf{D}\mathbf{H_{Tx}}(\mathbf{H_{Tx}})^{\mathbf{H}}\overline{\mathbf{DH_{Rx}}}$ is the time reversal operator in the absence of noise [Eq. (5)], M is the number of transmit pulses, σ^2 is the noise power, and $\mathbf{I}$ is the identity matrix. The eigenvectors of the time reversal operator remain unchanged and the eigenvalues are uniformly increased by $M\sigma^2$. Thus, a small amount of electronic noise has then no influence on the DORT or FDORT process.

The influence of acoustic noise is more complicated. Signals from scatterers other than the ones we want to detect (here the wires in the phantom) are considered as noise. In that case these are sub-resolution scatterers generating speckle signal. Two cases can be distinguished:

a. Ideally separated scatterers. These are not coupled with the targets; they are located outside the coupling area for each target, as depicted on Fig. 3. They give rise to new nonzero eigenvalues, but do not affect the eigenvectors corresponding to the targets that still enable perfect focusing. The focusing properties are preserved, but as there are more nonzero eigenvalues, it is more difficult to determine which eigenvectors correspond to the targets. The difference with sensor noise is that the eigenvalues are not uniformly increased: eigenvalues corresponding to noise can now have higher values than the ones from targets.

b. Nonideally separated scatterers. They are coupled with the targets. There are more nonzero eigenvalues, as in the previous case, but here the eigenvectors are affected: They are no longer equal to the Green functions of individual targets, but are linear combinations of the targets' Green functions and Green functions from noise scatterers coupled with the targets. The resulting eigenvectors can be very complex in the presence of speckle signals, as hundreds of subresolution scatterers are coupled with the target. Figure 3(b) represents, in a clinical image, all the scatterers coupled with

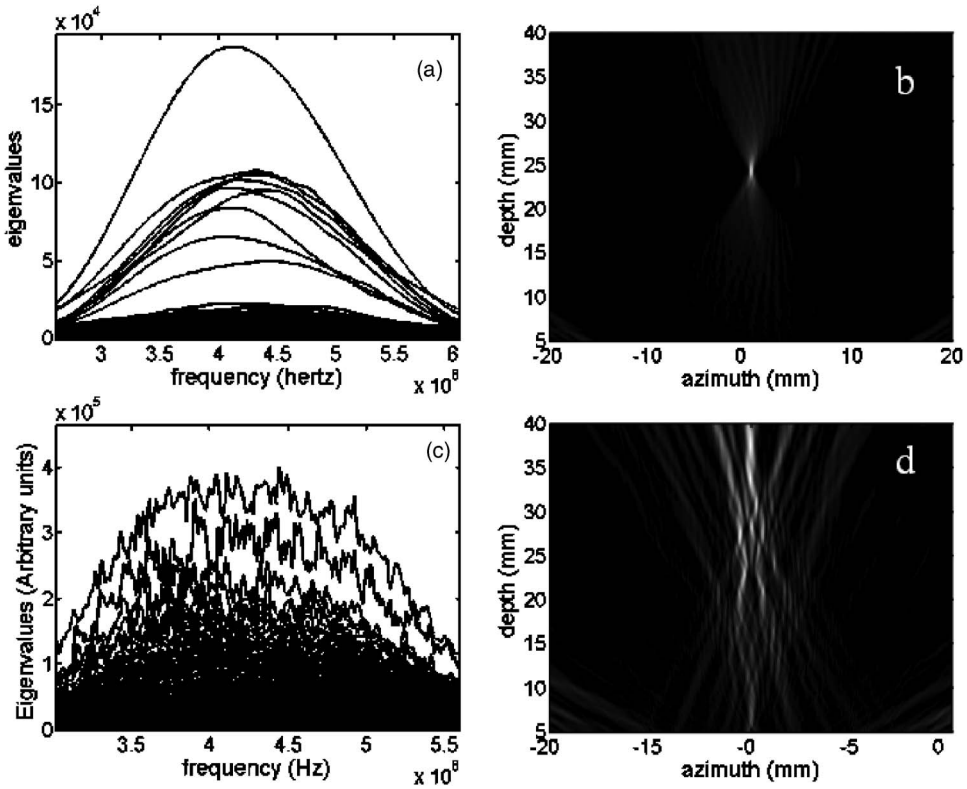


FIG. 8. The FDORT method has been performed on a tissue-mimicking phantom with wires. Top: Eigenvalues vs frequency (a) and intensity (b) of the field resulting from the backpropagation of the second eigenvector, when FDORT is restricted to a slice around depth z=25 mm where 9 wires are located, using time gating. Nine significant eigenvalues are observed and the eigenvectors focus accurately on the wires. Bottom: The same experiment but without time gating. The focusing property of the eigenvectors is dramatically reduced, because of the coupling with the speckle.

the target, located at a depth of 25 mm. The focusing properties are in this case dramatically degraded, as seen in Fig. 8.

B. A solution: FDORT with time gating

1. Principle

We are here interested to reduce the effect of the last kind of noise, which has the worst influence. The noise comes from the set of scatterers coupled with the target, contained in a conical shape centered on the target. Laterally, in the target plane, the zone of coupling is narrow. It was shown in Sec. II. this is determined by the classic point spread lateral extension. This limitation is due to the finite size of the array and affect usual beamforming imaging. Axially the coupled zone is much more extended. This limit is not present in classical beamforming. It is due to the fact that the temporal resolution is not exploited by DORT or FDORT because of their only monochromatic nature. An important amount of information is then lost. To solve this difficulty, impulsive, and monochromatic approaches have to mixed. At any given time τ, an incident ultrasound beam illuminates only a volumetric distribution of scatterers, called the isochronous volume.[21,22] Now, selecting within the echographic signal an analysis time window $[\tau, \tau+\Delta\tau]$ is equivalent to selecting only echoes from scatterers located in a well-defined volume whose lateral extension is equal to the lateral extension of the beam, and whose axial extension Δz is related to $\Delta\tau$. If we repeat this process for every beam, the union of all the volumes gives a slice of the medium of width Δz, represented in Fig. 7. From the point of view of FDORT, using time windows is similar to having an empty medium outside this slice. Therefore, there is no coupling with scatterers outside of the slice. Thanks to time gating, the zone of coupling narrows axially and tends to the limit set by the temporal resolution; targets from different depths can then be fully decoupled if advantage is taken from the impulsive approach.

Such a process is possible only with focused pulse transmission. If a single element transmission is used, as in the classical DORT method, the whole medium is insonified and it is no longer feasible to select signals from a given depth.

2. Experiments

Experiments are carried out on the medical phantom represented in Fig. 7, using a Philips HDI-5000 and a 1D linear array at 4.3 MHz center frequency.

For each transmission the signals on all N received elements are recorded. The signals are gated in time, keeping only the signal from $z-0.5\Delta z$ to $z+0.5\Delta z$. The window width is chosen to be slightly longer than the pulse width. For the mth transmission pulse, gating in depth is achieved using the geometrical focal law focusing along the beam m at depth z.

Figure 8 (top) shows the eigenvalue spectrum and the numerical backpropagation of the first eigenvectors obtained using the FDORT method in the medical phantom at the depth of 9 wires. It demonstrates a great improvement in the focusing ability of the first eigenvectors, compared to the results in Fig. 8 (bottom). To obtain the focal law of each wire, one needs to repeat the process for several depth ranges.

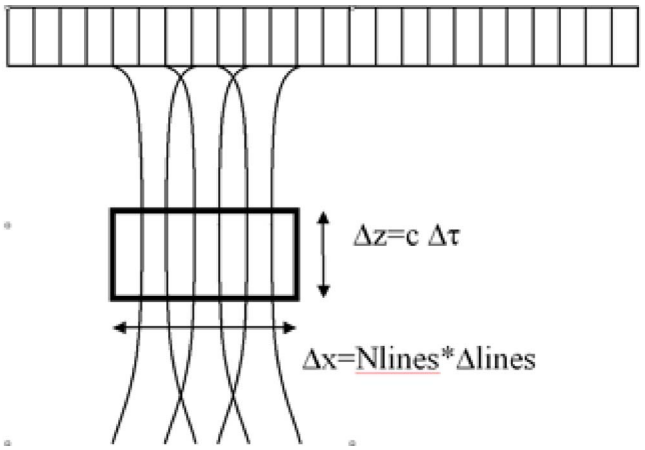


FIG. 9. Shows how FDORT can be restricted to a small area (the black window depicted): Windowing in depth (Δz) is achieved by time gating ($\Delta\tau$) of the signals and windowing in azimuth (Δx) is achieved by the number (Nlines) of consecutive lines used: On the picture, three lines are used. Δlines is the distance between two consecutive lines (the beam spacing).

In conclusion, the FDORT method can still be performed to detect point targets in the presence of speckle noise, but the medium needs to be sampled in thin slices to decouple the targets from the speckle noise. The number of significant eigenvalues is the number of targets in the slice, and the eigenvectors are the corresponding Green functions.

IV. THE LOCAL FDORT METHOD

A. Principles

In Sec. III B we proposed gating the signal in depth to reduce coupling. In addition, the volume can be shrunk along the lateral dimension. To perform the method in a zone contained between azimuth $x-\Delta x$ and $x+\Delta x$, only the lines whose azimuth lies between $x-\Delta x$ and $x+\Delta x$ are processed. Thus, using FDORT, we can specify the size of the volume where the method is performed: In azimuth by setting which lines are processed, and in depth by setting the width of the time window. This is shown in Fig. 9.

The lowest lateral extension of the area where one can perform FDORT without losing information about a point scatterer is here about 10 consecutive lines, corresponding to a width of 1.5 mm. This corresponds to the beam width. FDORT performed in such a small window, represented with solid line in Fig. 7, is termed Local FDORT. It processes the signal reflected only from a specific location and gives the first few eigenvalues and eigenvectors. This local algorithm offers the best option to detect a low reflectivity scatterer within speckle; it minimizes the influence of the noise resulting from both coupled speckle (by windowing in depth) and noncoupled speckle (by windowing in azimuth).

A Local DORT method has been performed from a full data set by Kerbrat *et al.*,[23] using a prefocused array. However, they need to repeat the acquisitions for each location where they want to perform DORT and all the transmissions

have to be processed. For the method presented in this paper, only one set of acquisitions is needed, and only 10 transmissions need to be processed for each location. This results in an important savings in processing cost. As it uses focused, and thus localized, transmissions, FDORT is better adapted for local processing than the conventional DORT method.

B. *In vivo* experiments: Application to microcalcification detection

One of the challenges of breast ultrasonic imaging is improved detection of small microcalcifications, which can be associated with cancer. Small microcalcifications are often hard to distinguish because of their small size and because they are embedded in speckle.

1. Influence of the scatterer nature in the DORT method

When the scatterer size is less than a wavelength, the scatterer is associated with only one nonzero eigenvalue. However, for deterministic scatterers whose size is greater than a wavelength, two or more eigenvalues are observed.[10] Also, for speckle signals, several eigenvalues of similar magnitude are observed, as seen in Fig. 10(a).

A way of detecting small deterministic scatterers like the microcalcifications is to use the Local FDORT method and to consider the ratio of the first two eigenvalues. If a point scatterer is present at the location, the first eigenvalue is higher than the second one, and the ratio is dramatic. If there is only speckle or an extended reflector, like a cyst edge, the eigenvalues are of similar magnitude and the ratio is close to 1. Considering only the first eigenvalue is not enough, as it is proportional to the echogenicity (reflectivity) which is often greater for extended objects or even speckle; the ratio takes into account the size difference.

2. The moving window FDORT

Local FDORT gives access to the local properties of the medium. Thus it gives the eigenvalues and the ratio of the eigenvalues at a specific location, the location of the window. To scan the medium, one needs to move the window at a series of locations. This can be done in depth by translating the time gate or in azimuth by changing the lines processed. For example, the first window may use lines 1–10, the second window lines 2 to 11, and so on. For each location, the sizes of the windows are the same. Eigenvalues are computed for each position. This is the moving window FDORT.

3. In vivo experiments and results

Experiments are carried out on clinical data. Acquisitions are performed on a healthy female volunteer, using a Philips HDI-5000 scanner with a 1D linear array probe, at 7.3 MHz center frequency. Local FDORT is performed in a window whose dimension is $\Delta z = 0.7$ mm and $\Delta x = 1$ mm (10 lines). The window is moved along the white line depicted on Fig. 10(b) at a constant depth, where a microcalcification has been identified, using a moving window FDORT. Figure 10(c) shows the variation of the first eigenvalues versus azimuth, at the center frequency. Looking only at the first eigenvalue, which is proportional to the reflectivity, we cannot distinguish the microcalcification from the other scatterers. However, when considering the ratio λ_1/λ_2, in Fig. 10(d), the microcalcification appears clearly. Averaging over several frequencies within the bandwidth improves the results. The process can be repeated for several depths, and a 2D color map can be plotted, but at significant computational cost.

Finally, the eigenvector corresponding to the identified microcalcification has been used in focusing. The resulting field in the medium has been computed and the results are displayed in Fig. 11. This demonstrates the good focusing property of FDORT's eigenvectors in a clinical application.

V. CONCLUSION

This paper demonstrates that the DORT method can be successfully generalized to acquisitions with focused transmission beams. It is then termed FDORT. This mode of transmission is the one routinely used by medical ultrasound scanners. If point scatterers are well resolved, each scatterer is associated with an eigenvector given by FDORT, and focusing on each scatterer can be performed by transmitting the corresponding eigenvector. DORT and FDORT would be equivalent if each focused pulse was transmitted using the whole aperture of the array (and if enough beams are used to span the region of interest), but this is usually not the case. The resolution is then lower with FDORT and is limited by the size of the transmit aperture.

In biological tissue, as in other complex media, speckle noise dramatically reduce the performance of DORT or FDORT. We show that this effect is due to speckle located in the zone of coupling (*shadow*) of the scatterers. The solution is to take advantage of the temporal resolution in an impulsive approach and to sample the medium in slices. Gating the signal in depth allows getting rid of most of the noise, and detection of wires in a tissue-mimicking phantom was performed. Eigenvectors in that case had good focusing properties. Depth gating takes advantage of the focused transmits of FDORT, and is not directly compatible with DORT. Eigenvectors associated with microcalcifications in clinical data were exhibited, demonstrating the feasibility of the method *in vivo*. Other applications may exist for nondestructive evaluation or detection of objects on ocean floor.

A method to help in microcalcification detection is also proposed. The difference of the FDORT *signature*, using the ratio of the two first eigenvalues, between point scatterers, whose size is less than the wavelength, and other scatterers, is exploited. Using a moving-window algorithm, a scan highlighting the position of the point scatterers can then be displayed. Again, FDORT is better suited than DORT for this application.

The criterion used to detect the microcalcifications is their size, which is not the most relevant criterion. It has been shown that DORT can also give information on the mechanical properties of the scatterer,[18] and further work should investigates a criterion based on the fact that the material of the microcalcifications is different from the material

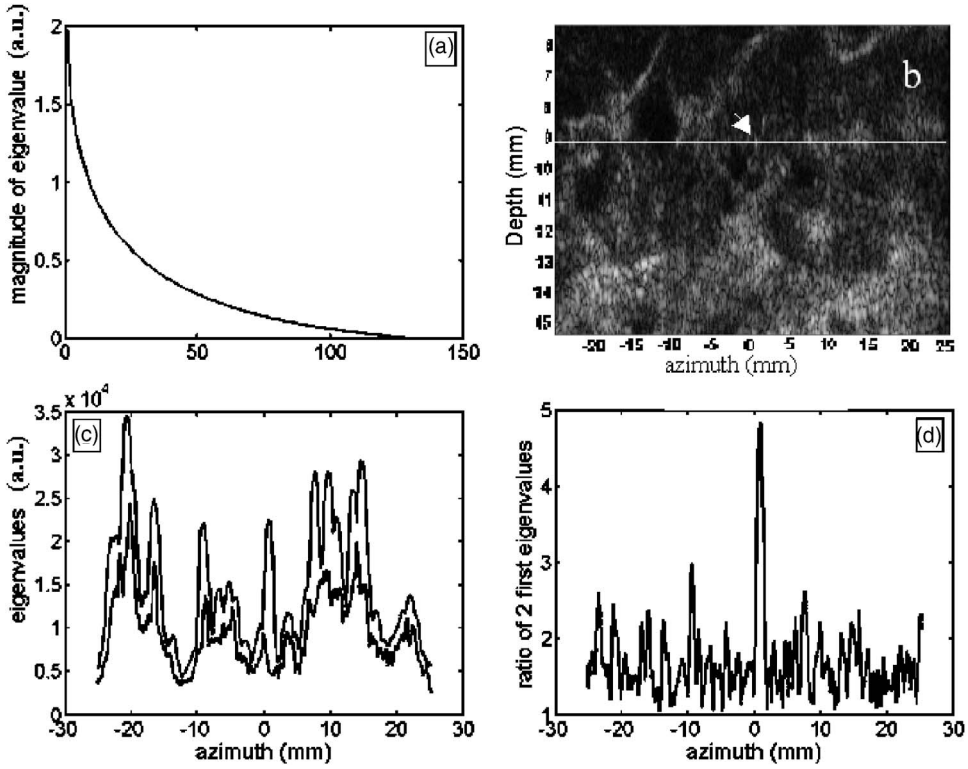


FIG. 10. (a) Eigenvalues in pure speckle in decreasing order: The spectrum is continuous, thus the first two eigenvalues have similar magnitude. FDORT was performed locally in a 0.7 mm deep and 1 mm wide area at several azimuth along the white line depicted on the breast ultrasound image. (b) A scatterer identified as a microcalcification is shown by the white arrow. (c) The first two eigenvalues are plotted vs the azimuth. The first eigenvalue is proportional to the echogenicity, but considering the second eigenvalue adds additional information. (d) Ratio of the first two eigenvalues vs the azimuth: The position of the microcalcification is indicated by a high ratio.

of tissue. It has also been shown that FDORT can focus well even in pure speckle regions.[24] This should be investigated and theoretically justified in further studies.

APPENDIX: COUPLING TERM BETWEEN TWO SCATTERERS WITH FDORT

For simplicity, the scatterers are first assumed to be in the focal plane. It is also supposed that all the beam patterns are identical to within a translation, which is usually the case without aberration and not too close to the array edge. Under the Fresnel approximation, the beam patterns in the focal plane are the Fourier transform of the aperture functions. The pattern centered at $x=0$ is given by $F(x)=e^{j\pi x^2/\lambda z}FT[A(X)]\times(x/\lambda z)$, where $A(X)$ is the aperture function and FT is the Fourier transform. We will neglect the complex exponential factor in the following. The beam pattern for a beam centered at azimuth x_i is then $F_i=F(x-x_i)$ and the field sensed by target P is $F_i(P)=F(x_P-x_i)$.

Now the coupling term of Eq. (16) becomes

$$C(Q,P)=\sum_{i=1}^{N}F_i(Q)F_i(P)^*=\sum_{i=1}^{N}F(x_Q-x_i)F(x_P-x_i)^*.$$

Using distributions, this can be rewritten as

$$\begin{aligned}C(Q,P)&=\int F(x_Q-x)F(x_P-x)^*\sum_{i=1}^{N}\delta(x-x_i)dx\\&=\int F(x)\sum_{i=1}^{N}\delta(-x-x_i-x_Q)\\&\quad\times F(x-(x_Q-x_P))^*dx,\end{aligned}\tag{A1}$$

where $\Sigma_{i=1}^{N}\delta(-x-x_i-x_Q)$ is a dirac comb multiplied by a gate of width $x_N-x_1=N$ times the beam spacing, and centered on x_Q. The influence of the gate is negligible if N is large enough so that the width of the beam pattern is neglectible compared to N times the beam spacing. The gate will be omitted in the following. The formula (A1) can be interpreted as the cross-correlation between the 2 functions

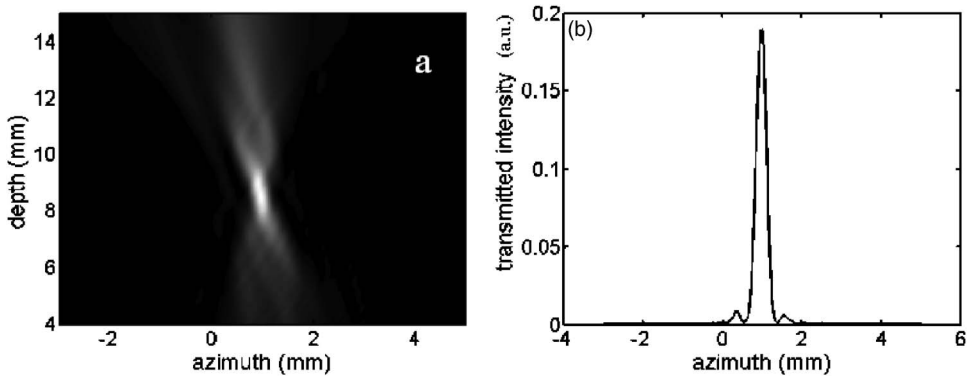


FIG. 11. Focusing achieved in breast clinical data using a microcalcification as a point scatterer. The first eigenvector has been numerically backpropagated. Left: Intensity of the resulting field. Right: Intensity vs azimuth at the depth of the microcalcification.

$F(x)\Sigma_{i=-\infty}^{+\infty}\delta(-x-x_i-x_Q)$ and $F(x)$ at lag x_Q-x_P. Using Fourier transforms it can be expressed as

$$C(Q,P) = FT\left[FT^{-1}\left(F(x)\sum_{i=-\infty}^{+\infty}\delta(-x-x_i-x_Q)\right) \times FT^{-1}(F(x))^*\right]\left(\frac{x_Q-x_p}{\lambda z}\right). \quad \text{(A2)}$$

Now, $FT^{-1}(F(x))=A(X)$ and $FT^{-1}(F(x)\Sigma_{i=-\infty}^{+\infty}\delta(-x-x_i-x_Q))$ is equal to $A(X)$ convoluted by a Dirac comb. Two Dirac peaks in the comb are separated by $d=\lambda z/\Delta x_i$, where f is the focal length and Δx_i is the beam spacing. If d is greater than the width of the aperture function D, or in other words, if the beam spacing is smaller than the resolution $\lambda z/D$, then Eq. (A2) simplifies to

$$C(Q,P) = C_0 FT[|A(X)|^2]\left(\frac{x_Q-x_p}{\lambda z}\right) = C_0 F(x-\Delta x)\otimes F(-x) = C_0 R_{FF}(\Delta x) = 0. \quad \text{(A3)}$$

where C_0 is a constant.

If no apodization is used, $A(X)$ and $|A(X)|^2$ are both the same rectangular windows. Therefore $C(Q,P)=\mathbf{F}(x_Q-x_p)$. The coupling is then weak if the targets' separation is greater than the beam width. In DORT the coupling is proportional to the Fourier transform of the whole array; in FDORT it depends on the aperture function used for each transmission.

If there are aberrations at the transducer surface, the aperture function becomes $A(X)e^{i\varphi(X)}$. As only the modulus of the aperture function is considered in the coupling term, it is not affected by the aberration as long as only a phase shift is introduced.

If the targets are out of the focal plane, the field in the plane of the targets can be expressed by the Fourier transform of $A(X)e^{i\phi(X)}$ with $\phi(X)=(\pi/\lambda z)X^2$. The phase term also disappears when the modulus is taken. The coupling should then not depend on the focal length.

[1] M. O'Donnell and S. W. Flax, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: Measurements," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 768–774 (1998).

[2] R. C. Gauss, M. S. Soo, and G. E. Trahey, "Wavefront distortion measurements in the human breast," IEEE Ultrasonic Symposium Proceedings, 1547–1551 (1997).

[3] R. C. Gauss, G. E. Trahey, and M. S. Soo, "Wavefront estimation in the human breast" Proc. SPIE **4325**, 172–181 (2001).

[4] D. L. Liu, J. A. Baker, and P. Von Behren, "A clinical study of adaptive beamforming using time-delay adjustments on a 1D array," IEEE Ultrasonic Symposium Proceedings, **1**, 339–342 (2003).

[5] J. J. Dahl and G. E. Trahey, "Off-axis scatterer filters for improved aberration measurements," IEEE Ultrasonic Symposium Proceedings, 343–347 (2003).

[6] M. Fink, "Time reversal of ultrasonic fields-Part I: Basic principles," IEEE Trans. Sonics Ultrason. **39**(5), 555–566 (1992).

[7] C. Prada, F. Wu, and M. Fink, "The iterative time reversal mirror: A solution to self-focusing in the pulse echo mode," J. Acoust. Soc. Am. **90**(2), 1119–1129 (1991).

[8] C. Prada and M. Fink, "Eigenmodes of the time reversal operator: A solution to selective focusing in multiple-target media," Wave Motion **20**, 151–163 (1994).

[9] C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Application to detection and selective focusing on two scatterers" J. Acoust. Soc. Am. **99**, 2067–2076 (1996).

[10] S. Komilikis, C. Prada, and M. Fink, "Characterization of extended objects with the DORT method," IEEE Ultrasonic Symposium Proceedings, 1401–1404 (1996).

[11] A. J. Devaney, "Computational time reversal and object localization," J. Acoust. Soc. Am. **110**, 2617 (2001); A. J. Devaney, "Super-resolution processing using time reversal and music," J. Acoust. Soc. Am. (unpublished).

[12] A. B. Baggeroer, W. A. Kupperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**(4), (1993).

[13] 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. **113**(6), 1–9 (2003).

[14] J. G. Berryman *et al.*, "Statistically stable ultrasonic imaging in random media," J. Acoust. Soc. Am. **112**, 1509–1522 (2002).

[15] T. D. Mast, A. L. Nachman, and R. C. Waag, "Focusing and imaging using eigenfunctions of the scattering operator," J. Acoust. Soc. Am. **102**, 715–725 (1997).

[16] C. F. Gaumond, D. M. Fromm, J. Lingevitch, R. Menis, G. Edelmann, D. Calvo, and E. Kim "Application of DORT to active sonar," MTS/IEEE TECHNO-OCEAN **4**, 2230–2235 (2004).

[17] C. Prada, M. Tanter, and M. Fink, "Flaw detection in solid with the DORT method" IEEE Ultrasonic Symposium Proceedings, 681–686 (1997).

[18] D. H. Chambers and A. K. Gautesen, "Time reversal for a spherical scatterer," J. Acoust. Soc. Am. **109**, 2616–2624 (2001).

[19] C. H. Frazier and W. D. O'Brien, "Synthetic aperture technique with a virtual source element," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(1), 196–207 (1998).

[20] K. L. Gammelmark and J. A. Jensen, "Multielement synthetic transmit aperture imaging using temporal encoding," IEEE Trans. Med. Imaging **22**(4), 552–563 (2003).

[21] M. Fink and J. Cardoso, "Diffraction effects in pulse-echo measurements," IEEE Trans. Sonics Ultrason. **31**, 313–329 (1984).

[22] R. Mallart and M. Fink, "Adaptative focusing in scattering media through sound speed inhomogeneities: The Van Cittert Zernike approach and focusing criterion," J. Acoust. Soc. Am. **96**(6), 3721-3732 (1994).

[23] E. Kerbrat, 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).

[24] Michael R. Burcher, Anna. T. Fernandez, and Claude Cohen-bacrie, "A novel phase aberration technique derived from the D.O.R.T. method: Comparison with correlation based method on simulated and *in vivo* data," IEEE Ultrasonic Symposium Proceedings, August 2004.

Adjoint-based acoustic inversion for the physical characterization of a shallow water environment[a)]

Jean-Pierre Hermand[b)] and Matthias Meyer[c)]
Department of Physics and Mathematics, Université libre de Bruxelles, Environmental Hydroacoustics Laboratory, Avenue Franklin D. Roosevelt, 50 - CP 194/05, B-1050 Brussels, Belgium and The Royal Netherlands Naval College, P. O. Box 10000, 1780 CA Den Helder, The Netherlands

Mark Asch[d)]
Université de Picardie Jules Verne, LAMFA-CNRS UMR 6140, 33 rue Saint Leu, 80039 Amiens, France

Mohamed Berrada[e)]
Laboratoire d'Océanographie et du Climat-Expérimentation et Approches Numériques, Université Pierre et Marie Curie, 4 Place Jussieu, 75252 Paris, France

(Received 7 December 2005; revised 24 March 2006; accepted 26 March 2006)

Recently the concept of adjoint modeling has been introduced in shallow water acoustics for solving inverse problems. Analytical adjoints have been derived for normal modes and for both the standard parabolic equation and Claerbout's wide-angle approximation (WAPE). This paper proposes the application of a semiautomatic adjoint approach that has been successfully applied in the past for multidimensional variational data assimilation in meteorological and climate modeling. Starting from a modular graph representation of the underlying forward model, a programming tool facilitates the generation and coding of both the tangent linear and the adjoint models. The potential of this numerical adjoint approach for the physical characterization of a shallow water environment is illustrated with two applications for geoacoustic inversion and ocean acoustic tomography using Claerbout's WAPE in combination with nonlocal boundary conditions. Furthermore, the adjoint optimization is extended to multiple frequencies and it is shown how a broadband approach can enhance the performance of the inversion process. For a sparse array geometry in particular, the generalization of the adjoint-based approach to a joint optimization across multiple frequencies is necessary to compensate for the lack of vertical sampling of the propagation modes. Results with test data synthesized from geoacoustic inversion experiments in the Mediterranean show that with the numerical adjoint approach the acoustic field, the sound speed profile in the water column and the bottom properties can be efficiently retrieved. [DOI: 10.1121/1.2197790]

PACS number(s): 43.60.Pt, 43.60.Rw, 43.30.Pc, 43.60.Uv [EJS] Pages: 3860–3871

I. INTRODUCTION

Over the past two years the method of adjoint modeling has gained an increasing interest in the ocean acoustics community.[1–7] Adjoint models have been derived for normal modes,[1–4] for the standard parabolic equation (SPE),[5,6] as well as for Claerbout's wide-angle approximation (WAPE).[7] Basically the idea in the different adjoint applications is to produce the necessary corrections to the respective model inputs that cause mismatches between the observations and model predictions either by using the adjoint of a normal mode model or the adjoint of a PE forward model. The adjoint approach enables an exact gradient calculation of the cost function by means of which the number of modeling runs required for the inversion of an acoustic field can be significantly decreased. The derivation of the adjoint in these cases is generally accomplished by an analytical transformation of the governing differential or integral equations in the continuous domain[1–3,6,7] or by manipulation of the corresponding finite difference matrices,[5] respectively.

In this context the present paper proposes a numerical approach for semiautomatic adjoint generation on the basis of a modular graph methodology. Starting from the modular graph, a process and data flow diagram which describes the underlying acoustic model, an algorithmic tool (YAO)[8] facilitates the generation and coding of both the tangent linear and the adjoint models. This allows a higher degree of complexity in the physical model compared to the analytical case where the adjoint has to be derived by hand. The exact gradient information obtained by means of numerical differentiation with YAO is then used in combination with an efficient quasi-Newton gradient technique for large-scale bound-constrained or unconstrained optimization.[9–11]

[a)]Portions of this work were presented at the International Conference on Underwater Acoustic Measurements: Technologies & Results, 28 June – 1 July, 2005, Crete, Greece and at the 7th International Conference on Theoretical and Computational Acoustics, 19–23 September, 2005, Hangzhou, Zhejiang, China.

[b)]Electronic mail: jhermand@ulb.ac.be
[c)]Electronic mail: mmeyer@ulb. ac. be
[d)]Electronic mail: mark.asch@u-picardie.fr
[e)]Electronic mail: mohamed.berrada@lodyc.jussieu.fr

 0001-4966/2006/119(6)/3860/12/$22.50

In the past the numerical adjoint approach has been successfully applied, e.g., for multidimensional variational data assimilation in meteorological and climate modeling, for variational data assimilation with several models in oceanography (three-dimensional modeling of phytoplankton growth),[12] and for land hydrology with the ISBA code of Météo-France.[13,14] Here, the potential of this numerical adjoint approach is illustrated for the physical characterization of a shallow water environment with two applications in ocean acoustic tomography and geoacoustic inversion using Claerbout's WAPE and the Yevick and Thomson nonlocal boundary conditions (NLBC).[15] The analytical optimal control approach proposed in an earlier paper by the authors[7] is extended numerically to multiple frequencies and it is shown how a broadband approach can enhance the performance of the inversion process, especially for the case of a sparse receiver array spanning part of the water column. The very concept of multiple frequency adjoint-based inversion of a locally reacting impedance boundary condition for the standard PE was first presented analytically in Ref. 16. Here, the multiple-frequency adjoint is generated numerically for Thomson's finite difference WAPE solver with NLBCs thus allowing a direct inversion of the embedded geoacoustic parameters of the bottom and the sound speed profile in the water column.

Following a general description of the automatic adjoint approach and the modular graph methodology in Sec. II, Sec. III introduces the PE propagation model and explains the corresponding modular graph decomposition. Details of the optimization technique that is applied in the remainder of the paper are given in Sec. IV. In particular, Sec. IV A deals with the definition of the cost function while Sec. IV B focuses on the minimization routines for bound-constrained and unbounded optimization. Section V starts with a short description of the experimental setup and environmental data collected during the YS94 trials in the Mediterranean sea, south of Elba island. The data are used as a shallow water test case to validate the numerical adjoint approach. The potential of this approach for the physical characterization of the south Elba site and other shallow water environments is illustrated for the two applications of ocean acoustic tomography (OAT) and geoacoustic inversion (GI) in Secs. V A and V B. An additional example combines the two applications and demonstrates the feasibility of geoacoustic inversion in the presence of an uncertain sound speed profile. Section VI concludes by discussing the advantages and limitations of the automatic WAPE-adjoint inversion and the possible extensions of this work.

II. AUTOMATIC ADJOINT MODELING

A. Adjoint formalism

Given an oceanic environment, a model G describing the acoustic propagation for a set of control variables $\chi \mapsto G(\chi)$ and a cost function J which quantifies the fit between the model solution $\nu = G(\chi)$ and a corresponding set of observations, the gradient of the cost function with respect to the control variables χ can be evaluated as follows:[7,17,18]

$$\nabla_\chi J = G'^* \nabla_\nu J. \tag{1}$$

The linear operator G' is the so-called tangent linear model (TLM) and its adjoint G'^* represents the adjoint model. With the gradient information thus obtained, the cost function can be efficiently minimized with respect to the control variables χ.

Both the tangent linear and the adjoint of the propagation model G can be obtained analytically from the governing differential equations by means of variational calculus. Alternatively, the model G can be described as the composition of a number of successive operations[18–20]

$$G = C_n \cdots C_2 C_1. \tag{2}$$

The derivative G' can then be obtained by differentiation according to the chain rule[21]

$$G' = C_n' \cdots C_2' C_1', \tag{3}$$

where, for each m, C_m' is the linear operator obtained by differentiation of C_m. The adjoint G'^* of this product of operators is by definition the product of the adjoints, taken in reverse order

$$G'^* = C_1'^* C_2'^* \cdots C_n'^*. \tag{4}$$

According to Eq. (4) the computation of G'^* requires that the adjoints of the operations which make up G have to be performed in reverse order. In particular, if G represents the range marching solution algorithm of the propagation model, the operators C_m may describe a succession of elementary range step integrations (in which case the corresponding adjoint integration will be performed backward in range). Likewise, the operations C_m do not necessarily need to represent the full operator that marches the model forward in range, but equally well small fractions of the model, which in the following will be referred to as modules.

This modular decomposition of the propagation model by means of a modular graph and the respective adjoint calculation scheme is covered in greater detail in the next section. For a more detailed overview of the different adjoint approaches in general and their applications also in other related fields the reader is referred to Refs. 22–26 and the review.[27]

B. Modular graph approach

1. General concept

The modular graph concept presented in this section has its origins in general graph theory[28,29] and is mainly based on the idea that it is much more efficient to decompose a complex problem into a set of simpler subproblems than to solve the full problem directly.

Rather than deriving the adjoint directly in the continuous domain by analytical transformation of the governing differential equations and subsequent programming of the adjoint code,[7] the operators C_m in Eq. (2) can be viewed as representations of individual lines of the numerical code describing the propagation model.

Variables and syntax of the numerical code can either be identified automatically by means of code parsing, data flow

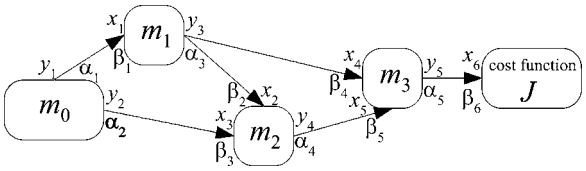


FIG. 1. Example of a modular graph. For each module m_n the input and output variables are denoted by x_k and y_j respectively. α_k and β_j represent the corresponding Lagrange multipliers that are used for the adjoint generation scheme in Sec. II B2.

analysis, and code assignment or, as described in the following, provided by the user in the form of a modular graph.

The so-called modular graph is a process and data flow diagram which describes the underlying acoustic model. It consists of a complete set of interconnected modules m, where the input of each module m_n is provided by the output of its predecessors $m_{p,p<n}$ (Fig. 1). In general, the first module m_0 provides the propagation model with the initial data and as such has no formal input variables x_k whereas the last module simply calculates the cost function J and has no other output variables y_j. The Lagrange multipliers, denoted here by α_j and β_k, are explained in greater detail in the next section.

There are no restrictions as to the size of each module; from a practical point of view the decomposition may depend on the module semantic. Each module represents a differentiable function, which may be a simple function or a complex one represented in turn by a subgraph. Especially, modifying the model, the cost function, or the adjoint at any time is straightforward due to the modular graph structure.

If the underlying model is properly decomposed by the user into a number of differentiable modules m, Eq. (4) in combination with the modular graph methodology presents a convenient way to generate the adjoint by encoding the local Jacobian and backpropagating the result to the preceding modules.

2. Lagrangian formalism

For the derivation of the reverse adjoint calculation scheme using Lagrangian formalism, $\mathcal{M},\mathcal{I},\mathcal{O}$, and $\mathcal{P}$ shall denote in the following the complete sets of indices of all modules, module input variables, module output variables and module parameters of the system, respectively. One can then define three mappings

$$X:\mathcal{I}\to\mathcal{M}, \quad k\mapsto X(k), \tag{5}$$

$$Y:\mathcal{O}\to\mathcal{M}, \quad j\mapsto Y(j), \tag{6}$$

$$W:\mathcal{P}\to\mathcal{M}, \quad i\mapsto W(i), \tag{7}$$

where $X(k)$ represents the index of the module for which k is the index of one of its input variables, $Y(j)$ the index of the module for which j is the index of one of its output variables, and $W(i)$ returns the corresponding index of the module for which i is the index of one of the module parameters. With these definitions the module m_n can be formally defined as

$$\forall j \in Y^{-1}(n), \quad y_j = f_j[(x_k)_{k\in X^{-1}(n)},(w_i)_{i\in W^{-1}(n)}]. \tag{8}$$

This is the formal statement of the constraint that each output variable y_j of a given module m_n be defined as a function f_j of the input variables $(x_k)_{k\in X^{-1}(n)}$ and the parameters $(w_i)_{i\in W^{-1}(n)}$ of that module. At the same time each input variable x_k of the module m_n is required to emanate from one and only one output variable of a preceding module $m_{l,l<n}$. This can be formally expressed as

$$\forall k \in X^{-1}(n), \quad x_k = y_{\phi(k)} \tag{9}$$

with

$$\phi(k) \in \bigcup_{p=0}^{n-1} Y^{-1}(p). \tag{10}$$

The Lagrangian $\mathcal{L}$ of the system can then be defined as the cost function J measuring the fit between the model result and the observations subject to the two constraints formulated in Eqs. (8)–(10)

$$\mathcal{L} = J - \sum_{j\in\mathcal{O}} \alpha_j\{y_j - f_j[(x_k)_{k\in X^{-1}[Y(j)]},(w_i)_{i\in W^{-1}[Y(j)]}]\} - \sum_{k\in\mathcal{I}} \beta_k(x_k - y_{\phi(k)}). \tag{11}$$

The Lagrange multipliers α_j and β_k can thus be obtained via

$$\frac{\partial\mathcal{L}}{\partial y_j} = -\alpha_j + \sum_{k\in\phi^{-1}(j)} \beta_k = 0 \tag{12}$$

and[30]

$$\frac{\partial\mathcal{L}}{\partial x_k} = -\beta_k + \sum_{j\in Y^{-1}[X(k)]} \alpha_j \frac{\partial f_j}{\partial x_k} = 0 \tag{13}$$

as

$$\alpha_j = \sum_{k\in\phi^{-1}(j)} \beta_k \tag{14}$$

and[30]

$$\beta_k = \sum_{j\in Y^{-1}[X(k)]} \alpha_j \frac{\partial f_j}{\partial x_k}. \tag{15}$$

The reverse calculation (backpropagation) of the Lagrange multipliers via Eqs. (14) and (15) is initiated at the last module for which β_k simply reduces to[30]

$$\beta_k = \frac{\partial J}{\partial x_k}. \tag{16}$$

Once all Lagrange multipliers $\{\alpha_j,\beta_k\}$ of the system are computed, the Lagrangian formalism allows the calculation of the local gradient of the cost function J with respect to any given model parameter w_i as

$$\frac{\partial J}{\partial w_i} = \sum_{j \in Y^{-1}[W(i)]} \alpha_j \frac{\partial f_j}{\partial w_i}. \tag{17}$$

Based on this reverse modular graph formalism an algorithmic tool can thus facilitate the generation and coding of the adjoint of the complex acoustic propagation model in accordance with Eq. (4).

YAO,[8] the tool that is used in this work further provides several routines to test the validity of the local derivatives of the different modules, the cost function and an automatic validation can also be performed for the tangent linear and the adjoint model.

In terms of memory requirements, it is in principle necessary for the reverse computation of the adjoint to store the results of all the intermediary computations leading from χ to $G(\chi)$ in memory since the coefficients of each linear operator C'^{*}_{m} depend on the results of the corresponding m-th step in the direct integration.[19] Alternatively, the values of the intermediary variables can be successively recomputed at each step of the reverse adjoint generation.

III. FORWARD MODEL

In continuation of the analytical optimal control approach introduced in Ref. 7 the acoustic propagation model G that is chosen to demonstrate the adjoint-based inversion approach for the physical characterization of a shallow water environment is the wide-angle PE due to Claerbout.[31–33] For a stratified medium with varying density $\rho(z)$, sound speed $c(z)$, and absorption loss $\alpha(z)$ the wide angle PE can be obtained as

$$2ik_0\left(1+\frac{1}{4}(N^2-1)\right)\frac{\partial \psi}{\partial r} + \rho\frac{\partial}{\partial z}\left(\rho^{-1}\frac{\partial \psi}{\partial z}\right) + \frac{i}{2k_0}\rho\frac{\partial}{\partial z}\left(\rho^{-1}\frac{\partial^2 \psi}{\partial z\,\partial r}\right) + k_0^2(N^2-1)\psi = 0, \tag{18}$$

where $k_0=\omega/c_0$ is a reference wave number, $N(z)=n(z)[1+i\alpha(z)]$ and $n(z)=c_0/c(z)$ the refractive index. The field ψ therein is related to the complex pressure p according to

$$p(r,z) = \frac{\psi(r,z)\exp(ik_0 r)}{\sqrt{k_0 r}}. \tag{19}$$

The boundary condition at the water-bottom interface $z_b=H$ is determined by a discrete nonlocal impedance boundary condition that exactly transforms the PE problem having a transverse radiation condition at infinity, into an equivalent one in a bounded domain. Dividing the interval $0\rightarrow r+\Delta r$ into $L+1$ intervals of width Δr, Yevick and Thomson[15] obtain the required impedance boundary condition

$$\left[\frac{\partial}{\partial z} - i\beta\right]\psi[(L+1)\Delta r, z_b] = i\beta\sum_{j=1}^{L+1} g_{1,j}\psi[(L+1-j)\Delta r, z_b] \tag{20}$$

with the convolution coefficients $g_{1,j}$ and[7]

$$\beta = \frac{\rho_w}{\rho_b}k_0\sqrt{\frac{(N_b^2-1)\left(1+\frac{1}{4}\nu^2\right)+\nu^2}{\left(1+\frac{1}{4}\nu^2\right)}}, \tag{21}$$

where $\nu^2=4i/k_0\Delta r$, and the subscripts w and b indicate the water column and bottom, respectively. Further details of the NLBC derivation including algebraic expressions for the coefficients $g_{1,j}$ are given in the Appendix. Equation (20) accounts for the total impedance jump (sound speed, attenuation, and density) encountered by waves that cross the lower boundary of the waveguide and allows the boundary value of the acoustic field at a given range-propagation step to be determined in terms of the history of boundary-field values for all previous range steps.[15]

To summarize, the system of the wide-angle PE model including a Dirichlet boundary condition at the surface, an analytical Thomson's source term,[33,34] and the nonlocal boundary condition at the bottom can be described as follows:

$$\begin{cases} 2ik_0\left[1+\frac{1}{4}(N^2-1)\right]\frac{\partial \psi}{\partial r} + \rho\frac{\partial}{\partial z}\left(\rho^{-1}\frac{\partial \psi}{\partial z}\right) \\ \quad + \frac{i}{2k_0}\rho\frac{\partial}{\partial z}\left(\rho^{-1}\frac{\partial^2 \psi}{\partial z\,\partial r}\right) + k_0^2(N^2-1)\psi = 0 \\ \psi(r,z=0) = 0 \\ \psi(r=0,z) = S(z) \\ \text{NLBC} \quad \left[\frac{\partial}{\partial z} - i\beta\right]\psi(z)\Big|_{z=H} = i\beta\sum_j g_{1,j}\psi_j(z)\Big|_{z=H} \end{cases} \tag{22}$$

For convenience the notation for the NLBC is simplified here by dropping the range coordinate and using $\psi_j(z_b)=\psi[(L+1-j)\Delta r, z_b]$.

In the following the control variables χ defined in Sec. II A will be determined by the sound speed profile in the water column $c(z)$[via $n^2(z)=c_0^2/c^2(z)$] (OAT, see Sec. V A), the geoacoustic parameters $\{\rho_b, c_b, \alpha_b\}$ of the sediment (GI, see Sec. V B), or a combination of both.

In contrast to the analytic control approach presented in Ref. 7 where it was necessary to introduce a generalized impedance boundary coefficient $F(r)$ in order to represent the complex convolution sum on the right-hand side of Eq. (20), the numerical adjoint approach enables direct inversion for the geoacoustic parameters.

A. Numerical implementation

The finite difference implementation of the direct problem given in Eq. (22) is an implicit Crank-Nicolson scheme. For the discretization in depth, an offset grid is used to avoid the need to compute the field along the top and bottom of the computational domain. The use of a heterogeneous finite-difference approximation of the variable-density operator in Eq. (18) further precludes the need for explicitly enforcing continuity of pressure and vertical particle velocity at any jump discontinuities in the medium properties.[33,35]

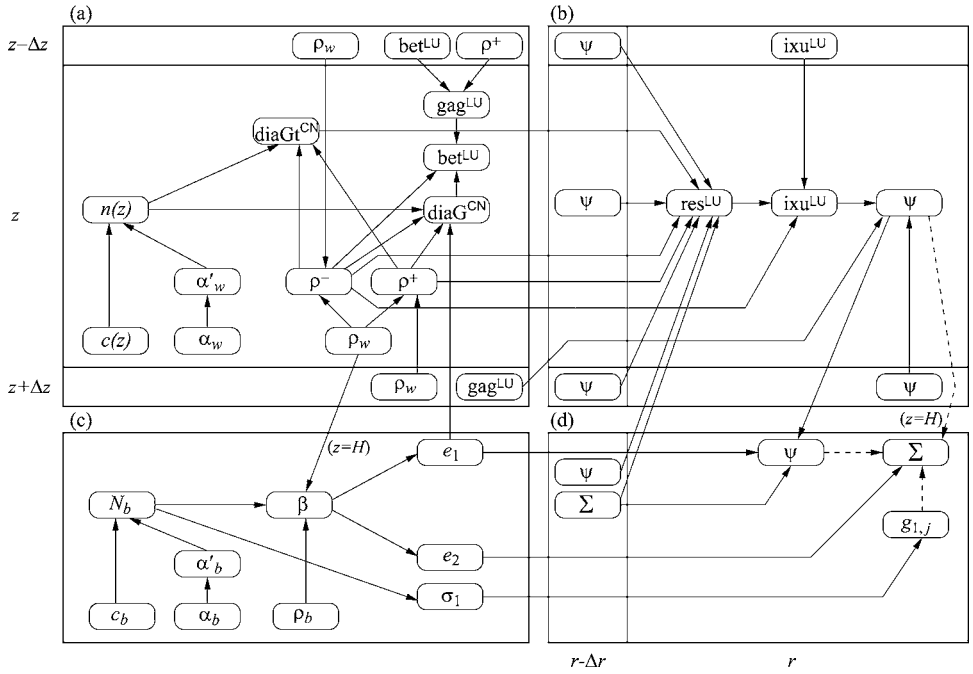


FIG. 2. Modular graph representation of the WAPE NLBC model. The nomenclature is consistent with the notation in Sec. III, particularly Eqs. (18), (20), (21), and (27). Modules with the superscript "LU" or "CN" implement the LU decomposition (Ref. 39) and the Crank-Nicolson scheme, respectively. Module "Σ" refers to the summation of the boundary-field values in Eq. (27).

The NLBC in Eq. (20) is treated as a first order ODE in depth and as such integrated with respect to the depth z.[36] Multiplication of both sides by $e^{-i\beta z}$ yields

$$\frac{\partial}{\partial z}[e^{-i\beta z}\psi(z)] = i\beta e^{-i\beta z}\sum_j g_{1,j}\psi_j(z), \tag{23}$$

and subsequent integration over the depth cell that is adjacent to the boundary at $z_b=H$ can be approximated[37] as

$$[e^{-i\beta z}\psi(z)]_{H-1/2\Delta z}^{H} \approx -\sum_j g_{1,j}\psi_j(H-1/4\Delta z) \times [e^{-i\beta z}]_{H-1/2\Delta z}^{H}. \tag{24}$$

Multiplying both sides by $e^{i\beta H}$ then gives

$$\psi(H) - e^{i\beta(1/2\Delta z)}\psi(H-1/2\Delta z) = (e^{i\beta(1/2\Delta z)}-1)\sum_j g_{1,j}\psi_j(H-1/4\Delta z) \tag{25}$$

$$= e^{i\beta(1/4\Delta z)}2i\sin(1/4\beta\Delta z)\sum_j g_{1,j}\psi_j(H-1/4\Delta z). \tag{26}$$

Substituting $\psi_j(H-1/4\Delta z)$ by $[\psi_j(H)+\psi_j(H-1/2\Delta z)]/2$ finally allows the calculation of the field on the boundary

$$\psi(H) = \underbrace{e^{i\beta(1/2\Delta z)}}_{\doteq e_1}\psi(H-1/2\Delta z) + \underbrace{ie^{i\beta(1/4\Delta z)}\sin(1/4\beta\Delta z)}_{\doteq e_2}\sum_j g_{1,j}[\psi_j(H) + \psi_j(H-1/2\Delta z)]. \tag{27}$$

Following the discretization of the direct WAPE system, the forward model can then be decomposed according to the modular graph concept described in Sec. II B. The resulting modular graph (Fig. 2) consists of 26 modules and is divided into four blocks (a)–(d), each of which can be further subdivided vertically and/or horizontally. Horizontal layering within a block indicates adjacent finite difference depth cells $(z, z\pm\Delta z)$ and vertical subdivision represents successive range steps $(r-\Delta r, r)$. The blocks (a)–(d) represent four different dimensional spaces; block (a) mainly serves for the initialization of the environmental parameters $[\alpha_w, \rho_w, c(z), n(z)]$ in the water column and the setup of the tridiagonal finite difference matrices (diaGt, DiaG), block (b) represents the actual range marching solution for the field ψ in the water column via LU decomposition (res, ixu).[39] Blocks (c) and (d) represent the corresponding counterparts for the initialization (c) and calculation (d) of the field in the bottom in accordance with the NLBC [Eqs. (20), (21), (27)]. The dashed arrow further indicates that the module Σ which represents the summation of the boundary-field to values in Eq. (27) depends on all the known values (history) of the source modules at previous range steps, not just on the actual value of the current instance. Compared to the analytical adjoint derivation in Ref. 7, the modular-graph approach clearly allows for a more systematic and efficient treatment of complex, general structures such as those associated with the nonlocality (spatial memory) of the boundary condition.

Further technical details and a more comprehensive description of the modular decomposition of the WAPE system can be found in Ref. 40.

IV. OPTIMIZATION ALGORITHM

A. Cost function

With YAO, the cost function is calculated automatically from the modules which are declared as cost modules (i.e., forecast state) and from the observations that are loaded from an external file

$$J(\chi) = \tfrac{1}{2}(G(\chi)-\psi_{\text{obs}})^T R^{-1}(G(\chi)-\psi_{\text{obs}}), \tag{28}$$

with χ the control parameters $[\chi=\{\rho_b, c_b, \alpha_b\}$ in geoacoustic inversion and $\chi=c(z)$ in ocean acoustic tomography], G the numerical model, and ψ_{obs} the observation.

J represents the least squares difference between the observations and their equivalent model results balanced by the covariance matrix R of the corresponding error.

Using the formalism derived for the regularization of an adjoint-based optimization in Ref. 7 one may consider the application of a minimum norm criterion to the gradient of the control parameter $\|\nabla\chi\|^2$ as a side constraint to the cost function to ensure maximum smoothness of the inverted solution.

An *a priori* estimate χ_{apr} of the solution can also be included as an additional side constraint in the cost function. Hence, with the two side constraints, the quantity to be minimized[41,42] becomes

$$J(\chi) = \tfrac{1}{2}[G(\chi) - \psi_{\text{obs}}]^T R^{-1}[G(\chi) - \psi_{\text{obs}}] + \tfrac{1}{2}a(\chi - \chi_{\text{apr}})^T B^{-1}(\chi - \chi_{\text{apr}}) + \tfrac{1}{2}b\,\|\nabla\chi\|^2, \quad (29)$$

where B represents the covariance matrix for the control parameter χ of the medium and (a,b) are the respective regularization parameters. In particular, if a prior measurement of the control parameter is available, e.g., in OAT a previous measurement of the sound speed profile [e.g., from a (CTD) cast] in the area under consideration, the cost functional in Eq. (29) forces the inverted profile to be close to this first guess.

A further extension of the cost function to multiple frequencies provides valuable additional observations especially in the case of sparse array observations and complex bottom conditions.[44,45] Considering a model with m different frequencies $f_i, i=1,\ldots,m$ and $\psi_{\text{obs},i}, i=1,\ldots,m$ the corresponding observations at each frequency, the cost function in Eq. (28) becomes

$$J(\chi) = \sum_{i=1}^{m} J_i(\chi) \quad (30)$$

$$= \sum_{i=1}^{m} \frac{1}{2}[G_i(\chi) - \psi_{\text{obs},i}]^T R_i^{-1}[G_i(\chi) - \psi_{\text{obs},i}] \quad (31)$$

and analogously for Eq. (29). The overall gradient of J is then obtained via

$$\nabla J(\chi) = \sum_{i=1}^{m} \nabla J_i(\chi). \quad (32)$$

An extensive analytical treatment of multiple-frequency adjoint-based inversion of a locally reacting impedance boundary condition for the standard PE can be found in Ref. 16.

B. Minimizer

With a cost function specified in Eqs. (28)–(31) the numerical implementation of the direct model described in Sec. III A can be differentiated using YAO in reverse mode to generate the adjoint code. Equation (17) then allows the computation of the gradient of the cost function with respect to the control variable χ. A Taylor test ensures that the derivatives generated with the adjoint code agree with the corresponding finite difference approximations for different directions of perturbation of the control variable. The minimization problem can be solved efficiently through the use of standard, iterative gradient methods, e.g., conjugate gradient or Newton-type methods.[43] The two routines M1QN3 and M2QN1 (Ref. 10) used here for the optimization process in the following sections are based on quasi-Newton methods. As for any optimization by a gradient method, initial data has to be provided that are not too far from true solutions. For an initial guess too far from the true solution gradient methods may converge to a local minimum, especially if the cost function is nonconvex.[43] Routine M1QN3, a solver for large-scale unconstrained minimization problems implements the limited-memory quasi-Newton technique L-BFGS.[10,11] The step-size is determined by means of the Fletcher-Lemaréchal algorithm, which in turn realizes the Wolfe conditions. Routine M2QN1, a solver of bound constrained minimization problems is designed to minimize functions depending on a small or medium number of variables x subject to bound constraints ($a <= x <= b$) and implements a quasi-Newton (BFGS) technique with line search. Both codes have been used in a number of applications in meteorology and oceanography, geophysical tomography, shape optimization, and seismic inversion.

V. INVERSION RESULTS

In order to demonstrate the performance of the semiautomatic, adjoint approach this section deals with the physical characterization of a shallow water environment through two types of inverse problems: ocean acoustic tomography (Sec. V A) and geoacoustic inversion (Sec. V B).

The Yellow Shark experiments,[44,45] carried out in the Giglio basin off the west coast of Italy during the fall of 1994 (YS94), are used in the following as a realistic test case to validate the adjoint-based inversion technique. The geometric and acoustic parameters used as an input for the WAPE forward model are identical to one of the runs along the main transect of the experiment: the 9-km run on September 10, 1994. A 200–800-Hz frequency modulated signal was emitted from a sound source at 69.2-m depth; the water depth was 113.1 m. The transmitted signal was received on a vertical array (VRA) of 32 hydrophones 2-m spaced from 37.2- to 99.2-m depth. The acoustic fields are synthesized accordingly with the WAPE model using ocean sound speed profiles and bottom geoacoustic properties measured *in situ*. The bottom is modeled as described in Ref. 46 and consists of a 7.5-m thick clay sediment layer with a compression-speed gradient ($\rho=1.5$ g cm^{-3}, $c=1470$ m s^{-1}, $g=2$ s^{-1}, $\alpha=0.03$ dB λ^{-1}), and a homogeneous fluid halfspace modeling a silty-clay sediment ($\rho=1.8$ g cm^{-3}, $c=1530$ m s^{-1}, $\alpha=0.15$ dB λ^{-1}).

A. Ocean acoustic tomography

The ocean acoustic tomography (OAT) application considered in this section is particularly suitable for a shallow water environment since it takes into account the bottom conditions, i.e., sound interaction with the bottom is modeled correctly. OAT results are presented which were obtained with seven different source frequencies {200, 250, 315, 400,

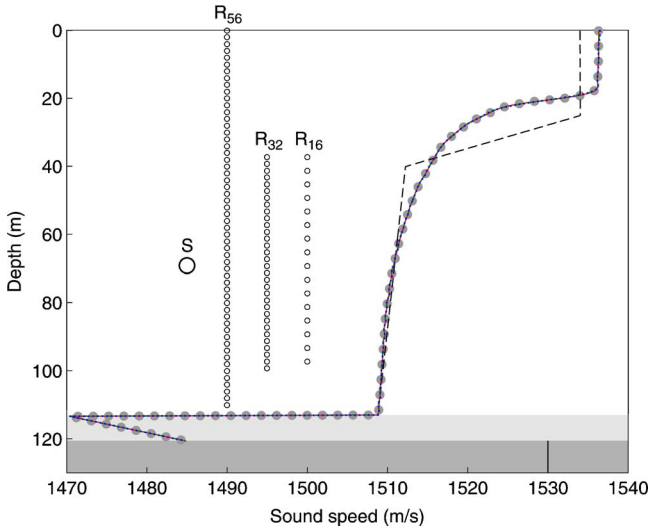


FIG. 3. (Color online) Regularized OAT results for three different VRA configurations at a range of 1.5 km, seven source frequencies $f=\{200,250,315,400,500,630,800\ \text{Hz}\}$ and a piecewise linear analytic initial SSP (dashed, black). The inverted profiles were obtained with a 56-element VRA (R_{56}) spanning the full water column from 0.1 to 110.1 m depth ($\Delta h=2$ m) (solid), a 32- (R_{32}) and a 16-element (R_{16}) VRA covering the column from 37.1 to 99.1 m depth, with $\Delta h=2$ m (dash-dot), and $\Delta h=4$ m (dotted), respectively. The source depth (S) and the individual hydrophone depths for the different arrays are indicated with small circles. The gray regions represent a clay layer overlying a silty clay subbottom; the bottom compression speed profile is indicated. All inverted profiles are almost identical with the true SSP (large dots, gray). The inversion errors are better visualized in Fig. 5.

500, 630, 800 Hz} and a VRA at a range of 1.5 km with three different configurations. For an efficient calculation of the overall gradient of the multiple-frequency cost function according to Eq. (32), the same finite difference grid is used for the forward and adjoint calculations at each frequency. The range and depth discretization steps $\Delta r=0.5$ m and $\Delta z=0.2$ m will be used for the GI applications in Sec. V B as well.

The inversion for the sound speed profile (SSP) is initialized with a piecewise linear analytical function. The ensemble average of all sound speed profiles measured within the vertical section during the acoustic transmissions is taken as the true profile whereas the ocean bottom is represented by the true geoacoustic properties obtained from sediment cores and through previous inversion of acoustic data using multitone matched-field and model-based matched filter processing.[44,45]

Figure 3 shows both the piecewise linear initial and the true profile versus depth, and the three VRA configurations used for the test; the source and the individual hydrophones are represented with small circles. Regularized OAT results are compared for a 56-element VRA (R_{56}) spanning the full water column with an aperture of 110 m and interelement spacing of $\Delta h=2$ m, and VRAs of 32 (R_{32}) and 16 (R_{16}) hydrophones spanning an aperture of 62 m in the lower part of the column. These configurations were chosen to compare the 32-element array of the actual experimental setup, where both source and VRA were positioned below a well-developed thermocline, to a smaller array with the same aperture but with twice the hydrophone spacing and a larger array covering in full the water column.

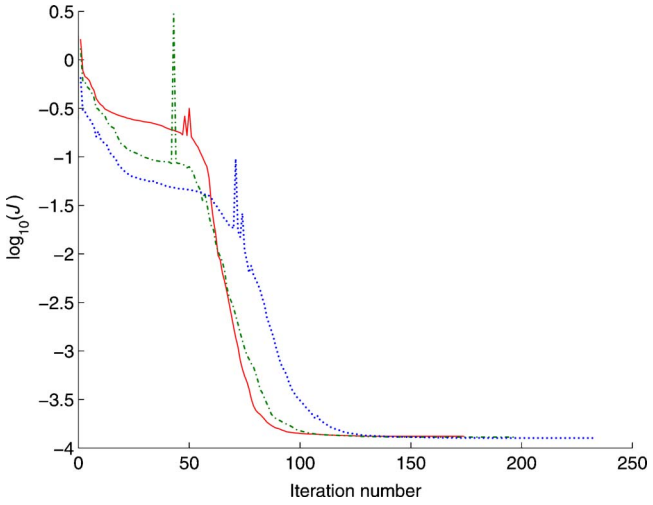


FIG. 4. (Color online) Comparison of the evolution of the log-scale cost function vs iteration number for the different array configurations shown in Fig. 3: 56-element, full water column (solid); 32-element (dash-dot), and 16-element (dotted), lower part of water column.

The initial profile is continuous but only piecewise differentiable and differs significantly from the true profile (Fig. 3). Nevertheless, the estimated profiles are almost identical with the true SSP for all three array configurations. Even for partial coverage of the water column with the 32- and 16-element array the inversion process succeeds in retrieving the true profile over the full water column. Regarding performance of the adjoint approach, Fig. 4 compares the evolution of the log-scale cost function versus iteration number for the three array configurations. As expected the required number of iterations is inversely proportional to the array size. However, good convergence is obtained in linear scale with all three array sizes after approximately 70 iterations. A comparison of the inversion errors between the true and estimated profiles is shown in Fig. 5 for the three array configurations. In analogy with the numerical performance, also the inversion error is inversely proportional to the array size. The

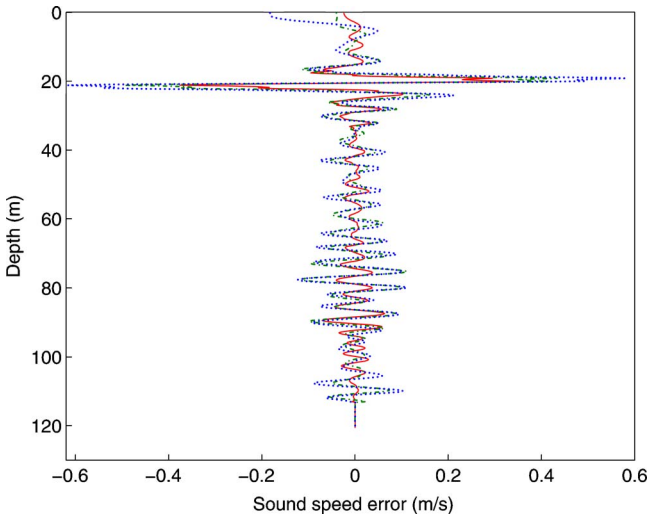


FIG. 5. (Color online) Comparison of the inversion errors between the true and the estimated profile for the 56-element VRA spanning the entire water column (solid), the 32-element (dash-dot), and the 16-element (dotted), VRA spanning the lower part of the column.

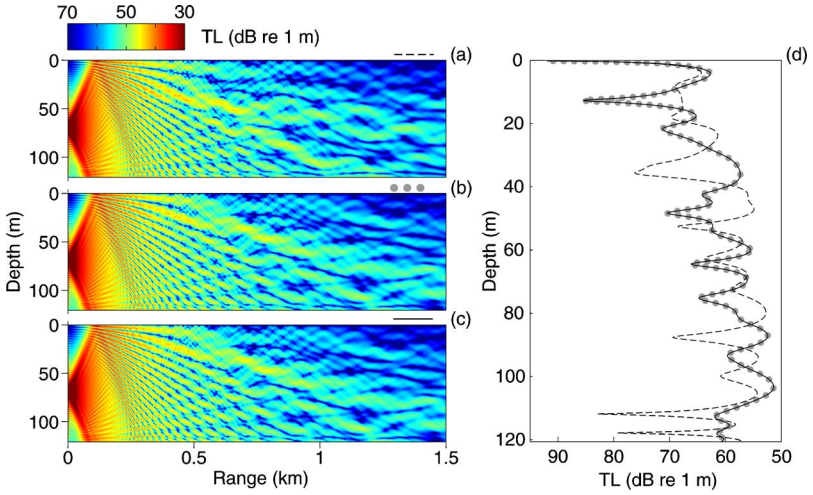


FIG. 6. (Color online) Initial (a), true (b), and inverted (c) acoustic field synthesized for the shallow water, muddy bottom site of the YS94 experiment (f =500 Hz) using the piecewise linear initial, the true and the estimated SSP obtained in Fig. 3. Transmission loss (TL) is shown. (d) Comparison of the TL vs depth at maximum range using the piecewise linear initial (dashed, black), the true (large dots, gray) and the estimated SSP (solid, black).

smallest error is obtained with full 56-hydrophone sampling of the water column. For all array configurations the greatest error occurs at the depth $z \approx 20$ m directly below the mixed layer where the sound speed gradient is maximal. Figure 6 shows the sound pressure fields that were computed for the shallow water, acoustically soft bottom site of the YS94 experiment with a source frequency of 500 Hz using the true, the piecewise linear initial, and the estimated SSP (Fig. 3). Figure 6(d) compares the three fields versus depth at maximum range. In all simulations a half angle of 40 deg is used for Thomson's source in Eq. (22). For comparison Fig. 7 shows the regularized inversion results obtained at a single frequency (500 Hz) for the sampling of all water depths with the 56-element array. The result clearly reveals that the inversion process at a single frequency is trapped in a local minimum and fails to retrieve the correct profile due to the lack of available field information even though the largest array is used. The same effect can be observed for the initial case of seven source frequencies (Fig. 3) when the array size is further reduced from 16 to 8 hydrophones (not shown). In these cases a combination of the quasi-Newton approach (based on the adjoint) and a global search method (e.g., Metropolis, genetic algorithms) in a hybrid optimization scheme is a possible solution to local convergence problems.

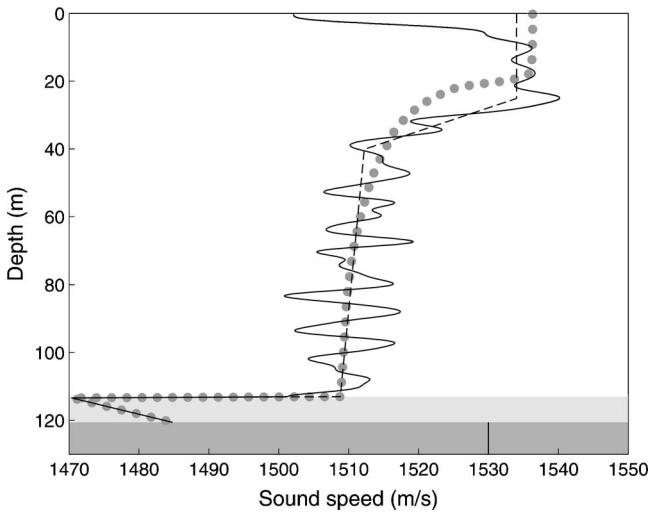


FIG. 7. OAT results using a 56-element VRA at 1.5 km range spanning the full water column between 0.1 and 110.1 m depth (Δh=2 m). For the inversion a single source frequency f=500 Hz was used. Starting with a piecewise linear analytic initial SSP (dashed, black) the inverted profile (solid, black) and the true (large dots, gray) differ significantly.

B. Geoacoustic inversion

This section presents the geoacoustic inversion (GI) results for the two cases of a known and an uncertain SSP in the water column. In a first example (Figs. 8–10) the 32-element VRA introduced in the Sec. V A is located at a range of 9 km and three different source frequencies $\{200, 400, 500\}$ Hz are used for the inversion of the geoacoustic parameters. The sound speed profile in the water column and the low-absorbing, fluid sediment layer are assumed to be known while the parameters $\{\rho_b, c_b, \alpha_b\}$ associated with the layer-subbottom interface are inverted for. In a second

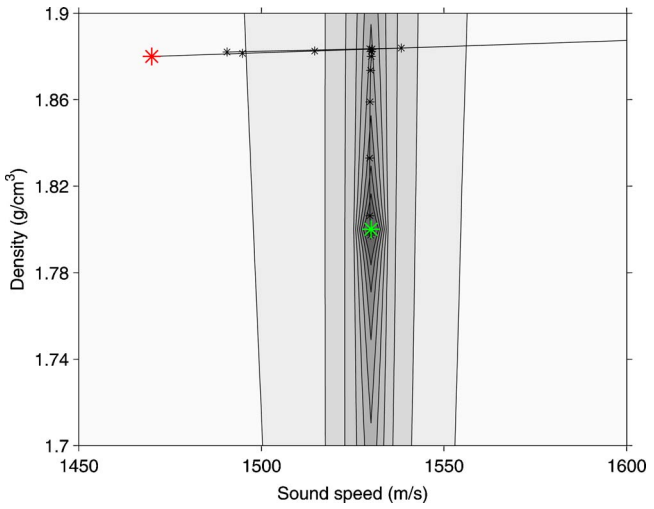


FIG. 8. (Color online) Contour plot of the log-scale cost function for the case of the 32-element VRA at 9-km range and three source frequencies $\{200, 400, 500\}$ Hz. The iterative bound constrained optimization process (routine M2QN1) is shown in the $\rho_b - c_b$ plane. For illustration purposes the sound attenuation is kept fixed at the true value α_b=0.15 dB λ^{-1}. The starting and ending points of the inversion process are marked with a red and green asterisk.

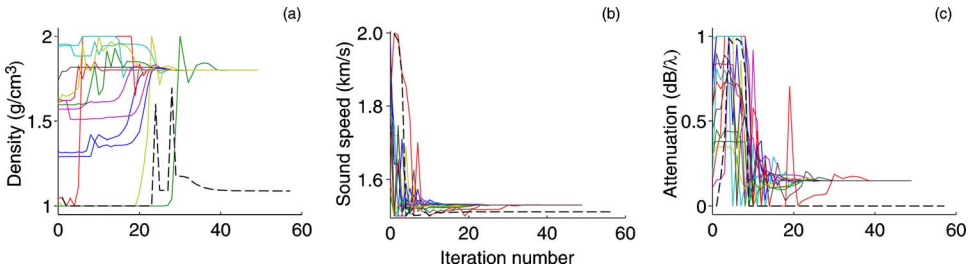


FIG. 9. (Color online) GI results for the 32-element VRA located below the thermocline at 9-km range, and for joint optimization across three source frequencies {200,400,500} Hz. The estimated geoacoustic parameters ρ_b, c_b, and α_b are shown as a function of the iteration step and for random combinations of the initial conditions. The dashed curves show a representative single-frequency (500 Hz) inversion result for one selected initial condition.

example (Figs. 11 and 12) the water-column SSP is included in the parameter search space as well, to adjust for an uncertain SSP during the geoacoustic inversion. For this purpose inversion results are compared for the three configurations of partial water-column spanning arrays with 32, 16, and 8 hydrophones, whereas the same frequency set and experimental geometry will be used as in the OAT case in Sec. V A.

In the first example, to test convergence random values are chosen for the initial conditions in the following intervals: for the density, $1\ \mathrm{g\ cm^{-3}} < \rho_{b,\mathrm{init}} < 2\ \mathrm{g\ cm^{-3}}$; for the compression speed, $1500\ \mathrm{m\ s^{-1}} < c_{b,\mathrm{init}} < 2000\ \mathrm{m\ s^{-1}}$ and for the sound absorption, $0\ \mathrm{dB\ \lambda^{-1}} < \alpha_{b,\mathrm{init}} < 1\ \mathrm{dB\ \lambda^{-1}}$. Figure 8 illustrates the iterative bound constrained optimization process in the ρ_b-c_b plane using routine M2QN1 (Sec. IV B): For illustration purposes the sound attenuation is kept fixed at the true value $\alpha_b = 0.15\ \mathrm{dB\ \lambda^{-1}}$ and the inversion is started with the initial values of $\rho_b = 1.88\ \mathrm{g\ cm^{-3}}$ and $c_b = 1470\ \mathrm{m\ s^{-1}}$. The convergence of all three geoacoustic parameters versus iteration number is shown in Fig. 9. Even though the joint optimization was performed across only three frequencies all three parameters converge to the true values for the many initial conditions chosen randomly. The control parameters further display a clear parameter hierarchy which relates to the relative sensitivity of the acoustic field to the physical parameters. The compressional speed converges first after approximately 20 iterations before the attenuation and the density; the latter converge after approximately 30 iterations. For comparison a set of representative single frequency (500 Hz) inversion results are included in Fig. 9 for one selected initial condition (dashed line). With a single-frequency cost function none of the three geoacoustic parameters can be retrieved correctly.

As an illustration Fig. 10 contains a comparison of the true, the inverted, and the initial field over the entire range and depth for one of the initial conditions shown in Fig. 9. Figure 10(d) then compares the three acoustic fields versus depth at the receiver range. As a concluding example Figs. 11 and 12 show the GI results that include the effects of uncertainty in the range-average SSP of the water column. The latter is added as an additional unknown to the parameter search space. One of the CTD profiles that were measured along the transect during the acoustic transmissions is used as an initial estimate. This particular profile is chosen since it is the one that most deviates from the ensemble average which serves as the true solution. Like in the OAT case in Sec. V A the combined GI/OAT results are obtained with three different VRAs at a range of 1.5 km using the same seven source frequencies as before. As regards the array configurations three partial water column spanning VRAs are used with 32, 16, and 8 hydrophones, 2-, 4-, and 8-m spaced at depths between 37.2–99.2 m. Figure 11 shows both the initial and the true profile and compares the three estimated profiles that were corrected during the GI run. A detailed view in the left part of Fig. 11 further displays the inversion error between the estimated and the true profiles. The overall performance of the combined inversion scheme is presented in Fig. 12. For the evolution of the estimated SSP versus iteration number the depth-integrated error

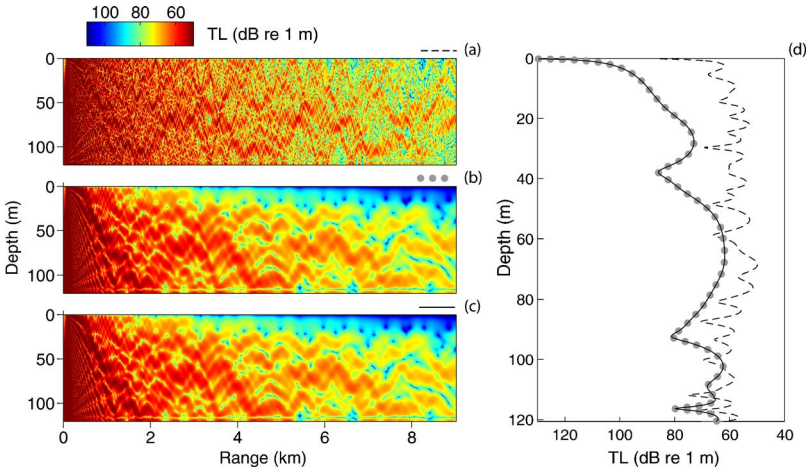


FIG. 10. (Color online) Initial (a), true (b), and inverted (c) acoustic pressure field at 500 Hz. Transmission loss (TL) is shown. The initial conditions are $c_b = 1795\ \mathrm{m\ s^{-1}}$, $\alpha_b = 0.95\ \mathrm{dB\ \lambda^{-1}}$, and $\rho_b = 1.05\ \mathrm{g\ cm^{-3}}$. (d) Comparison of the TL depth at maximum range using the initial (dashed, black), the true (large dots, gray), and the estimated (solid, black) geoacoustic parameters.

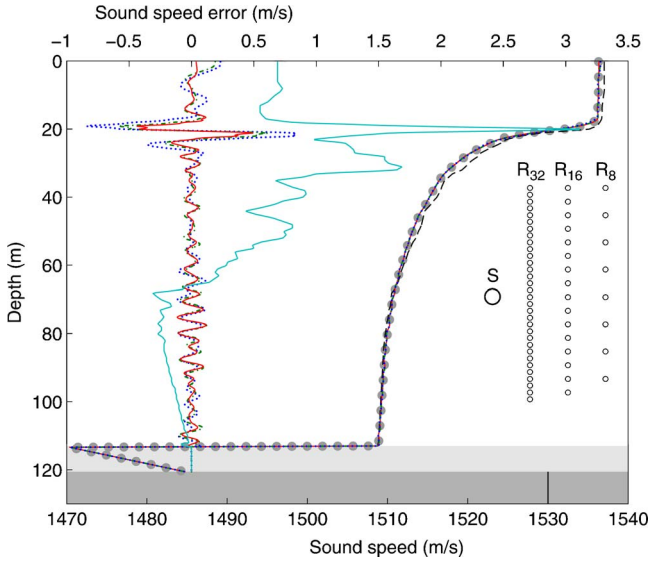


FIG. 11. (Color online) Correction for an uncertain SSP during GI using three partial water column spanning VRAs with 32 (R_{32}), 16 (R_{16}) and 8 (R_8) hydrophones at 1.5 km range. For the inversion seven frequencies were used (see Fig. 3); the initial SSP profile (dashed, black) is calculated from the CTD cast that deviates the most from the true ensemble average (large dots, gray). The inverted profiles are obtained with a 32 (solid), a 16 (dash-dot), and an 8-element (dotted), VRA. Left: Inversion errors for the three estimated and the initial profile (solid, turkey) (top scale).

$$\text{Error}[c(z)] = \frac{\sqrt{\sum_i |c_{\text{true}}(z_i) - c_{\text{calc}}(z_i)|^2}}{\sqrt{\sum_i |c_{\text{true}}(z_i)|^2}} \tag{33}$$

is calculated at each iteration. Upon completion of the inversion process both the geoacoustic parameters of the bottom and the sound speed profile of the water column are retrieved correctly. For the eight-hydrophone array a small degradation of the inversion results can be observed. Inversion error and convergence rate are again found to be inversely proportional to the array size and also in this case a clear parameter hierarchy can be observed: Attenuation α_b and density ρ_b of the bottom initially remain constant until the bottom compression speed c_b and the water-column SSP are preliminarily adjusted. In overall order the first parameter to converge is c_b before α_b, the water-column SSP, and finally ρ_b.

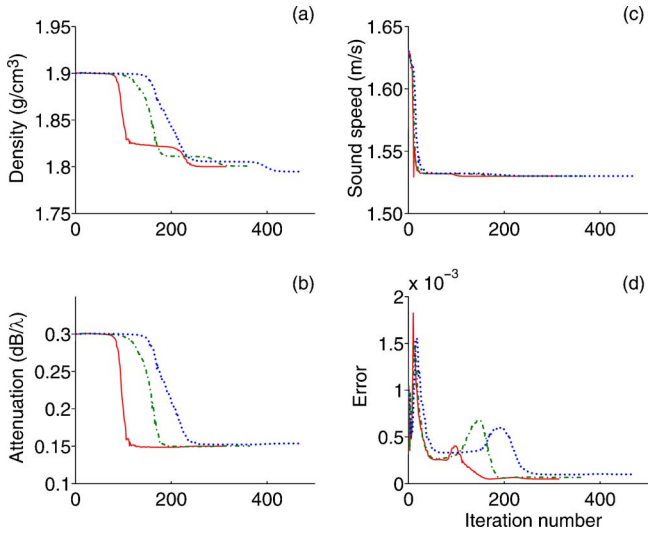


FIG. 12. (Color online) Results of geoacoustic inversion combined with simultaneous correction for an uncertain SSP. The evolution of the estimated geoacoustic parameters vs iteration number is shown together with the depth-integrated error of the water-column SSP in Eq. (33). Inversion results correspond to the 32 (solid), the 16 (dash-dot), and the 8-element (dotted), VRA in Fig. 11.

VI. CONCLUSION AND PERSPECTIVE

The semiautomatic adjoint approach described in this paper facilitates the generation and coding of the adjoint WAPE model using Yevick and Thomson's NLBCs and allows for a direct control of either the sound speed profile of the water column $c(z)$ or the geoacoustic parameters $\{\rho_b, c_b, \alpha_b\}$ embedded in the NLBC. In both cases the control variables represent physical parameters that can be used to characterize an unknown ocean environment and to construct a geoacoustic model of the material properties of the bottom.

In a previous paper[7] the authors applied an analytical optimal control approach to retrieve generalized coefficients $\{\beta, F(r)\}$ of the nonlocal impedance boundary condition. Those can be used for model tuning to correctly predict the acoustic propagation without knowing the physical parameters of the environment. Such a "through-the-sensor" inversion approach allows, e.g., to generate an effective bottom model for use in sonar signal processing algorithms, but not to obtain the underlying geoacoustic (or geophysical) parameters of the sediment.

In the present paper the feasibility of the semiautomatic adjoint approach is demonstrated with two applications in OAT and GI using environmental data collected in Mediterranean shallow waters. The performance of the multifrequency inversion process is studied for different experimental configurations, particularly taking into account sparse array geometries and partial coverage of the water column.

Furthermore, a combined GI/OAT inversion is proposed that includes the SSP as an additional unknown in the parameter space. Uncertainty in the range-average sound speed profile of the water column section, as typically due to temporal or spatial variability, can significantly degrade the inversion results. For this purpose an initial estimate of the SSP based on a previous CTD measurement has been efficiently included in an augmented cost function through the use of regularization. By means of adjoint modeling exact gradient information can be obtained to determine the sediment model parameters of the half-space bottom and at the same time adjust for the SSP uncertainty.

In a next step following the present demonstration of feasibility the aim is to test the adjoint PE inversion method with experimental acoustic data. Particularly for the case of sparse receiver arrays, a combination of a global search method (e.g., Metropolis, genetic algorithms) and a quasi-Newton approach (based on the adjoint) into a hybrid optimization scheme is a possible solution to prevent convergence to local optima.

Work is currently ongoing to include the effect of shear in the sediment and especially for the handling of real data, more sophisticated regularization schemes need to be applied.

ACKNOWLEDGMENTS

The authors wish to thank David J. Thomson for his helpful collaboration on discrete nonlocal boundary conditions and for sharing his PE programming code. Special thanks also to Charles Sorror, Sylvie Thiria, and Fouad Badran (LOCEAN, Univ. Paris VI) who provided valuable assistance with the implementation of the YAO code. The use of optimization routines M1QN3 and M2QN1 was kindly granted by the Institute National de Recherche en Informatique et Automatique (INRIA). The research work presented in this paper is supported by the Royal Netherlands Navy, the Service Hydrographique et Océanographique de la Marine Française in the framework of AO-BUOY joint research project with NATO Undersea Research Centre, Italy, and AQUATERRA integrated project funded by the European 6th Framework Programme.

APPENDIX: ALGEBRAIC EXPANSION OF THE COEFFICIENTS G_{1J} OF THE NLBC CONVOLUTION SUM

Starting from the z-space vertical wave number operator Γ_1 obtained for the WAPE NLBC in Ref. 15 [Eqs. (21) and 22) therein], both numerator and denominator are expressed in the following separately in powers of the translation operator $R=\exp(-\Delta r\partial_r)$

$$\sqrt{1+\sigma_1 R}=\sum_{k=0}^{\infty}\binom{1/2}{k}(\sigma_1 R)^k \tag{A1}$$

and

$$\frac{1}{\sqrt{1+\sigma_2 R}}=\sum_{k=0}^{\infty}\binom{-1/2}{k}(\sigma_2 R)^k, \tag{A2}$$

where

$$\sigma_1=\frac{(N_b^2-1)\left(1-\frac{1}{4}\nu^2\right)-\nu^2}{(N_b^2-1)\left(1+\frac{1}{4}\nu^2\right)+\nu^2} \tag{A3}$$

and

$$\sigma_2=\frac{1-\frac{1}{4}\nu^2}{1+\frac{1}{4}\nu^2}. \tag{A4}$$

Convolution of the coefficients finally gives the algebraic expressions for the coefficients $g_{1,j}$. Below the first five coefficients are given, the zeroth order coefficient $g_{1,0}$ which is equal to 1 is omitted, since it is tacitly carried over to the left-hand side in Eq. (20)

$$g_{1,1}=\tfrac{1}{2}(-\sigma_2+\sigma_1), \tag{A5}$$

$$g_{1,2}=\tfrac{1}{8}(3\sigma_2^2-2\sigma_2\sigma_1-\sigma_1^2), \tag{A6}$$

$$g_{1,3}=\tfrac{1}{16}(-5\sigma_2^3+3\sigma_2^2\sigma_1+\sigma_2\sigma_1^2+\sigma_1^3), \tag{A7}$$

$$g_{1,4}=\tfrac{1}{128}(35\sigma_2^4-20\sigma_2^3\sigma_1-6\sigma_2^2\sigma_1^2-4\sigma_2\sigma_1^3-5\sigma_1^4), \tag{A8}$$

$$g_{1,5}=\tfrac{1}{256}(-63\sigma_2^5+35\sigma_2^4\sigma_1+10\sigma_2^3\sigma_1^2+6\sigma_2^2\sigma_1^3+5\sigma_2\sigma_1^4+7\sigma_1^5). \tag{A9}$$

In practice, all the coefficients $g_{1,j}$ which are needed for propagating out to a maximum range can be calculated conveniently in terms of mathematical gamma functions. It is important to note that the coefficients $g_{1,j}$ depend in a nontrivial way on the geoacoustic parameters $\{c_b,\alpha_b\}$ via N_b in σ_1. This is exactly the reason why a direct inversion for these geoacoustic parameters is not analytically tractable. By means of generalized impedance parameters $\beta,F(r)$ this problem was avoided in the earlier papers.

[1] A. Thode and K. Kim, "Multiple-order derivatives of a waveguide acoustic field with respect to sound speed, density, and frequency," J. Acoust. Soc. Am. **116**(6), 3370–3383 (2004).
[2] A. Thode, "The derivative of a waveguide acoustic field with respect to a three-dimensional sound speed perturbation," J. Acoust. Soc. Am. **115**(6), 2824–2833 (2004).
[3] E. K. Skarsoulis and B. D. Cornuelle, "Travel-time sensitivity kernels in ocean acoustic tomography," J. Acoust. Soc. Am. **116**(1), 227–238 (2004).
[4] I. Charpentier and P. Roux, "Mode and wavenumber inversion in shallow water using an adjoint method," J. Comput. Acoust. **12**(4), 521–542 (2004).
[5] P. Hursky, M. B. Porter, W. S. Hodgkiss, and W. A. Kuperman, "Adjoint modeling for acoustic inversion," J. Acoust. Soc. Am. **115**(2), 607–619 (2004).
[6] J.-C. Le Gac, Y. Stéphan, M. Asch, P. Helluy, and J.-P. Hermand, "A variational approach for geoacoustic inversion using adjoint modeling of a PE approximation model with non local impedance boundary conditions," in *Theoretical and Computational Acoustics 2003*, edited by A. Tolstoy, Y. Teng, and E. Shang (World Scientific Publishing, Singapore, 2004), pp. 254–263.
[7] M. Meyer and J.-P. Hermand, "Optimal nonlocal boundary control of the wide-angle parabolic equation for inversion of a waveguide acoustic field," J. Acoust. Soc. Am. **117**(5), 2937–2948 (2005).
[8] C. Sorror and S. Thiria, *YAO User's guide. Version 1.0*, Technical report TR-123, LOCEAN, Paris, France.
[9] R. H. Byrd, P. Lu, and J. Nocedal, "A limited memory algorithm for bound constrained optimization," SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **16**(5), 1190–1208 (1995).
[10] J. Gilbert and C. Lemaréchal, "Some numerical experiments with variable-storage quasi-newton algorithms," Math. Program. **45**, 407–435 (1989).
[11] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," Math. Program. **45**, 503–528 (1989).
[12] G. Madec, P. Delecluse, M. Imbard, and C. Lévy, "OPA 8.1 Ocean General Model reference manual," Technical note 11, LODYC/IPSL, Paris, France.
[13] J. Noilhan and J.-F. Mahfouf, "The ISBA land surface parameterization scheme," Glob. Planet. Change **13**, 145–159 (1996).
[14] A. Boone, J.-C. Calvet, and J. Noilhan, "The inclusion of a third soil layer in a land surface scheme using the force-restore method," J. Appl. Meteorol. **38**, 1611–1630 (1999).
[15] D. Yevick and D. J. Thomson, "Nonlocal boundary conditions for finite-difference parabolic equation solvers," J. Acoust. Soc. Am. **106**(1), 143–150 (1999).
[16] M. Meyer, J.-P. Hermand, M. Asch, and J.-C. Le Gac, "An analytic multiple frequency adjoint-based inversion algorithm for parabolic-type approximations in ocean acoustics," Inverse Probl. Sci. Eng. **14**(3), 245–265 (2006).
[17] O. Talagrand and P. Courtier, "Les équations adjointes - Application à la modélisation numérique," Atelier modélisation de l'atmosphère, Direction de la météorologie, Toulouse, France.
[18] R. Giering and T. Kaminski, "Recipes for adjoint code construction," ACM Trans. Math. Softw. **24**(4), 437–474 (1998).
[19] O. Talagrand, "Application of optimal control to meteorological problems," *Variational Methods in Geosciences*, Vol. **5** of Developments in Geomathematics (Elsevier, New york, 1986) pp. 13–28.
[20] R. M. Errico, "What is an adjoint model?," Bull. Am. Meteorol. Soc. **78**, 2577–2591 (1997).
[21] A more rigorous explanation of the chain rule of elementary calculus

considering a sequence of operations comparable to Eq. (2) is given, e.g., in Ref. 20, [Eqs. (5)–(9) therein].
[22]S. J. Norton, "Iterative inverse scattering algorithms: Methods of computing Fréchet derivatives," J. Acoust. Soc. Am. **106**(5), 2653–2660 (1999).
[23]S. J. Norton, "Iterative algorithms for computing the shape of a hard scattering object: Computing the shape derivative," J. Acoust. Soc. Am. **116**(2), 1002–1008 (2004).
[24]P. R. Gillivray and D. W. Oldenburg, "Methods for calculating Fréchet derivatives and sensitivities for the non-linear inverse problem: A comparative study," Geophys. Prospect. **38**, 499–524 (1990).
[25]A. Tarantola, "Inversion of seismic reflection data in the acoustic approximation," Geophysics **49**), 1259–1266 (1984).
[26]M. B. Giles and N. A. Pierce, "An introduction to the adjoint approach to design," Flow, Turbul. Combust. **65**, 393–415 (2000).
[27]M. Meyer and J.-P. Hermand, "Backpropagation techniques in ocean acoustic inversion: Time reversal, retrogation and adjoint modelling—A review," in *Acoustic Sensing Techniques for the Shallow Water Environment: Inversion Methods and Experiments*, edited by A. Caiti, R. Chapman, J.-P. Hermand, and S. Jesus (Springer, Dordrecht, 2006), ISBN: 1-4020-4372-4 (in press).
[28]J. L. Gross and J. Yellen, *Graph Theory and its Applications, CRC series on Discrete Mathematics and its Applications* (CRC Press, Florida, 2005).
[29]R. Diestel, *Graph Theory, Graduate Texts in Mathematics*, Vol. **173** (Springer, Heidelberg, 2005).
[30]If the index $k \in \mathcal{I}$ belongs to an input variable of the last module (cost function) the derivative reduces to $\frac{\partial \mathcal{L}}{\partial x_k} = \frac{\partial J}{\partial x_k} - \beta_k = 0 \Leftrightarrow \beta_k = \frac{\partial J}{\partial x_k}$.
[31]J. Claerbout, "Coarse grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure," Geophysics **35**), 407–418 (1970).
[32]F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics Press, New York, 1994).
[33]D. J. Thomson, "Wide-angle parabolic equation solutions to two range-dependent bench mark problems," J. Acoust. Soc. Am. **87**(4), 1514–1520 (1990).
[34]D. Thomson and C. Bohun, "A wide angle initial field for parabolic equation models," J. Acoust. Soc. Am. **83**, S118 (1988).
[35]G. H. Brooke, D. J. Thomson, and G. R. Ebbeson, "PECAN: A Canadian parabolic equation model for underwater sound propagation," J. Comput. Acoust. **9**(1), 69–100 (2001).
[36]A similar treatment of Papadakis' original spectral integral formulation of the NLBC (Neumann to Dirichlet map) is proposed in Eqs. (2.20)–(2.23) in Ref. 37 while an improved handling of the singularity of the inverse square-root operator therein was presented later on in Ref. 38.
[37]D. J. Thomson and M. E. Mayfield, "An exact radiation condition for use with the a posteriori PE method," J. Comput. Acoust. **2**(2), 113–132 (1994).
[38]D. J. Thomson, "Notes on the evaluation of the outer integral in PE NLBCs," (unpublished).
[39]W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge Univ. Press, Cambridge, U.K., 1992).
[40]J.-P. Hermand, M. Meyer, M. Asch, M. Berrada, C. Sorror, S. Thiria, F. Badran, and Y. Stéphan, "Semi-automatic adjoint PE modeling for geoacoustic inversion," in *Seventh International Conference on Theoretical and Computational Acoustics*, edited by D. Lee, A. Tolstoy, E. Shang, and Y. Teng (World Scientific Publishing, 2006), Hangzhou, Zhejiang, China, 19–23 September 2005 (to appear).
[41]M. S. Zhdanov, "Geophysical inverse theory and regularization Problems," No. 36 in *Methods in Geochemistry and Geophysics* (Elsevier, Amsterdam, 2002).
[42]M. Hanke and P. C. Hansen, "Regularization methods for large-scale problems," Surv. Math. Ind. **3**, 253–315 (1993).
[43]S. J. Norton, "The inverse-scattering problem and global convergence," J. Acoust. Soc. Am. **118**(3), 1534–1539 (2005).
[44]J.-P. Hermand and P. Gerstoft, "Inversion of broadband multitone acoustic data from the YELLOW SHARK summer experiments," IEEE J. Ocean. Eng. **21**(4), 324–346 (1996).
[45]J.-P. Hermand, "Broad-band geoacoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: Theory and experimental results," IEEE J. Ocean. Eng. **24**(1), 41–66 (1999).
[46]J.-P. Hermand and C. W. Holland, "Geoacoustic characterisation of fine-grained sediments using single and multiple reflection data," Mar. Geophys. Res. **21**, 267–274 (2005).

Spectral fine-structures of low-frequency modulated distortion product otoacoustic emissions

Lin Bian[a)]
Department of Speech and Hearing Science, Arizona State University, 3470 Coor Hall, Tempe, Arizona 85287-0102

(Received 27 January 2006; revised 2 April 2006; accepted 4 April 2006)

Biasing of the cochlear partition with a low-frequency tone can produce an amplitude modulation of distortion product otoacoustic emissions (DPOAEs) in gerbils. In the time domain, odd- versus even-order DPOAEs demonstrated different modulation patterns depending on the bias tone phase. In the frequency domain, multiple sidebands are presented on either side of each DPOAE component. These sidebands were located at harmonic multiples of the biasing frequency from the DPOAE component. For odd-order DPOAEs, sidebands at the even-multiples of the biasing frequency were enhanced, while for even-order DPOAEs, the sidebands at the odd-multiples were elevated. When a modulation in DPOAE magnitude was presented, the magnitudes of the sidebands were enhanced and even greater than the DPOAEs. The amplitudes of these sidebands varied with the levels of the bias tone and two primary tones. The results indicate that the maximal amplitude modulations of DPOAEs occur at a confined bias and primary level space. This can provide a guide for optimal selections of signal conditions for better recordings of low-frequency modulated DPOAEs in future research and applications. Spectral fine-structure and its unique relation to the DPOAE modulation pattern may be useful for direct acquisition of cochlear transducer nonlinearity from a simple spectral analysis. [DOI: 10.1121/1.2200068]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Bt, 43.64.Ld [BLM] Pages: 3872–3885

I. INTRODUCTION

Hearing is initialized in the inner ear via signal transduction, i.e., the transformation of mechanical energy in sound to electrical signal in the cochlear hair cells. There is also a reverse electromechanic transduction serving as a feedback mechanism to enhance hearing sensitivity. Since both forward and reverse transduction processes are highly nonlinear (Hudpeth, 1989; Evans *et al.*, 1991), estimating cochlear nonlinearity is one way to evaluate the functional status and structural integrity of the cochlear transducers: the hair cells and their coupling structures. A consequence of the cochlear nonlinearity is the generation of distortion products (DPs) under two-tone stimulation ($f_1, f_2, f_1 < f_2$). These mechanical DPs are measurable in the ear canal as faint sounds termed distortion product otoacoustic emissions (DPOAEs). Because of the objective and noninvasive nature of DPOAE measurement, and its fundamental relation with the cellular structure of the inner ear, various attempts have been made to correlate DPOAEs and the cochlear functions (Kemp, 2002). These researches have brought about several important clinical applications in diagnostic audiology (Lonsbury-Martin and Martin, 2003).

Recently, a method has been introduced to obtain a cochlear transducer function (F_{tr}) by low-frequency modulation of DPOAEs (Bian *et al.*, 2002, 2004; Bian, 2004). In this method, a high-level low-frequency tone was used to bias the cochlear partition and in turn modulate the magnitudes of DPOAEs. As a result, the DPOAE magnitudes were either enhanced or suppressed depending on the variation of the low-frequency tone amplitude. In principle, the envelopes of the DPOAE waveforms or modulation patterns are related to the absolute values of the appropriate derivatives of the cochlear F_{tr}, and hence the cochlear nonlinearity can be derived. Moreover, for one cycle of the bias tone, there are two modulation patterns forming a hysteresis loop. It has been speculated that the loop can be attributed to a feedback mechanism for an active force production in the outer hair cells which consists of a "negative damping" component (Bian *et al.*, 2004; Bian and Chertoff, 2006). Therefore, low-frequency modulation of DPOAEs could be useful as a clinical tool to estimate the function of the "cochlear amplifier."

Because the previous studies on low-frequency modulation of DPOAEs (Frank and Kössl, 1996, 1997; Scholz *et al.*, 1999; Bian *et al.*, 2002, 2004; Bian, 2004) focused on the most prominent DP, the cubic difference tone (CDT, $2f_1 - f_2$) and quadratic difference tone (QDT, $f_2 - f_1$), little is known about the wide spectrum of other DPs. DPOAEs other than these are useful for getting a more integrated view of cochlear nonlinearity (Cheng, 1999). Their interrelation could reveal the exact nonlinear characteristics of the cochlear transducer. Moreover, studying the temporal behaviors of the DPOAEs and their spectral correlation could provide insight into the mechanisms of the DP generation (Kemp and Brown, 1986; Brown, 1994; Moulin and Kemp, 1996). Simulations of DP generations from nonlinear systems indicate that much more complex interactions among the harmonics of the primary tones and their intermodulations can give rise to the various DPOAEs measured in the ear canal (Fahey *et al.*, 2000). Since a key element of these complicated mechanisms is the cochlear nonlinearity, attempts

a)Electronic mail: lin.bian@asu.edu

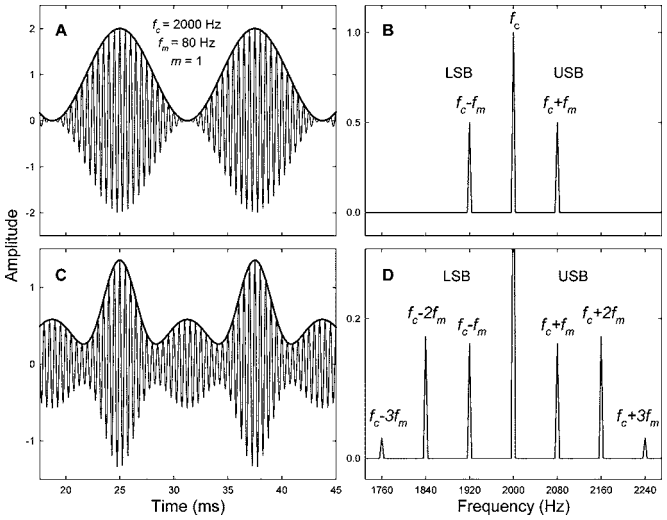


FIG. 1. Amplitude modulation (AM) and the effect of nonlinearity. A waveform of an AM signal (A) and its corresponding spectral representation (B). Modulation depth is 100%. USB: upper sideband; LSB: lower sideband. When such an AM signal is submitted to a saturating nonlinear system, the output spectrum shows multiple sideband components around the carrier (D). The waveform of the carrier in the nonlinear output demonstrates a complex AM (C).

should be made to examine the different behaviors of the odd- and even-order DPOAEs under low-frequency biasing of the cochlear structures. Therefore, the purpose of this study is to explore the low-frequency modulation of all measurable DPOAEs in spectral and temporal domains.

II. THEORETICAL CONSIDERATION

The temporal effect of low-frequency biasing of cochlear partition is a result of modulating cochlear gain by shifting the operating point (OP) along the cochlear F_{tr}. This periodic variation of cochlear gain enhances or suppresses the magnitude of the DPOAEs, i.e., produces an amplitude modulation (AM) of DPOAEs in time domain. There are direct relations between the spectral contents of an AM signal and the temporal patterns of the waveform [Figs. 1(a) and 1(b)]. If a carrier of frequency f_c is amplitude-modulated by a sinusoid with frequency f_m, the time-domain waveform can be represented by

$$S_{\mathrm{AM}}(t) = [1 + m\cos(2\pi f_m t)]\cos(2\pi f_c t), \tag{1}$$

where a positive value of $m(\leqslant 1)$ controls the amount of modulation in carrier amplitude or modulation depth. From Fig. 1(a), it is noticeable that $1+m\cos(2\pi f_m t)$ is the time-varying amplitude or the envelope of the carrier signal. Expanding Eq. (1), one can obtain an alternative expression of the AM signal in terms of three frequency components:

$$S_{\mathrm{AM}}(f) = \cos(2\pi f_c t) + \frac{m}{2}[\cos 2\pi(f_c \pm f_m)t], \tag{2}$$

where $f_c \gg f_m$. In the spectrum (panel B), the AM signal contains two additional components at a frequency difference of f_m on both sides of the carrier (f_c). The magnitude of the upper sideband (USB, f_c+f_m) or the lower sideband (LSB, f_c-f_m) equals 1/2 of that of the modulator.

For a nonlinear system, the spectral and temporal behavior of an AM signal could be more complex. With the AM signal in Eq. (1) as input, the nonlinear output can be expanded with a Taylor's series, $F_{\mathrm{NL}}=\Sigma_{n=0}^{\infty} a_n S_{\mathrm{AM}}^n(t)$, where a_n is a coefficient related to the nth derivative of F_{NL}, i.e., $F_{\mathrm{NL}}^{(n)}/n!$. Applying the binomial theorem, the nth term of the power series is

$$S_{\mathrm{AM}}^n(t) = a_n \sum_{k=0}^{n} \left[m^k \binom{n}{k} \cos^k(2\pi f_m t) \right] \cos^n(2\pi f_c t). \tag{3}$$

In the spectral domain, generally $\cos^n\theta = \Sigma_{i=1}^{n} B_i \cos(n\theta)$, where $B_i = \binom{n}{i}/2^{n-1}$ and $\binom{n}{i}$ is a binomial coefficient. The frequency components relating only the carrier f_c can be determined

$$S_{\mathrm{AM}}^n(f_c) = a_n \sum_{k=0}^{n} \left[m^k \binom{n}{k} \sum_{j=1}^{k} B_j \cos 2\pi(k f_m)t \right] \cos(2\pi f_c t), \tag{4}$$

where $B_j = \binom{k}{j}/2^{k-1}$. When multiplied through, Eq. (4) can be modified as

$$S_{\mathrm{AM}}^n(f_c) = a_n \sum_{k=0}^{n} \left[m^k \binom{n}{k} \sum_{j=1}^{k} \frac{B_j}{2} \cos 2\pi(f_c \pm k f_m)t \right]. \tag{5}$$

Therefore, the carrier-related spectral contents of the nonlinear output consist of sidebands of $f_c \pm k f_m$, i.e., the sidebands components are at integer multiples of the f_m from the carrier [Fig. 1(d)]. The waveform of the nonlinear AM signal shows complex patterns that are determined by the transfer characteristics of the system [Fig. 1(c)]. Thus, a simple spectral analysis of the amplitude modulated DPOAEs may help to identify the temporal modulation pattern. This interrelation between the structure of the spectral sidebands and their temporal modulation patterns could be useful in deriving cochlear F_{tr}.

III. METHODS

Data were collected in the Auditory Processing Laboratory at the University of Kansas Medical Center (KUMC) from eight Mongolian gerbils anesthetized with ketamine and xylazine (100 and 2 mg/kg, i.p.). The experimental procedures were approved by the Institutional Animal Care and Use Committee of the KUMC and described previously (Bian *et al.*, 2004). Briefly, a plastic speculum was sealed onto the bony ear canal after removing of the right pinna. A probe microphone (Etymotic Research, ER-10B) and two coupled earphones (ER-3A), was inserted into the speculum. The bias tone was wave-guided from a subwoofer (Boston Acoustics) via a silicon tube to a pore at the tip of the speculum. Two primary tones and a low-frequency bias tone were presented simultaneously and the ear-canal acoustics were recorded. To ensure stability and reliability, prior to each recording the CDT magnitude was verified to be no less than 40 dB SPL when the primaries were presented at 80 dB SPL.

The bias tone and the two primaries were presented separately with their intensity levels controlled by three attenuators (Tucker-Davis Tech., TDT PA4). Two pairs of pri-

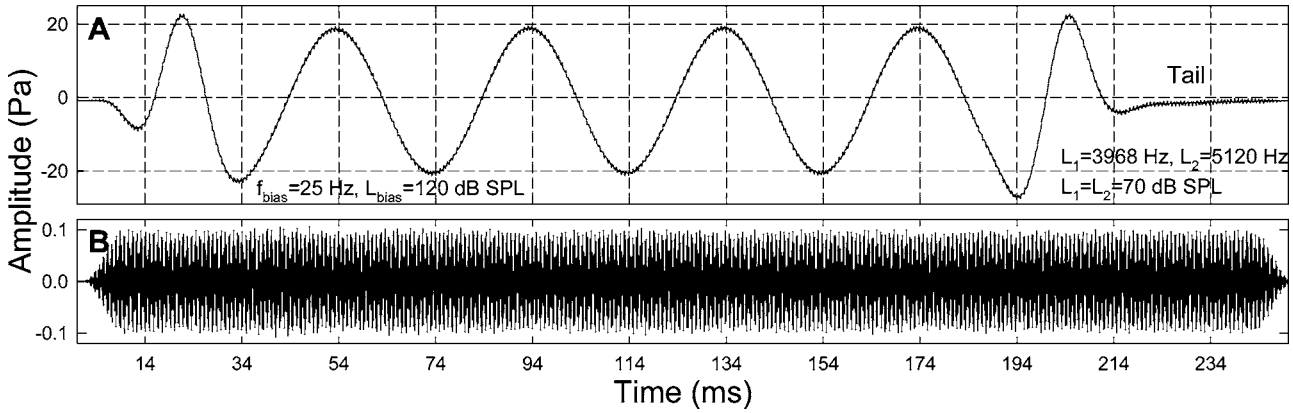


FIG. 2. Ear canal acoustical signal and analysis. (A) The bias tone at 120 dB pSPL and two tone signal at 70 dB SPL. The tail portion (about 40 ms) of the signal is flat with no biasing. (B) The acoustical signal in (A) is high-pass filtered at 200 Hz to eliminate the bias tone. Note: there is no modulation in the amplitude of the high-passed signal.

mary frequencies were used, i.e., 3968/5120 Hz, (f_2/f_1 =1.29) and 5888/7552 Hz, (f_2/f_1=1.28). The primary levels were varied from 50 to 75 dB SPL in 5-dB steps (L_1=L_2). A bias tone of 25 Hz (f_{bias}) was amplified (Optimus, SA-155) and calibrated at 20 Pa peak amplitude (120 dB pSPL). A 12-dB correction was made to compensate for the frequency response of the ER-10B microphone at 25 Hz. The bias tone was 250 ms in duration with a flat tail of about 40 ms serving as a reference for no biasing condition [Fig. 2(a)]. The peak amplitude of the bias tone was descended automatically in 21 steps from 20 to 0 Pa using a quadratic scaling so that the level was attenuated in more steps at lower intensities. The recorded signals were amplified 29 dB (Stewart VBF 10M), averaged up to 100 times, sampled at 65536 Hz (TDT AD2).

Spectral analysis and temporal DPOAEs extraction routines were performed in MATLAB (v. 6.1 MathWorks). The full-length acoustical signals were high-pass filtered at 200 Hz [Fig. 2(b)] to eliminate the bias tone and submitted to a fast Fourier transformation (FFT) to obtain the frequency spectra. All DPOAE components 5 dB above the whole-spectral noise floor were identified from the spectrum using an algorithm based on the formulae described by Bian (2004). Explicitly, for two-tone input of frequencies θ_1 and θ_2, the DP frequencies are determined by recursively solving $m_h\theta_1 \pm m_k\theta_2$, where $m_h=h-2i$, $m_k=k-2j$, $h=n-k$, and i,j are integers below $h/2$ and $k/2$, respectively. Whether a particular DP is odd—or even-order is determined by the sum of m_h and m_k. Once a DP component was identified, its spectral fine structure in a vicinity of ±200 Hz was examined and magnitudes of the USBs and LSBs were measured. Two types of sideband magnitudes were obtained, namely, the absolute amplitude in dB SPL and the relative amplitude in dB with the DP as the reference. Temporal features of the DP were extracted using a 300-Hz-wide band-pass filter (zero-phase Butterworth filter) centered at the DP or a 512-point moving-window FFT method (Bian *et al.*, 2004; Bian, 2004). Two different methods were used to validate that the results were free from signal processing artifact.

To examine the effects of varying bias levels and the primary levels, the absolute and relative magnitudes measured from the USBs and LSBs of the DPs were submitted to a three-way (sideband type × primary level × bias level) analysis of variance (ANOVA) in STATISTICA (StatSoft). To evaluate the effects of the primary levels, the sideband magnitudes obtained at the highest bias tone level, 20 Pa, were examined together for the two pairs of primary frequencies with a three-way (sideband type × primary frequency × primary level) ANOVA. The highest bias tone level was used because the modulation of DPOAEs was significant for most of the primary levels. A p-value of 0.01 was adopted as a criterion for statistical significance.

IV. RESULTS

A. Whole spectral contents

Frequency analysis of the acoustical signal and identification of DP components revealed that a whole spectrum of various intermodulation or harmonic DPs were generated (Fig. 3). For frequency region below f_2, a notable finding was that a series of related DP components presented. These DPs included QDT, its second- [$2(f_2-f_1)$ or 2QDT], and third-order harmonics [$3(f_2-f_1)$ 3QDT] with descending magnitudes. A fourth harmonic of the QDT [$4(f_2-f_1)$ 4CDT] was often found between the higher primary frequency (f_1 =5888 Hz) components [Fig. 3(b)]. Within the same frequency range ($<f_2$), odd-order DPs were CDT and $3f_1-2f_2$. The magnitude of CDT was the largest and that of the QDT came next. For frequencies higher than f_2, a third-order DP, $2f_2-f_1$, was the largest and the summation tone (f_1+f_2) the second. An even-order DP $3f_1-f_2$ was often observed along with the harmonics of the primaries, such as $2f_1$, $3f_1$, and $2f_2$. For lower frequencies (f_1=3968 Hz), there were more additional DP components [Fig. 3(a)], e.g., a fifth-order DP ($3f_2-2f_1$) and the second harmonic of CDT [$2(2f_1-f_2)$ or 2CDT].

Most of the identifiable intermodulation or harmonic DPs shared some common features. The amplitudes of these components were at least 10 dB above the averaged noise floor across the whole spectra which was from approximately −15 to −20 dB SPL. Moreover, there was a substantial amount of energy around the DP frequency component,

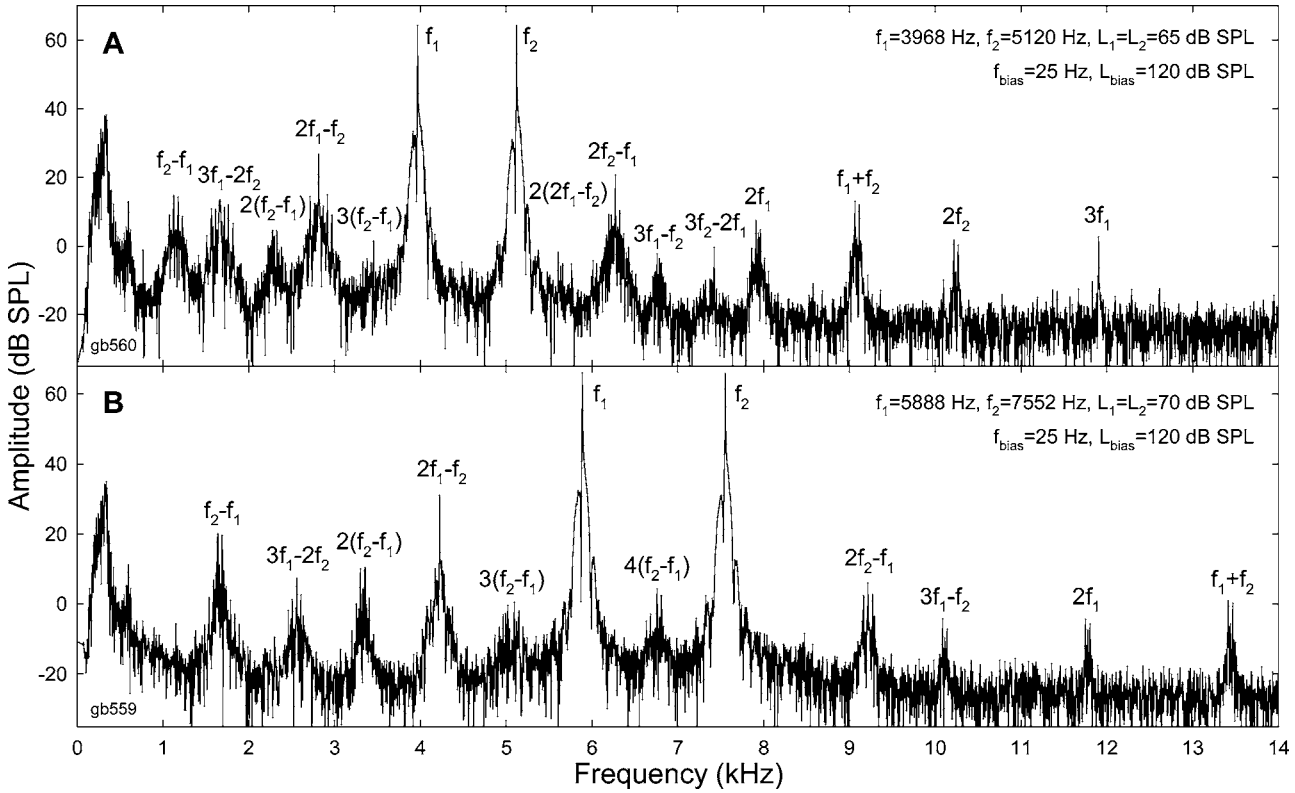


FIG. 3. Spectra of the ear-canal acoustic signals for different primary frequencies. (A) $f_1=3968$ Hz; (B) $f_1=5888$ Hz. The original time waveforms were high-pass filtered at 200 Hz. Compositions of the harmonic and DPs are labeled. Note: the multiple sideband peaks and the elevated baseline around each DP component.

forming a noticeable spectral lobe underneath the DP and its sideband complex (Fig. 3). The width of the lobe seemed to be proportional to the amplitude of the DP. The most important spectral feature was the complicated fine-structures around the DP components.

B. Temporal modulation patterns and spectral fine structure

1. Temporal modulation patterns

The presence of the typical modulation patterns depended on the magnitude of the DP, the type of the particular DP component, and the power of the bias tone. When the DP magnitude without bias tone was below 0–5 dB SPL, it tended to be very noisy and show hardly any typical modulation pattern. For most DPs, the general observation was that odd- and even-order DPs showed different modulation patterns (Fig. 4). Comparing to the peaks and troughs of the bias tone indicated by the vertical grid lines in Fig. 2(a), one can relate the DP modulation patterns with the instantaneous bias tone pressures. For odd-order DPs (upper panels of Figs. 4 and 5), the effect of the bias tone was a phase-dependent suppression and modulation of the DP magnitude compared to the tail portions of the waveforms where the DP was relatively large and stable. A peak of DP magnitude presented between two adjacent biasing extrema, and deep notches occurred immediately before and after the peak forming a spindle-shaped center lobe. At the bias tone peaks or troughs, smaller lobes presented that were often asymmetrical, i.e., they were larger and wider at the troughs or biasing in the negative sound pressure direction. This modulation pattern showing one center lobe and two side-lobes generally was similar to the shapes of the absolute odd-order derivative of a sigmoid-shaped function. For $3f_1-2f_2$ as an example, the lobes were narrower and consistent with the absolute value of the fifth derivative of the cochlear F_{tr} (Lukashkin and Russell, 2005; Bian and Chertoff, 2006).

For even-order DPs, the bias tone produced both enhancement and modulation of the DP compared to the tail portions of the waveforms (lower panels of Figs. 4 and 5). Enhancements occurred at the biasing peaks or troughs and notches presented slightly behind the zero-crossings (center between vertical grids) due to delays of cochlear responses. The enhancement of the DPs was asymmetrical so that the lobes corresponding to biasing toward the positive sound pressures are larger. The summation tone appeared to be enhanced in the opposite biasing direction. At the lower primary frequencies ($f_1=3968$ Hz), the enlarged magnitudes of QDT and 2QDT could not sustain and slipped when biasing pressure further increased in both directions, thus showing double-peaked lobes, especially during biasing in the positive direction (lower panel of Fig. 4).

Exceptions were observed for a third-order DP, $2f_2-f_1$, whose waveform was very similar to those of the even-order DPs (lower panel of Fig. 4). Moreover, the DP magnitude was very small in the tail portion. Such a behavior was observed in several animals under various signal conditions, especially at the lower primary frequencies ($f_1=3968$ Hz). Another observation was related to the reduced modulation ability of the bias tone when increasing the primary level and

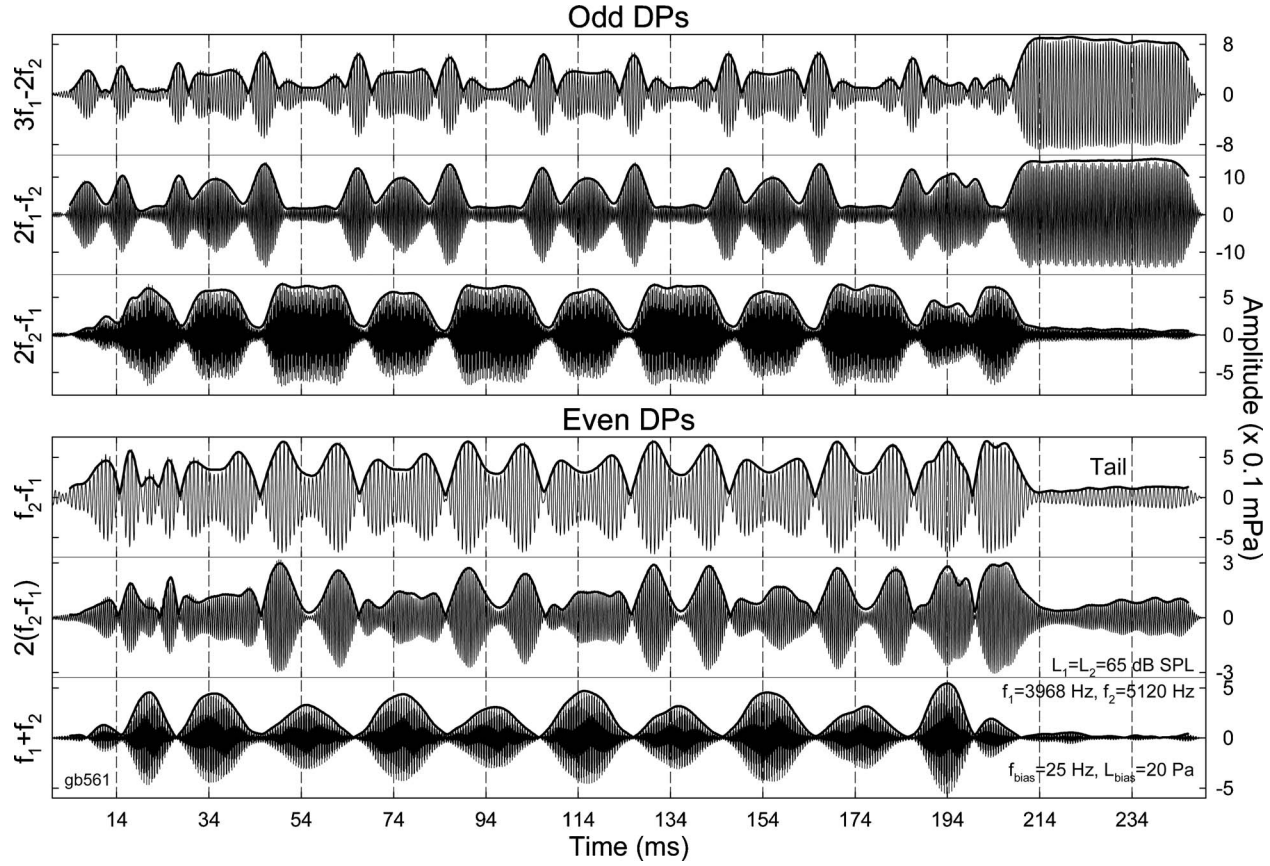


FIG. 4. Temporal modulation patterns at the lower primary frequencies (f_1=3968 Hz). Band-passed waveforms of the odd- and even-order DPs are displayed in the upper and lower panels, respectively. The bandwidth of the filters (zero-phase Butterworth) is 300-Hz centered at the DP frequencies. Envelopes of the waveforms are derived from the moving-window method. The DPs are labeled in each panel. Vertical grid lines indicate the peaks and troughs of the bias tone (see Fig. 2 for a reference) where magnitudes of the odd-order DPs are suppressed and even DPs are enhanced. Refer to the tail portions for DPOAE magnitudes without biasing. Note: the modulation pattern of $2f_2-f_1$ is similar to even DPs.

frequencies. At higher primary levels (⩾70 dB SPL), the CDT could be strong enough that the bias tone could only partially modulate the DP magnitude (upper panel of Fig. 5), compared to the typical modulation pattern at 65 dB SPL primary level.

2. Spectral fine-structures

Spectral features of the low-frequency modulation of DPOAE can be characterized by the fine-structures around the DP components (Figs. 6 and 7). Just as predicted by Eq. (5), there were multiple sidebands presented at multiples of the f_{bias} (25 Hz) from the DP with reducing amplitudes. The USBs are generally similar to the LSBs, but they could be asymmetrical. However, only every other sideband components that were 50 Hz or $2f_{\text{bias}}$ apart from each other were prominent for both odd- and even-order DPs. These enhanced sideband components might correlate with the temporal modulation patterns of the DPs, since the locations of the prominent components for the odd- and even-DPs were different. Three sideband components were identified for the DPs. For odd-DPs, the first sideband peak, termed sideband I, was 50 Hz ($2f_{\text{bias}}$) from the DP components on either side (left-hand panels of Figs. 6 and 7). The second sideband peaks, sideband II, were 100 Hz or $4f_{\text{bias}}$ from the DP. In contrast, for even-order DPs, the first sideband peak or sideband I, was 25 Hz (f_{bias}) from the DP component and sideband II was 75 Hz or $3f_{\text{bias}}$ from the DP (right-hand panels of Figs. 6 and 7). The third components, sideband III, were 50 Hz further away from sideband II on both sides. Because sideband III was relatively smaller and more variable, the analysis was focused on sidebands I and II.

Generally, magnitudes of the odd-order DPs were greater than that of the sidebands (left-hand panels of Figs. 6 and 7). Usually, magnitude of sideband I was greater than sideband II, but that was not always the case for CDT and $3f_1-2f_2$ when f_1 was 3968 Hz (Fig. 6). The highest sideband magnitude could be well above 10 dB SPL. The size of the sidebands relative to DP magnitude could indicate the presence or absence of the typical modulation pattern. At the primary level of 70 dB SPL for an example, the CDT was slightly modulated with no typical modulation pattern in contrast to the CDT at 65 dB SPL (upper panel of Fig. 5). The sidebands at 70 dB SPL were about 20 dB below the CDT magnitude as oppose to nearly equal at 65 dB SPL (top left-hand panel of Fig. 7). It was observed that when the magnitude difference between the sideband and the DP was within approximately 10 dB, there was a presence of the temporal modulation pattern.

A surprising finding was that for even-order DPs the sideband magnitudes were usually greater than the size of the DP (right-hand panels of Figs. 6 and 7). The largest sideband component, sideband I, could be more than 10 dB above the

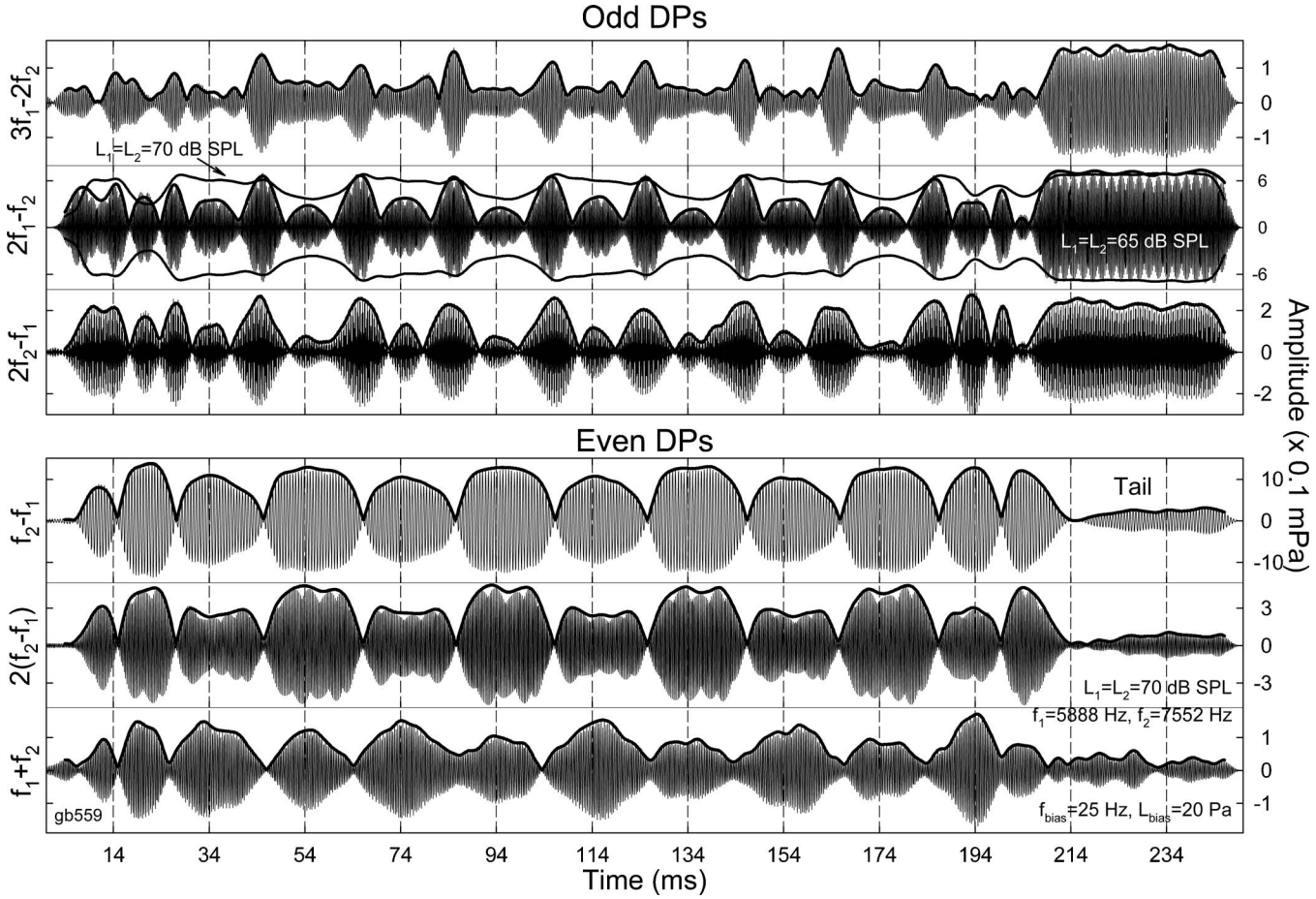


FIG. 5. Temporal modulation patterns at the higher primary frequencies (f_1=5888 Hz). Band-passed waveforms and envelopes of the odd- and even-order DPs are displayed in the upper and lower panels, respectively. Vertical grid lines indicate the peaks and troughs of the bias tone. The temporal modulation patterns are fundamentally different for odd and even DPs. Signal conditions are indicated in the lower panel. Note: (1) the CDT ($2f_1-f_2$) waveform is from a lower primary level (65 dB SPL) and the envelope of this DP at 70 dB SPL primary level is outlined with a pair of solid lines; (2) the CDT at higher primary level is partially modulated.

DP magnitude. These sidebands could be a marker of the presence of an even-DP. As in the case of summation tone when f_1 was 5888 Hz (bottom right-hand panel of Fig. 7), the magnitude of f_1+f_2 was approximately −5 dB SPL, and the sideband magnitudes were above 1 dB SPL. This made the otherwise difficult-to-detect DP more easily to be identified and measured, i.e., a 6-dB enhancement. The spectral enhancement was greater at the lower primary frequencies ($F1$=3968 Hz), and could be more than 10 dB for QDT and summation tones (right-hand panels of Fig. 6).

C. Contours of the modulation sidebands

A three-way (sideband type × primary level × bias level) ANOVA showed that there was no significant difference between the magnitudes of USBs and LSBs. Therefore, the magnitudes of the USBs and LSBs were pooled and averaged to focus on the level effects. Dependence of the sideband magnitudes on both the primary and bias tone levels was examined with a contour map constructed by plotting the averaged sideband magnitudes across animals against these signal levels (upper portions of panels in Figs. 8 and 9). Since the DP modulation can be better described by sideband enhancements, contours of the sideband magnitudes relative to the corresponding DPs were also plotted for a comparison (lower portions of panels in Figs. 8 and 9). Because the CDT and QDT are among the largest DPs with small variability, sidebands of these DPs were selected to represent the odd- and even-DPs, respectively.

1. Sidebands of CDT

Generally, the absolute sideband magnitudes remained relatively low (<0 dB SPL) when bias level was below 4 Pa or 106 dB SPL (upper portions of panels in Fig. 8). In this bias-level range, increasing the primary level elevated the sideband up to approximately 15 dB. Beyond these low bias levels, the major effect was the increase in sideband magnitudes with both bias and primary levels. The sideband magnitudes reached their peak values (5–10 dB SPL) when the bias level was above 12 Pa or 116 dB SPL and the primary level higher than 55–60 dB SPL. For the higher primary frequency (f_1=5888 Hz), further increase in the primary level did not result in greater sideband amplitudes. There was a significant interaction between the primary and bias tone levels for sidebands I and II. It can be observed from Fig. 8

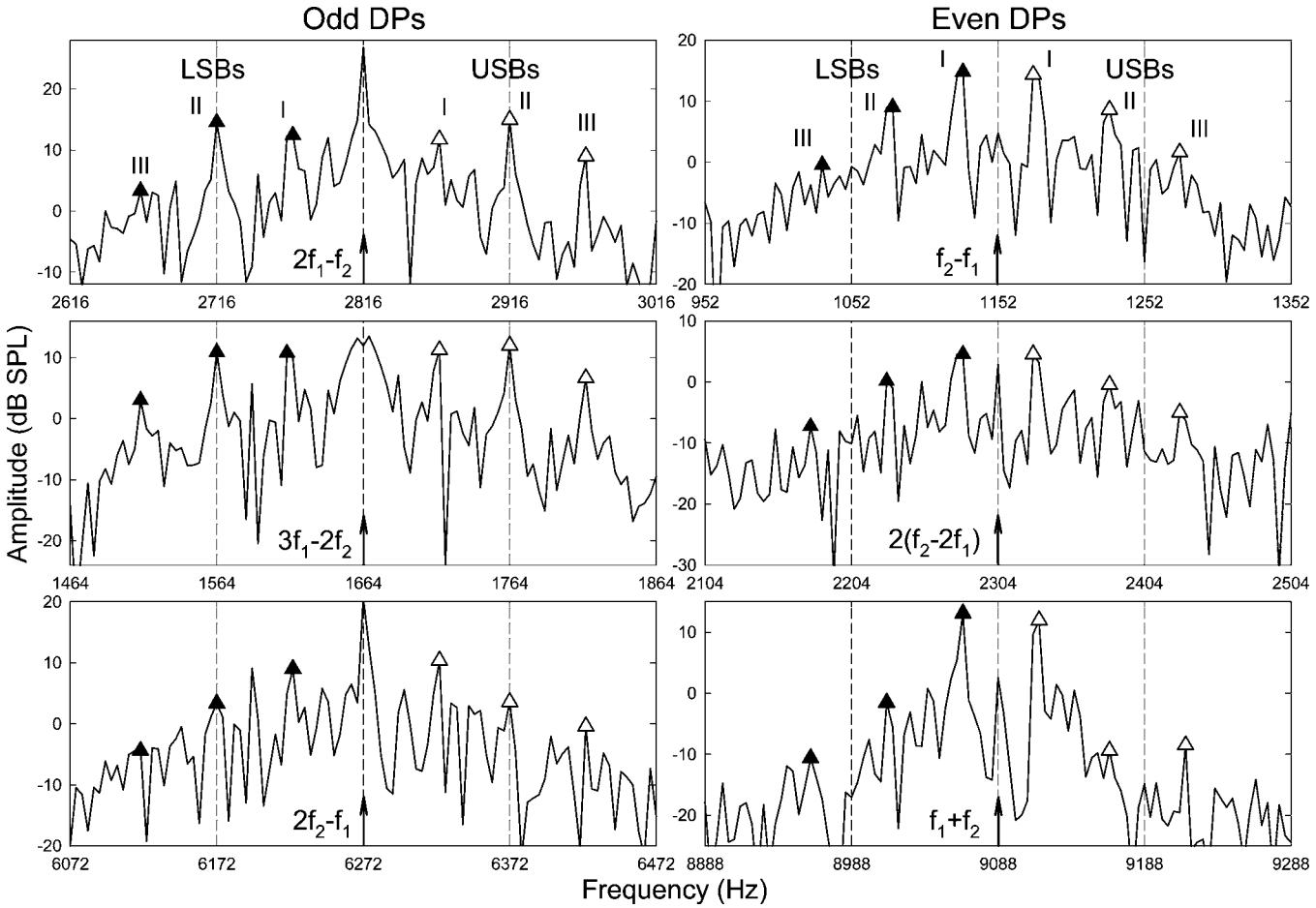


FIG. 6. Spectral fine-structures of low-frequency modulated DPOAEs (f_1=3968 Hz). Left column: odd-order DPs; right column: even DPs. Labels of the DPs are indicated in the center of each panel. Enhanced components of the LSB (closed symbols) and USB (open symbols) are labeled with Roman numerals according to the spectral distance from the DP. There is a 50-Hz interval between adjacent sidebands with labels. Sideband I is 25 Hz (f_{bias}) from the even-DP components and 50 Hz ($2f_{bias}$) from odd DPs. USBs and LSBs are generally symmetrical. Note: there are also smaller sideband components between the labeled ones, especially for $2f_2-f_1$. Time waveforms of the DPs are displayed in Fig. 4.

that the growths of the absolute sideband magnitudes with the bias tone are steeper at lower primary levels than higher levels.

The sideband magnitudes relative to the CDT showed significant effects of the primary and bias tone levels. As shown in the lower portions of panels in Fig. 8, the sidebands remain well below (from −25 to −30 dB) the CDT magnitude when bias level is lower than 4 Pa. The interaction between the primary and the bias tone levels can be more easily observed, i.e., the effect of bias level is greater for low primary levels than higher levels. For primary levels below 60 dB SPL, the relative sideband magnitudes rose sharply to reach a plateau from approximately −10 to −5 dB as bias level passed 8–12 Pa (112–116 dB SPL). For higher primary levels (>65 dB SPL), increase in bias level resulted in very limited growth in the sideband magnitudes. This effect indicated that the maximal modulation occurred when the primary levels were below 60 dB SPL and the bias level was above 116 dB SPL or 12 Pa.

2. Sidebands of QDT

Increasing both the primary and bias tone level can maximize the absolute magnitudes of QDT sidebands (upper portions of the panels in Fig. 9). It is worth noting that sideband I could be maximized when the bias level was as low as 4 Pa or 106 dB SPL [Fig. 9(a) top] for the lower primary frequencies (f_1=3968 Hz). Moreover, this maximal sideband magnitude was confined between the primary levels of 60 and 70 dB SPL. Further increase in the primary level did not produce higher sideband I magnitude. For the higher primary frequencies (f_1=5888 Hz), the primary level should be at least 65 dB SPL to generate a sideband I that is greater than 10 dB SPL (panel C top). To maximize sideband II, however, the bias level should be greater then 10 Pa or 114 dB SPL (top portions of panels B and D) and the primary level at least 60 dB SPL.

The relative magnitudes of QDT sidebands showed that the maximal modulation of QDT occurred at some clearly defined signal spaces. For sideband I at f_1 of 3968 Hz [lower portion of Fig. 9(a)], the maximal relative magnitude of sideband I was restricted between a narrow primary level range between 60 and 70 dB SPL and the bias level above 8 Pa (112 dB SPL). At the higher frequencies (f_1=5888 Hz), the sideband I was maximized at primary levels below 60 dB SPL and bias levels as low as just 5 Pa or 108 dB SPL (lower of panel C). Any additional increase in either the pri-

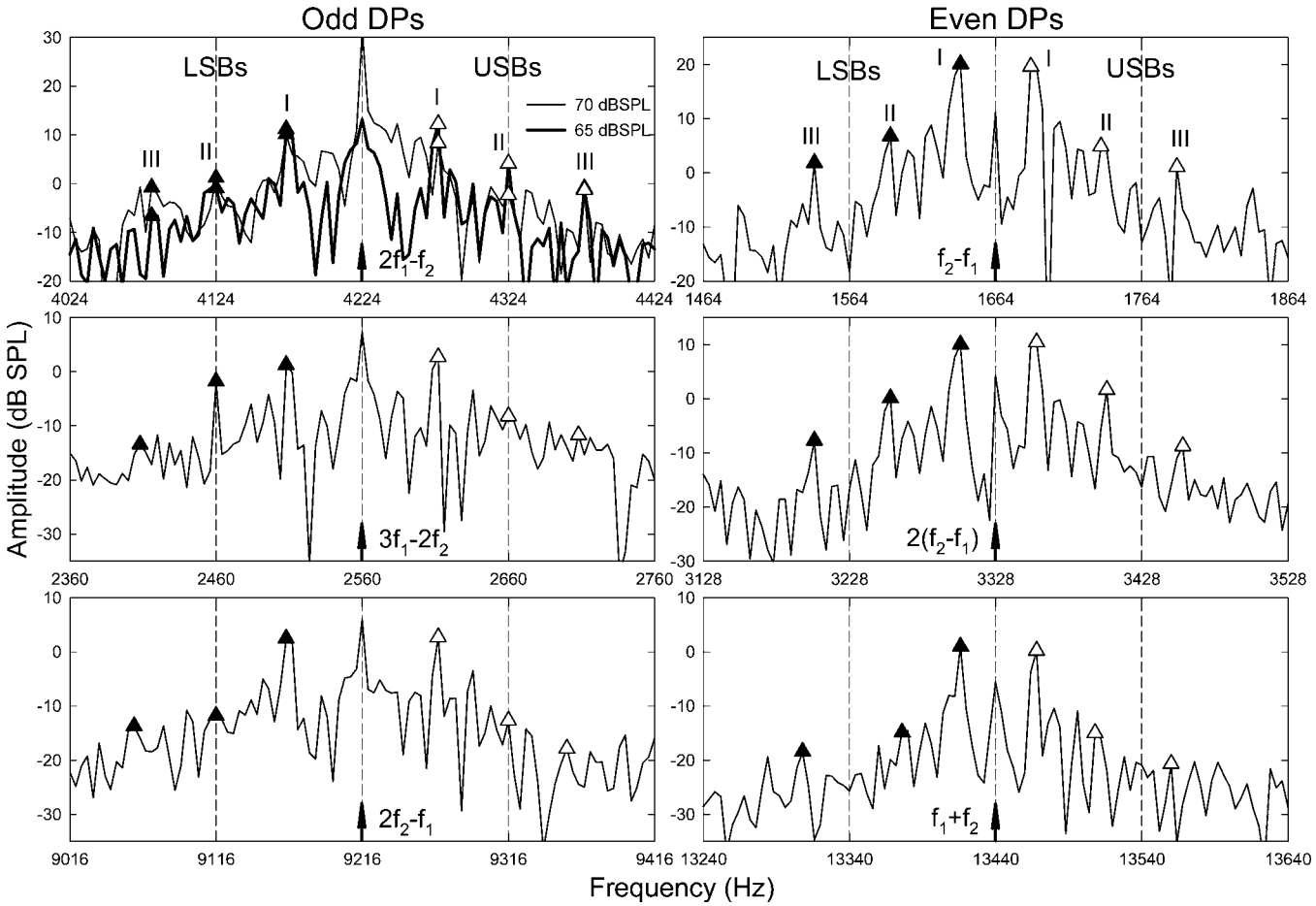


FIG. 7. Spectral fine-structures of low-frequency modulated DPOAEs (f_1=5888 Hz). Left column: odd-order DPs; right column: even DPs. Sideband components present at multiples of f_{bias} (25 Hz) from the DP (center of each panel). Enhanced sideband components (labeled) present at odd multiples of f_{bias} away from the even-order DPs and even multiples of f_{bias} for odd DPs. The spectral interval between adjacent labeled sidebands is 50 Hz. Time waveforms of the DPs and signal conditions are displayed in Fig. 5. For CDT, spectra from two primary levels are shown. Note: the greater sideband magnitudes relative to CDT at the lower primary level (65 dB SPL, thick line).

mary or bias level would result in lower relative sideband I magnitude or a weakened modulation. For sideband II, the relative amplitude was maximized at slightly lower primary levels, i.e., from 50 to 60 dB SPL at the lower primary frequencies (lower of panel B) and below about 52 dB SPL at the higher frequencies (lower of panel D). The lower limit of bias level for producing the maximal relative sideband II magnitude was just below 8 Pa.

D. Growths of modulation sidebands

Because the highest bias tone level (20 Pa) can produce AM of DPs at most primary levels, the growth properties of the DP sidebands with the primary levels were studied. Statistical ANOVA on sidebands I and II magnitudes, showed no difference between USBs and LSBs. The major effects were the influences of primary level and primary frequencies. The absolute sideband amplitudes of CDT showed a saturating growth with primary level [Figs. 10(a) and 10(b)]. The saturation started at approximately 60–65 dB SPL primary level after reaching a peak of approximately 10–13 dB SPL. There was a significant interaction between the primary frequency and primary levels. The amplitudes of sideband I and II for the higher primary frequency (f_1=5888 Hz) reached their maxima at a lower primary level (60 dB SPL), and declined beyond this level. At the lower frequencies, the sidebands showed a wider growth range (~70 dB SPL).

However, the relative amplitudes of the sidebands demonstrated an opposite trend, a declination with primary level [Figs. 10(c) and 10(d)]. The relative sideband amplitudes were among the highest below about 60 dB SPL, and reduced significantly above this level. This effect was greater for the higher primary frequencies (f_1=5888 Hz), where there was a more than 25 dB drop in the relative amplitude with the primary level increased from 60 to 75 dB SPL. The changes in the relative amplitudes were relatively smaller (<20 dB) for the lower primary frequencies (f_1=3968 Hz). In other words, the effect of primary level on the amount of modulation was frequency dependent. For a given bias tone level, increasing the level of primaries at higher frequency had more influence on the modulation patterns than at lower frequency, or the bias tone was weaker to modulate high-frequency DPs.

The absolute sideband amplitudes of QDT also grew significantly with the primary levels [Figs. 11(a) and 11(b)]. Sideband I showed a larger increase (about 30 dB) over a 25 dB rise in primary level. Growth of the absolute sideband I amplitude for the lower primary frequency (f_1=3968 Hz)

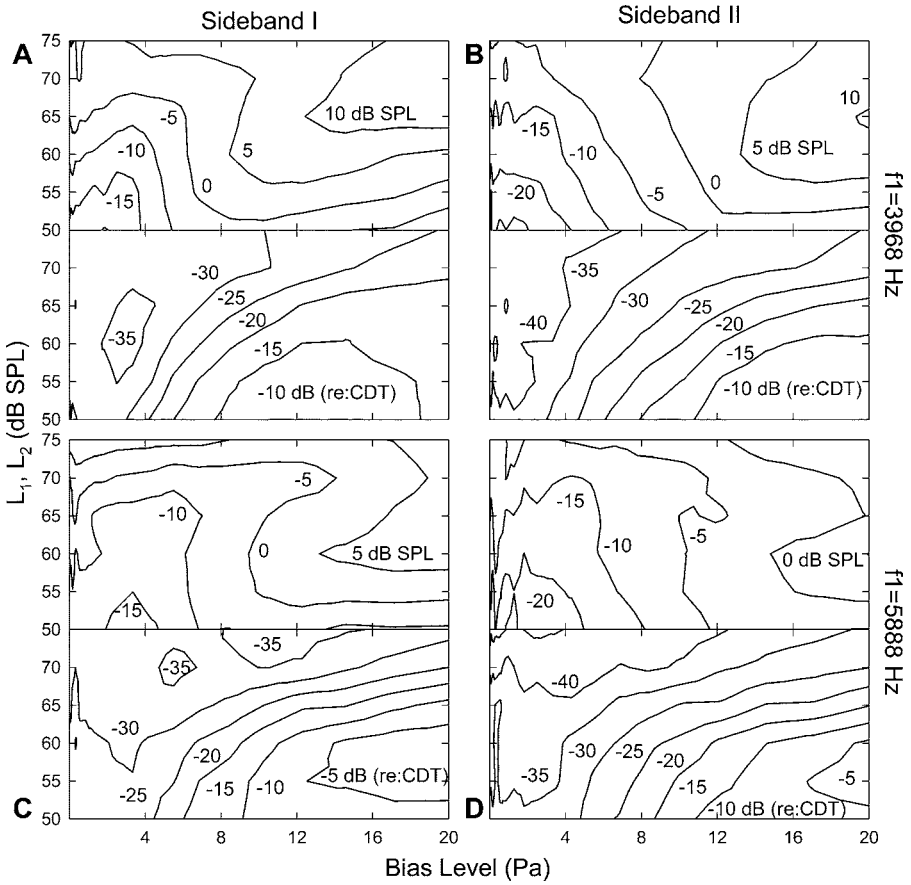


FIG. 8. Modulation contours of CDT. Absolute sideband magnitudes and their relative amplitude with reference to the CDT are plotted as functions of bias tone and primary levels in the upper and lower portions of each panel, respectively. Data represent an average of the sideband I [(A) and (C)] and II [(B) and (D)] across USB and LSB. Elevations of the contours are indicated. The contour lines reflect the mean sideband amplitude ($n=16$). Effective signal conditions can be observed from the contours of the relative sideband amplitudes (lower portion of each panel). Data from the lower primary frequencies are displayed in the upper [(A) and (B)] panels and higher primary frequencies in the lower panels [(C) and (D)], respectively.

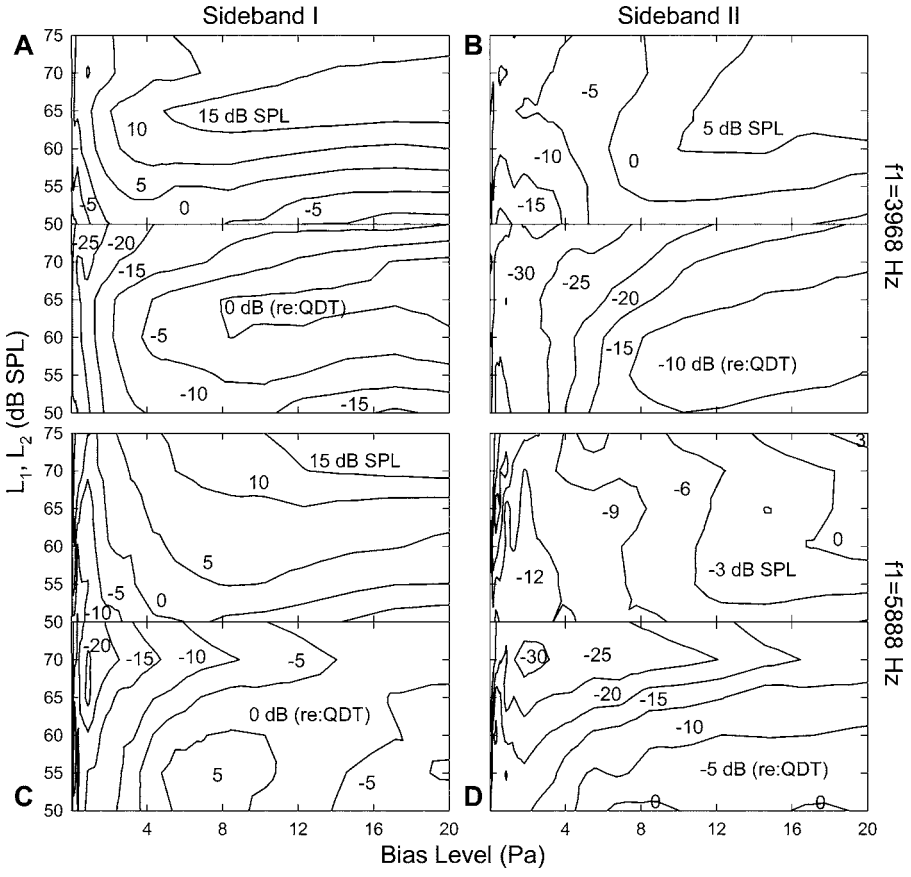


FIG. 9. Modulation contours of QDT. Absolute sideband magnitudes and their amplitude relative to QDT are plotted in the upper and lower portions of each panel, respectively. Data represent the mean of the sideband I [(A) and (C)] and II [(B) and (D)] magnitudes across USB, LSB, and all animals ($n=16$). Sideband amplitude from different primary frequencies are displayed in the upper [(A) and (B)] and lower panels [(C) and (D)], respectively.

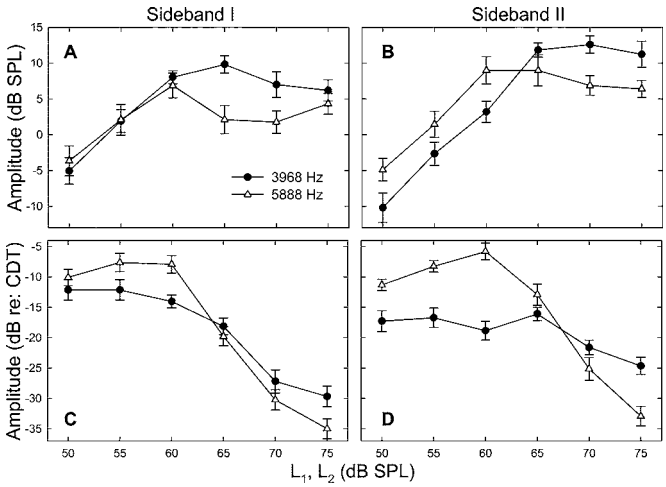


FIG. 10. Growths of CDT sideband amplitudes. (A) and (B) Growth functions of the absolute sideband amplitudes of the CDT as a function of the primary level. (C) and (D) Growths of the sideband magnitudes relative to the CDT. Data of sideband I averaged across USB and LSB are displayed in (A) and (C), averaged sideband II magnitude in (B) and (D). Data reflect the mean±standard error (SE) from all animals (n=16). Bias tone level is 120 dB pSPL or 20 Pa.

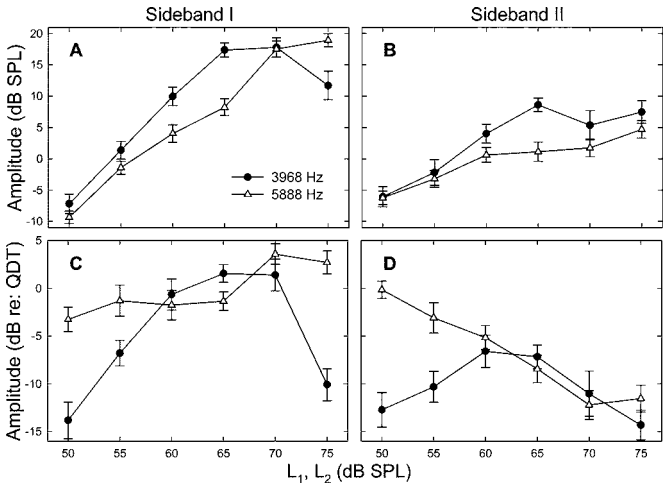


FIG. 11. Growths of QDT sideband amplitudes. Growth functions of the absolute amplitudes of the QDT sidebands are shown in (A) and (B), those of the relative sideband magnitudes in (C) and (D). Average amplitudes of sideband I and II across USB and LSB are displayed in the left and right panels, respectively. Data represent the mean±SE (n=16). Bias tone level is 20 Pa.

reached a maximum at 70 dB SPL and declined with further increase in the primary level. At the higher frequencies (f_1 =5888 Hz), the rise of sideband I amplitude was slower and showed no saturation. The absolute amplitude of sideband II (panel B) was smaller compared to sideband I. Increase in the amplitude of sideband II was limited within 20 dB, and the largest magnitude was about 10 dB SPL for both pairs of primary frequencies.

The relative amplitudes of the QDT sidebands for the two primary frequencies showed some mixed effects [Figs. 11(c) and 11(d)]. At the lower the frequencies (f_1 =3968 Hz), the relative amplitudes of sidebands I and II reached maximal around 65 dB SPL and declined thereafter. The increase of relative amplitudes was more than 15 dB for sideband I, and less than 10 dB for sideband II. However, at the higher frequencies (f_1=5888 Hz), the results were different for different sidebands. The relative sideband I magnitude increased about 10 dB, while sideband II amplitude dropped more than 10 dB within the primary level range from 50 to 70 dB SPL. Thus, the optimal primary level for modulation of QDT at the higher primary frequencies depended on the sideband components. As the sideband I magnitude was larger than sideband II, greater modulation of QDT was more likely to occurr at approximately 65–70 dB SPL.

V. DISCUSSION

A. Spectrum of DPOAEs and the notion of sidebands

A unique approach of the present study is to reveal and identify a whole spectrum of various DPOAEs. The most prominent DP is CDT, followed by QDT. Other DPs also show relatively strong amplitudes, such as, $3f_1-2f_2$, f_1+f_2, $3f_2-2f_1$, $2f_2-f_1$, etc. A considerable number of DP elements can be attributed to the harmonics of these prominent DPs, e.g., 2QDT, 3QDT, 4QDT, and 2CDT. These distortion or harmonic components are more or less modulated by the bias tone, indicated by the presence of multiple sidebands. In a recent report, Lukashkin and Russell (2005) observed that DP components corresponding to the second harmonics of QDT $[2(f_2-f_1)]$ and CDT $[2(2f_1-f_2)]$ were amplitude modulated by the bias tone.

Early studies devoted to the wide spectral presentation of DPOAEs (Kemp and Brown, 1986; Brown, 1986, 1994) showed that when the primary frequencies are closely spaced, the response spectrum reveals a wide spread of many DP components on both sides of the primaries (also termed USBs and LSBs). These DPs are considered to result from the interference of the beating of the two-tone signal, thus related to the QDT, namely, $f_1-n(f_2-f_1)$ for LSBs and $f_2+n(f_2-f_1)$ for USBs (n is an integer), respectively. The contents of these sideband DPOAEs (Brown, 1994) and intracochlear pressures (Dong and Olson, 2005) are comparable to the DP components found in the present study (Fig. 3), especially, the harmonics of QDT. Since the two-tone signal has a fluctuating envelope, it was proposed by Brown (1994) and recently formulated by van der Heijden (2005) that the cochlear transducer can adjust its gain based on the instantaneous envelope of the primaries, i.e., compression occurs at the peaks and amplification at the nulls. Such an envelope-following gain control results in a response with a fundamental frequency of the QDT, and the cochlear transducer nonlinearity facilitates the harmonics (Cooper, 1998; Olson, 2004). These harmonic distortions, in turn, modulate the primaries to give rise to a family of odd-order DPs of $(n+1)f_1-nf_2$ (LSBs) and $(n+1)f_2-nf_1$ (USBs), respectively. Moreover, varying gain of the cochlear transducer produces "pulsed" release of DPOAEs in the time domain (Kemp and Brown, 1986; Brown, 1994) that is similar to the temporal modulation pattern of low-frequency biasing reported here and previously (Bian *et al.*, 2002, 2004).

However, such a DP generation hypothesis cannot account for the observed even-order DPs that are not related to

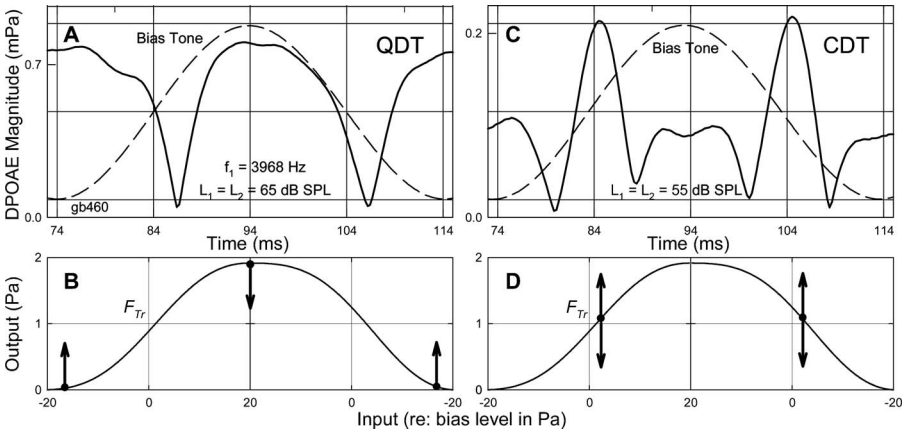


FIG. 12. Temporal modulation pattern in relation to the cochlear F_{tr}. (A) The modulation pattern of QDT during a period of the bias tone (dashed line). (B) Schematic representations of the OP change on the F_{tr} for the rise and fall of the bias pressure. The arrows indicate the compressive portions of the F_{tr} that correspond to the periodic generations of QDT in (A). (C) The modulation pattern of CDT within a biasing cycle (dashed line). (D) Illustration of the cochlear F_{tr} for the cycle of biasing. Arrows indicate the two centers of the F_{tr} with optimal gain where the peaks of CDT are produced (C). Note: the numbers of DP generation stages are different for the CDT and QDT.

QDT, e.g., f_1+f_2 and $3f_1-f_2$ (Fig. 3). These intermodulation DPs are merely the DPs from the compressive nonlinearity of the cochlear F_{tr}, because such a sigmoid-shaped nonlinear function is capable of producing both odd- and even-order DPs (Kemp and Brown, 1986; Lukashkin and Russell, 1999; Fahey *et al.*, 2000; Bian *et al.*, 2002; Bian, 2004). Fluctuation of the two-tone envelope could alter the gain of the cochlear transducer and the gain control follows the nonlinearity of the F_{tr}. The regulation of cochlear transducer could be achieved by adjusting the OP (Evans *et al.*, 1991), local feedback (Lukashkin and Russell, 1999), and adaptation of transduction (Fettiplace and Ricci, 2003). The notion of modulation sidebands indicates how the cochlear transducer operates in time because of the discrete relation between AM and spectral features (Fig. 1). Moreover, the response time inherited in such an envelope-following transducer could account in part for the finite time delay of various DPs (van der Heijden, 2005) and the formation of hysteresis in low-frequency modulated DPOAEs (Bian *et al.*, 2004; Bian and Chertoff, 2006). Therefore, DPs are consequences of the temporal behavior of the cochlear transducer that is determined by the compressive nonlinearity and its regulation.

B. Fine-structure of low-frequency modulated DPOAEs

An important result of the study is the presence of sidebands at multiples of f_{bias} around the DP component. This is not only in agreement with the nonlinear AM theory [Eq. (5)], but also with other experimental results in OAEs. Modulation sidebands are presented in the electrically evoked otoacoustic emissions (EEOAEs) when the cochlear partition is biased with a low-frequency tone (Kirk and Yates, 1998; Kirk, 2002). Spectral sidebands are also observed when stimulus-frequency otoacoustic emission (SFOAE) is manipulated with an intracochlear electrical current injection (Hubbard and Mountain, 1983), or an amplitude-varying acoustic suppressor tone (Neely *et al.*, 2005). Moreover, when stimulated with an AM signal, the spectra of basilar membrane vibration show sidebands around the f_c (Rhode and Recio, 2001). In addition, the present study shows that for even- or odd-DPs, sidebands with odd or even multiples of f_{bias} from the DPs are enhanced and others suppressed, respectively (Figs. 6 and 7). If an odd- or even-DP is treated as a carrier due to the AM of the cochlear transducer at f_{bias}, from Eq. (5), its sidebands $f_c \pm kf_m$ are of odd-orders when k is even, or even-orders when k is odd.

In the time domain, this harmonic distribution of sideband components is determined by the behavior of the cochlear transducer under cyclic biasing. As indicated in Figs. 12(a) and 12(b), during a complete cycle of rise and fall in biasing pressure, the cochlear transducer should experience one stage of even-DP generation at the peak of biasing, since these DPs are produced from the asymmetrical compression of the cochlear transduction (van Emst *et al.*, 1997; Bian, 2004). If including the minima of biasing pressure at the beginning and end of the cycle, there are three stages of compression (panel B), thus enhancing the sidebands related to the odd harmonics of the f_{bias} in the spectrum. In contrast, the odd-order DPs are generated mostly from the central portions around the inflection points of the F_{tr} (Bian *et al.*, 2002, 2004) within the cycle [Fig. 12(d)], there should be two bursts of odd-DPs (panel C) as previously observed "pulsed" CDT (Kemp and Brown, 1986; Brown, 1994). Consequently, the second harmonics of f_{bias} in the sidebands would be enlarged. Such a behavior is based on the principle that changes in DP magnitudes due to the OP shift follows the shape of the corresponding derivative(s) of the cochlear F_{tr} (Cheatham and Dallos, 1997; Frank and Kössl, 1997; Kirk and Patuzzi, 1997; Bian *et al.*, 2002). The shape of each F_{tr} derivative contains a certain number of lobes (Bian and Chertoff, 2006) that can contribute to the harmonic sidebands of DPs, as opposed to the derivatives of the biasing pressure (Zwicker and Scherer, 1987), because the latter cannot account for the differential behaviors of odd—versus even-DPs.

It is interesting to note that the magnitudes of the sidebands could be as high as or even greater than the DP (Figs.

6 and 7). There could be a reduction in DP magnitude as a result of suppression by the bias tone. Two prominent effects of the cochlear nonlinearity are the suppression of responses to one tone by the presence of another and distortion generation with two-tone input (Robles and Ruggero, 2001). The effects produced by a lower frequency tone demonstrate tonic and phasic suppressions (Geisler and Nuttall, 1997). The reduced spectral magnitude of odd-DPs represents the averaged amplitude during the whole length of the bias tone (Figs. 4 and 5), i.e., the tonic suppression. This suppression is a direct consequence of reduction in cochlear gain due to compression of the cochlear transducer induced by the bias tone. For even-order DPs, the greater sideband magnitudes also reflect the phase-dependent DP generation in the time domain. These additional distortions in the sidebands due to biasing of the cochlear transducer can lead to easier detection of the even-DP in the spectral domain. Therefore, these enhanced sidebands of even DPs may pose diagnostic utility under pathological conditions where odd DPs are significantly impaired (Choi *et al.*, 2004; Sirjani *et al.*, 2004).

C. Modulation contours and effective signal levels

The contours of the sidebands (Figs. 8 and 9) generally suggest that the magnitudes of the sidebands increase with both primary and bias tone levels. However, this trend varies depending on the type of sidebands and the type of DPs. For CDT, sideband I becomes measurable when bias tone level is greater then 4 to 8 Pa (106 to 112 dB SPL). This is consistent with the effective bias tone level of 105 dB SPL used to modulate DPOAEs in guinea pigs (Frank and Kössl, 1996, 1997), and 110 dB SPL in humans (Scholz *et al.*, 1999). Using low-frequency tone to modulate EEOAEs in guinea pigs, Kirk and Yates (1998) found that from 100 dB SPL the amplitude and the number of the EEOAE sidebands begin to rise. This is due to the sound pressure in the bias tone can push the cochlear transduction channels further into saturating range and produce more nonlinear interactions and suppressions among the DPs and primary tones, thus increasing the energy in the sidebands. Moreover, if DPs are the effects of a two-tone interaction, likewise modulation sidebands are the products of bias tone and the corresponding DP. Particularly, the sideband contours are found to be similar to the DPOAE contour maps as functions of L_1 and L_2 space in humans (Whitehead *et al.*, 1995), the gerbils (Mills, 2002), and the frogs (Meenderink and van Dijk, 2005). This suggests that DPOAEs and their low-frequency modulation sidebands are generated from the same mechanism: nonlinearity in hair cell transduction.

The sideband magnitude relative to the corresponding DP is an indicator of the extent of the AM of the DP. A value of greater than −10 dB marks the presence of significant modulation. Thus, contours of the relative sideband magnitudes reveal the most effective signal levels to produce the modulation of DPOAEs. From the contours of the CDT sideband at both pairs of primary frequencies (Fig. 8), it can be noted that the modulation of CDT is largest at low primary levels, preferably below 60 dB SPL. The bias tone level required to produce the maximal modulation has to be greater than 8–12 Pa (112–116 dB SPL). For QDT, the contours of sideband I indicate that to fully modulate QDT the bias tone level can be lowered to 4 Pa (106 dB SPL). This is because the magnitude of the QDT is quite small without biasing which is less than 10 dB SPL for primary level up to 70 dB SPL (Brown, 1993; Bian, 2004). For smaller DPOAE amplitudes, it requires less acoustical energy to effectively modulate the emissions. In future, effects of low-frequency biasing should be explored for primary levels below 50 dB SPL.

When the bias tone level is fixed, the DP sideband magnitudes grow with the primary level in a nonlinear manner (top panels of Figs. 10 and 11). The CDT sidebands grow at a rate less than 1 dB/dB and saturate at approximately 60–65 dB SPL primary level. This is comparable to the growths of SFOAE sidebands found in electrical biasing of cochlear transducer (Hubbard and Mountain, 1983). In comparison, the QDT sidebands rise with a steeper rate and saturate at higher primary levels. The growths of the absolute sideband magnitudes are consistent with the overall nonlinearity in the cochlear transducer. Saturations of DP sidebands and the DPOAE growth functions (e.g., Dorn *et al.*, 2001) indicate that the cochlear transducer begins to reduce its gain or compress its output around 60 dB SPL primary level. Therefore, as shown by their relative sideband amplitudes, the modulation is maximal at primary levels below 60 dB SPL for CDT [Figs. 10(c) and 10(d)] and above this level for QDT [Figs. 11(c) and 11(d)]. It is noteworthy that the growths of the sideband amplitudes for CDT and QDT are similar to those of the side lobes in their temporal modulation patterns (Bian *et al.*, 2004; Bian, 2004), thus the DP sidebands maybe correlate with the derivatives of the cochlear F_{tr}.

D. Frequency effect

It can be observed from Figs. 8 and 9 that there is a frequency effect on the low-frequency modulation of DPOAEs. As primary frequencies decrease, the bias tone becomes more effective as demonstrated in one or all of the three aspects: (1) the maximal sideband amplitude is greater; (2) the bias tone level that just produces a sideband of 10 dB above the noise floor is lower; and (3) for a given bias tone level, it can modulate DPs evoked at high primary levels. The differential effectiveness of the bias tone could be due to the distribution of the bias tone energy along the cochlear partition. For a tone at 25 Hz, the traveling-wave envelope peaks at the cochlear apex and the subsequent displacement of the cochlear partition gradually tapers off toward the base (Nilsen and Russell, 2000; Robles and Ruggero, 2001). The cochlear place where lower primary frequencies interact to generate DPOAEs receives more biasing than higher frequency locations, therefore, producing more significant modulation of DPs. To estimate cochlear F_{tr} at more basal regions, the frequency of the bias tone could be increased to improve the efficiency for modulating DPOAEs evoked with higher primary frequencies. When bias tone frequency is moved closer to the primaries, the effective bias tone level could be significantly reduced. Comparing the fact that the level of the bias tone has to be greater than the primaries to

the empirical practice that the DPOAEs can be maximized if the level of the lower frequency tone is elevated by 5–10 dB (Wiederhold *et al.*, 1986; Hauser and Probst, 1990), suggests that the DPOAEs and their modulation by the bias tone arise from the same nonlinear mechanism.

E. Issue of $2f_2-f_1$

There is an exception of an odd DP that shows the behavior of even DPs at lower primary frequencies (f_1 =3968 Hz). A third-order DP, $2f_2-f_1$, demonstrates a different modulation pattern from other odd DPs (Fig. 4). However, its spectral fine structures (Fig. 6 bottom left) indicate the features of both odd- and even-order DPs. One reason for this exception is probably that the DP contains a mixture of a primarily generated component and a secondary composition produced from the feedback mechanism (Yates, 1990). The cochlear transducer nonlinearity produces harmonics of the primaries (Cooper, 1998; Olson, 2004) and DPs containing $2f_2-f_1$ (Fig. 3). These harmonic distortions can feedback to the cochlear transducer as inputs to form other DPs, such as the difference tone of $2f_2$ and f_1. As $2f_2-f_1$ could be generated basal to f_2 place with shorter latency (Martin *et al.*, 1998), mechanisms involving the interactions between the harmonics of the primaries and other DPs may be possible (Fahey *et al.*, 2000). Thus, the secondary component could also be generated as a difference tone at the $2f_2$ cochlear place by $2f_2$ and the f_1 stimulus or as a summation tone at the f_2 location by the backward-traveling QDT and f_2 stimulus. In any case, these even-order components of $2f_2-f_1$ can be enhanced by the bias tone and become the dominant components in the waveform (Fig. 4) as oppose to being out-of-phase and canceling with the primary element without biasing (tail portion of the waveform). At higher frequencies, a more sharply tuned cochlear transducer could significantly attenuate the secondary components above f_2 thus reduces the contribution of the even-order components to $2f_2-f_1$. This is an example of the complexity and variation of DPOAEs generation mechanisms (Knight and Kemp, 2001; Shera, 2004) even though the fundamental source of the DPs is the cochlear nonlinearity (Fahey *et al.*, 2000; Lukashkin *et al.*, 2002).

VI. SUMMARY AND CONCLUSION

Low-frequency modulation of multiple component DPOAEs are examined in both time and frequency domains. The magnitudes of DPOAEs are either suppressed or enhanced depending on the bias tone phase and the type of DPs. Temporal envelopes of odd and even-order DPs demonstrate fundamentally different modulation patterns. Spectral analysis shows multiple modulation sidebands with intervals equal to multiples of the f_{bias} from the DP component. The amplitudes of the sidebands can be as high as or even greater than their centered DP if there is a significant AM in the time domain. These spectral fine structures and the variation of sidebands among odd- and even-DPs reflect a unique relation with their different temporal modulation patterns. Therefore, a possible application of measuring the modulation sidebands is to recover accurately the DPOAE modulation patterns using inverse transformation techniques, such as spectral heterodyne analysis (Kim *et al.*, 2001) or harmonic-matching the elements in Eq. (5). The cochlear F_{tr} can, in turn, be derived from these modulation patterns (Bian *et al.*, 2002, 2004; Bian, 2004), or more importantly, from a simple frequency analysis of the acoustic signal in the ear canal. Moreover, the modulation contours of the sideband magnitudes can provide a guide for selection of signal conditions for the optimization of low-frequency modulation recordings in a given subject. Findings of the present study suggest that neural perception of natural sounds in the central auditory system by detecting AM (Joris *et al.*, 2004) initializes at the onset of hearing: hair cell transduction. Furthermore, low-frequency modulation of EEOAEs found in amphibians (Manley *et al.*, 2001) with no differentiation of hair cells implies that the "active" and nonlinear process responsible for hearing sensitivity may universally present in the hair cell transduction channels.

ACKNOWLEDGMENTS

The author wishes to thank Mark Chertoff for help in the experiments and laboratory support of the study. Discussions and suggestions from Stephen Neely are appreciated. Comments from Jay Braun and two JASA reviewers have greatly improved the manuscript. This work was supported by a grant (R03 DC006165) from the National Institute on Deafness and Other Communication Disorders of the NIH.

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., and Chertoff, M. E. (**2006**). "Modulation patterns and hysteresis: Probing cochlear dynamics with a bias tone," in *Auditory Mechanisms: Processes and Models*, edited by A. L. Nuttall *et al.* (World Scientific, Singapore).

Bian, L., Chertoff, M. E., and Miller, E. (**2002**). "Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **112**, 198–210.

Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). "Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions," J. Acoust. Soc. Am. **115**, 2159–2172.

Brown, A. M. (**1986**). "Some experimental observations of responses evoked from the cochlear during two-tone stimulation," ORL **48**, 124–134.

Brown, A. M. (**1993**). "Distortion in the cochlea: Acoustic f2-f1 at low stimulus levels," Hear. Res. **70**, 160–166.

Brown, A. M. (**1994**). "Modulation of the hair cell motor: A possible source of odd-order distortion," J. Acoust. Soc. Am. **96**, 2210–2215.

Cheatham, M. A., and Dallos, P. (**1997**). "Low-frequency modulation of inner hair cell and organ of Corti responses in the guinea pig cochlea," Hear. Res. **108**, 191–212.

Cheng, J. (**1999**). "Estimation of active cochlear nonlinearity by multicomponent distortion-product otoacoustic emissions," Acust. Acta Acust. 85, 721–727.

Choi, C. H., Chertoff, M. E., Bian, L., and Lerner, D. (**2004**). "Constructing a cochlear transducer function from the summating potential using a low-frequency bias tone," J. Acoust. Soc. Am. **116**, 2996–3007.

Cooper, N. P. (**1998**). "Harmonic distortion on the basilar membrane in the basal turn of the guinea-pig cochlea," J. Physiol. (London) **509**, 277–288.

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

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Evans, B. N., Hallworth, R., and Dallos, P. (**1991**). "Outer hair cell electromotility: The sensitivity and vulnerability of the dc component," Hear.

Res. **52**, 288–304.
Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). "Nonlinear interaction that could explain distortion product interference response areas," J. Acoust. Soc. Am. **108**, 1786–1802.
Fettiplace, R., and Ricci, A. J. (**2003**). "Adaptation in auditory hair cells," Curr. Opin. Neurobiol. **13**, 446–451.
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.
Frank, G., and Kössl, M. (**1997**). "Acoustic and electrical biasing of the cochlear partition. Effects on the acoustic two tone distortions f_2-f_1 and $2f_1-f_2$," Hear. Res. **113**, 57–68.
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.
Hauser, R., and Probst, R. (**1990**). "The influence of systematic primary-tone variation L_2-L_1 on the acoustic distortion product emissions $2f_1-f_2$ in normal human ears," J. Acoust. Soc. Am. **89**, 280–286.
Hubbard, A. E., and Mountain, D. C. (**1983**). "Alternating current delivered into the scala media alters sound pressure at the eardrum," Science **222**, 510–512.
Hudspeth, A. J. (**1989**). "How the ear's works work," Nature (London) **341**, 397–404.
Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.
Kemp, D. T. (**2002**). "Exploring cochlear status with otoacoustic emissions," in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–47.
Kemp, D. T., and Brown, A. M. (**1986**). "Wideband analysis of otoacoustic intermodulations," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, New York), pp. 306–313.
Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (**2001**). "Adaptation of distortion product otoacoustic emission in humans," J. Assoc. Res. Otolaryngol. **2**, 31–40.
Kirk, D. L. (**2002**). "Interaction between adenosine triphosphate and mechanically induced modulation of electrically evoked otoacoustic emissions," J. Acoust. Soc. Am. **111**, 2749–2758.
Kirk, D. L., and Patuzzi, R. B. (**1997**). "Transient changes in cochlear potentials and DPOAEs after low-frequency tones: The 'two-minute bounce' revisited," Hear. Res. **112**, 49–68.
Kirk, D. L., and Yates, G. K. (**1998**). "Enhancement of electrically evoked oto-acoustic emissions associated with low-frequency stimulus bias of the basilar membrane towards scala vestibuli," J. Acoust. Soc. Am. **104**, 1544–1554.
Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.
Lonsbury-Martin, B. L., and Martin, G. K. (**2003**). "Otoacoustic emissions," Curr. Opin. Otolaryngol. Head Neck Surg. **11**, 361–366.
Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). "One source for distortion product otoacoustic emissions generated by low- and high-level primaries," J. Acoust. Soc. Am. **111**, 2740–2748.
Lukashkin, A. N., and Russell, I. J. (**1999**). "Analysis of the f2-f1 and 2f1-f2 distortion components generated by the hair cell mechanoelectrical transducer: Dependence on the amplitudes of the primaries and feedback gain," J. Acoust. Soc. Am. **106**, 2661–2668.
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.
Manley, G. A., Kirk, D. L., Köppl, C., and Yates, G. K. (**2001**). "*In vivo* evidence for a cochlear amplifier in the hair-cell bundle of lizards," Proc. Natl. Acad. Sci. U.S.A. **98**, 2816–2831.
Martin, G. K., Jassir, D., Stagner, B. B., and Whitehead, M. L. (**1998**). "Locus of generation for the 2f1-f2 vs 2f2-f1 distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlation," J. Acoust. Soc. Am. **103**, 1957–1971.
Meenderink, S. W., and van Dijk, P. (**2005**). "Characteristics of distortion product otoacoustic emissions in the frog from L_1, L_2 maps," J. Acoust. Soc. Am. **118**, 279–286.
Mills, D. M. (**2002**). "Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response," J. Acoust. Soc. Am. **112**, 1545–1560.
Moulin, A., and Kemp, D. T. (**1996**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation," J. Acoust. Soc. Am. **100**, 1640–1662.
Neely, S. T., Johnson, T. A., Garner, C. A., Gorga, M. P. (**2005**). "Stimulus-frequency otoacoustic emissions measured with amplitude-modulated suppressor tones," J. Acoust. Soc. Am. **118**, 2124–2127.
Nilsen, K. E., and Russell, I. J. (**2000**). "The spatial and temporal representation of a tone on the guinea pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **97**, 11751–11758.
Olson, E. S. (**2004**). "Harmonic distortion in intracochlear pressure and its analysis to explore the cochlear amplifier," J. Acoust. Soc. Am. **115**, 1230–1241.
Rhode, W. S., and Recio, A. (**2001**). "Basilar-membrane response to multicomponent stimuli in chinchilla," J. Acoust. Soc. Am. **110**, 981–994.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Scholz, G., Hirschfelder, A., Marquardt, T., Hensel, J., and Mrowinski, D. (**1999**). "Low-frequency modulation of the $2f_1-f_2$ distortion product otoacoustic emissions in the human ears," Hear. Res. **130**, 189–196.
Shera, C. A. (**2004**). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," Ear Hear. **25**, 86–97.
Sirjani, D. B., 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.
van der Heijden, M. (**2005**). "Cochlear gain control," J. Acoust. Soc. Am. **117**, 1223–1233.
van Emst, M. G., Klis, S. F. L., and Smoorenburg, G. F. (**1997**). "Identification of the nonlinearity governing even-order distortion products in cochlear potential," Hear. Res. **114**, 93–101.
Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L_1, L_2 space," J. Acoust. Soc. Am. **97**, 2359–2377.
Wiederhold, M. L., Mahoney, J. W., and Kellogg, D. L. (**1986**). "Acoustic overstimulation reduces $2f_1-f_2$ cochlear emissions at all levels in the cat," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer, New York), pp. 322–329.
Yates, G. K. (**1990**). "Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions," Hear. Res. **50**, 145–162.
Zwicker, E., and Scherer, G. (**1987**). "Correlation between time function of sound pressure, masking, and OAE suppression," J. Acoust. Soc. Am. **81**, 1043–1049.

Hybrid measurement of auditory steady-state responses and distortion product otoacoustic emissions using an amplitude-modulated primary tone

Johann A. Oswald
Lehrstuhl für Realzeit-Computersysteme, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany

Thomas Rosner and Thomas Janssen[a)]
Hals-Nasen-Ohrenklinik, Technische Universität München, Ismaningerstraße 22, 80675 Munich, Germany

(Received 2 December 2005; revised 20 March 2006; accepted 23 March 2006)

A maximum auditory steady-state response (ASSR) amplitude is yielded when the ASSR is elicited by an amplitude-modulated tone (f_c) with a fixed modulation frequency (f_m=40 Hz), whereas the maximum distortion product otoacoustic emission (DPOAE) level is yielded when the DPOAE is elicited using a fixed frequency ratio of the primary tones (f_2/f_1=1.2). When eliciting the DPOAE and ASSR by the same tone pair, optimal stimulation is present for either DPOAE or ASSR and thus adequate simultaneous DPOAE/ASSR measurement is not possible across test frequency f_2 or f_c, respectively. The purpose of the present study was to determine whether the ASSR and DPOAE can be measured simultaneously without notable restrictions using a DPOAE stimulus setting in which one primary tone is amplitude modulated. A DPOAE of frequency $2f_1$-f_2 and ASSR of modulation frequency 41 Hz were measured in ten normal hearing subjects at a test frequency between 0.5 and 8 kHz ($f_2=f_c$). The decrease in the DPOAE level and the loss in ASSR amplitude during hybrid mode stimulation amounted, on average, to only 2.60 dB [standard deviation (SD)=1.38 dB] and 1.83 dB (SD=2.38 dB), respectively. These findings suggest simultaneous DPOAE and ASSR measurements to be feasible across all test frequencies when using a DPOAE stimulus setting where the primary tone f_2 is amplitude modulated. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2197789]

PACS number(s): 43.64.Jb, 43.64.Ri [BLM] Pages: 3886–3895

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) and auditory steady-state responses (ASSRs) both provide frequency specific and quantitative assessment of hearing capability (Dimitrejvic *et al.*, 2001; Herdman and Stapells, 2001; Gorga *et al.*, 2003; Boege and Janssen, 2002). DPOAEs are acoustical responses generated within the cochlea at the outer hair cell level when the ear is stimulated simultaneously with two sinusoidal primary tones f_1 and f_2 with a fixed frequency ratio (f_2/f_1=1.2) across test frequency f_2. DPOAEs reflect the nonlinear cochlear sound amplification at a specific place, which is the overlapping region of the primary tones near f_2 and can be measured by means of a microphone placed in the outer ear canal (Kemp *et al.*, 1986). ASSRs are electrical responses measured from the scalp generated in the brainstem and subsequent cortical structures by the modulation frequency (f_m) of the eliciting amplitude-modulated tone with carrier frequency f_c reflecting cochlear activity at the f_c place (Galambos *et al.*, 1981; Herdman *et al.*, 2002).

Tone pairs are a possible choice for a simultaneous measurement of DPOAEs and ASSRs. However, the optimal tone separation for yielding maximum amplitude of the two measures is different. For ASSRs, a constant frequency separation across test frequency is required because the ASSR amplitude is highest at a fixed modulation frequency, i.e., f_m=40, 80, or 180 Hz (Galambos *et al.*, 1981; Cohen *et al.*, 1991). In contrast, for yielding maximum DPOAE amplitudes, frequency separation across test frequency f_2 is not constant because optimal frequency ratio of the primary tones is found to be f_2/f_1=1.2 (Gaskill and Brown, 1990; Brown *et al.*, 2000). Thus, when eliciting both DPOAE and ASSR by the same tone pair, optimal tone separation is only present for one of the two measures. Therefore, compromises with regard to an optimal stimulus setting are necessary when measuring DPOAE and ASSR simultaneously using tone pairs (Purcell *et al.*, 2002).

One way out of this dilemma is a stimulus setting in which one tone of the tone pair is amplitude modulated. In doing this, both optimal tone separation (f_2/f_1=1.2) and optimal modulation frequency (f_m=40, 80, or 180 Hz) across test frequency can be achieved. The question however is: How large are the decreases in amplitude of both cochlear and brainstem responses when using such a stimulus setting? Therefore, in the present study, influences of the modified stimuli on each response were quantified with respect to a change in amplitude and phase.

[a)]Electronic mail: t.janssen@lrz.tum.de

 0001-4966/2006/119(6)/3886/10/$22.50

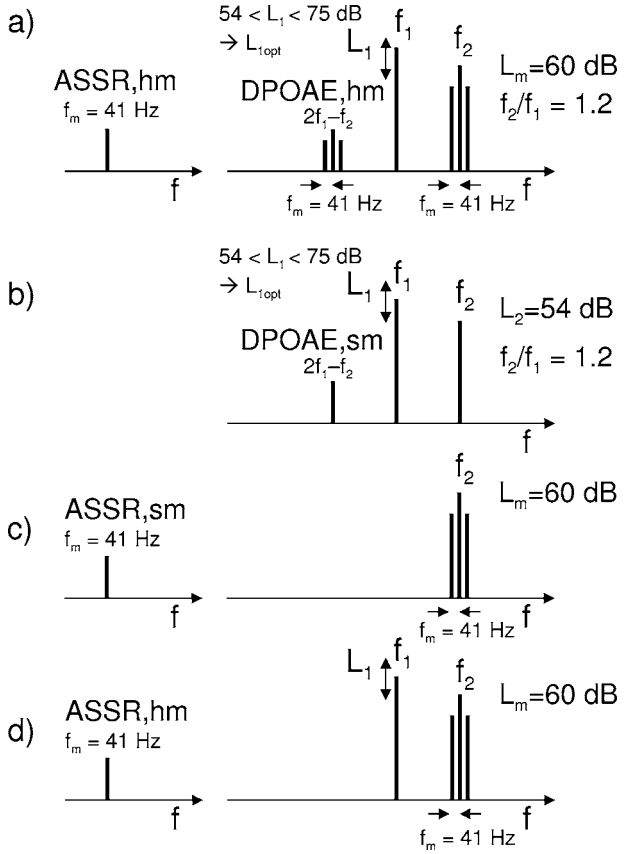


FIG. 1. Frequency spectrum (schematic) of the stimuli and the physiological responses in sm [(b) DPOAE, (c) ASSR] and hm [(d) DPOAE/ASSR]. The primary tone $f_2=f_c$ was amplitude modulated with a modulation frequency of 41 Hz. To find the optimal primary tone level separation at which the DPOAE level is maximum during ASSR/DPOAE measurement, L_1 was varied from 54 to 75 dB SPL, whereas L_2 was kept constant at 54 dB SPL. For investigating the impact of the primary tone f_1 on the ASSR amplitude (experiment 3), L_1 was varied from 54 to 81 dB SPL (d).

II. METHODS

A. Subjects

DPOAEs and ASSRs were measured in 10 ears of 10 awake human subjects with normal hearing (4 females, 6 males, aged between 24 and 30 years). According to pure-tone audiometry, audiometric thresholds were equal to or lower than 15 dB hearing loss between 125 Hz and 10 kHz. Middle ear and retrocochlear disorders were excluded by tympanometry and click-evoked auditory brainstem responses. None of the subjects had a history of otological or neurological disorders. Measurements were performed in a sound-insulated booth, while subjects were seated in a comfortable recliner and encouraged to read a book. Additionally, the reproducibility of the two measures and the suppressive effect of the primary tone f_1 on the ASSR amplitude were investigated in one subject.

B. Acoustic stimulation and recording

DPOAE measurements in single mode (sm) were performed at a fixed primary tone level [$L_2=54$ dB sound-pressure level (SPL)] and a fixed primary tone frequency ratio ($f_2/f_1=1.2$) [see Fig. 1(b)]. For a fixed L_2 and a fixed f_2/f_1 there is a distinct primary tone level L_1 which yields a maximum emission level (Gaskill and Brown, 1990; Whitehead *et al.*, 1995; Abdala, 1996; Kummer *et al.*, 2000). Thus, for searching for the highest DPOAE level in each individual ear the primary tone level L_1 was varied between 54 and 75 dB in 3 dB steps. The primary tone level L_1 at which the DPOAE level was highest is referred to as the optimal primary tone level $L_{1\ \mathrm{opt,sm}}$. The search for the highest DPOAE level was done at each test frequency f_2.

For eliciting ASSR measurements in sm, the primary tone f_2 was amplitude modulated [see Fig. 1(c)]. The level of the tone complex was set to $L_m=60$ dB SPL. The modulation depth was 100% so that level L_2 equaled 54 dB SPL for ensuring the same spectral intensity at the f_2 place in the cochlea in sm and hybrid mode (hm). For yielding high ASSR amplitudes in the awake subjects, a modulation frequency of $f_m=41$ Hz was chosen. The exact value was 41,0156 Hz.

Simultaneous DPOAE/ASSR measurements in hm were performed by applying the amplitude modulated primary tone f_2 (for eliciting ASSR) in combination with the nonamplitude modulated primary tone f_1 [see Fig. 1(a)]. Similar to DPOAE sm stimulation, L_1 was varied between 54 and 75 dB in 3 dB steps to find the optimal primary tone level $L_{1\ \mathrm{opt,hm}}$.

DPOAE and ASSR measurements both in sm and hm were performed at test frequencies $f_2=f_c=0.5$, 1, 2, 4, 6, and 8 kHz (exact values were: 499.02, 998.05, 1999.51, 3999.02, 5998.54, and 7998.05 Hz). In humans, the $2f_1$-f_2 DPOAE has the highest amplitude and was therefore used in this study. Due to the different compression of the primary tone traveling waves on the basilar membrane, the optimal overlapping of the traveling waves and hence the most contributing cochlea site to DPOAE generation is near the f_2 place (e.g., Kummer *et al.*, 1995). The carrier frequency f_c of the amplitude modulated tone represents the cochlear site at which the ASSR signal is elicited. Therefore, both the DPOAE and ASSR amplitudes were plotted across test frequency f_2 or f_c, respectively.

Before each measurement, in-the-ear calibration (Whitehead *et al.*, 1995) was performed for yielding constant sound pressure across the test frequency. Hard- and software for acoustic stimulation, DPOAE and ASSR measurements, and online data analysis were laboratory constructed using a digital signal processor (ADSP 2185, Analog Devices) and an analog digital converter (AD1819B, Analog Devices). The sound probe for stimulus generation and DPOAE recording was an ER-10C ear probe (Etymotic-Research). Aquired data were transmitted via infrared to a personal computer for data visualization, online analysis, and sequence control. The software was made using gnu c/c++ and QT under Linux and Visual DSP++ for programming the DSP. An additional off-line data analysis was done using MATLAB.

Testings were made in a tube (2 ccm) for excluding acoustic and technical distortions. No acoustic distortion was observed in the spectrum of the ear canal signal in either sm or hm. Also, when the electro encephalogram (EEG) was measured without acoustic stimulation, no equipment related distortion was observed.

DPOAEs and ASSRs were analyzed in the frequency

domain by the fast Fourier transform (FFT) of the average time-domain wave forms. The FFT resolution for the ASSR performed on the DSP was 3.4 Hz. The frame length was 2048 data points, and the recording sampling rate was 7 kHz (for all f_c, except at f_c=8 kHz where the ASSR was sampled with 8.4 kHz resulting in a frequency resolution of 4.1016 Hz). The frame length for DPOAE measurements was 1024 resulting in a frequency resolution of 6.8360 Hz, 13.6719 Hz, or 16.4062 Hz for the different frequency bands ($f_2<2$ kHz, 2 kHz$\leqslant f_2\leqslant$5 kHz, and $f_2>5$ kHz). Stimuli were digitally synthesized with a frame length of 2048, resulting in a frequency resolution of 3.4180 Hz, 6.8360 Hz, and 8.2031 Hz for $f_2<2$ kHz, 2 kHz$\leqslant f_2\leqslant$5 kHz, and $f_2>5$ kHz, respectively (stimulus sampling rate: 7 kHz, 14 kHz, and 16.8 kHz).

The averaging time for recording a DPOAE response was variable, its maximum being 10 s. DPOAEs were accepted as valid if the signal-to-noise ratio (SNR) exceeded 6 dB. The SNR was calculated using the ratio between the spectral amplitude of the DPOAE response at $2f_1$-f_2, and the mean of spectral amplitudes at the first, second, and fourth neighboring frequency at both sides. The third neighboring frequency was skipped because DPOAE sidebands at this frequency occurred due to the amplitude modulation of the primary tone f_2 [see Fig. 1(a)]. Consequently, the used "noise" bandwidth was variable from 27.2 Hz, 54.4 Hz to 65.6 Hz, below and above the DPOAE frequency (for $f_2<2$ kHz, 2 kHz$\leqslant f_2\leqslant$5 kHz, and $f_2>5$ kHz, respectively). The DPOAE measurement at the next test frequency was continued when the precedent DPOAE response reached validity (SNR=6 dB) or after time-out criterion occurred (10 s).

ASSR were obtained from electrodes at the vertex (Cz) and mastoid with a ground on the neck. Interelectrode impedance was less than 30 k Ohms at 200 Hz. The EEG was amplified 30,000 times and filtered using a high-pass filter of 33 Hz (12 dB/octave). An EEG recording sweep lasted for up to 179.9314 s, containing 615 sections of 0.2926 s each. Adaptive online artifact rejection was performed to eliminate muscle artifacts by comparing the root-mean square (RMS) of the current measured section of 0.2926 s to an individually estimated detection threshold. Once per session, an EEG sequence of 37.4491 s was recorded without stimulation. The mean RMS plus the standard deviation of the RMS of this sequence were chosen as the estimated detection threshold. The mean artifact rejection limits were set to 7.8 μV with a standard deviation of 1.6 μV. The validity of each ASSR was proven by using the F-technique (Picton *et al.*, 1987; Dobie and Wilson, 1993; Valdes *et al.*, 1997). To estimate the significance of each ASSR, the power of the response frequency was compared to the power of six neighboring frequencies (three above and three below the response). By using a spectral resolution of 3.4180 Hz (4.1016 Hz for f_2=8 kHz), the bandwidth of estimated noise level was thus 10.254 Hz (12.3047 Hz) above and below the response frequency. The significance of this power ratio was assessed using the F-test with 2 and 12 degrees of freedom (Zureck, 1992; John and Picton, 2000). An ASSR was considered to be valid if the F-ratio of the signal to noise was uninterruptedly significant at the $p<0.05$ level for longer than 30 s. Testing every 6 s, the ASSR recording was cut short only after five continuously estimated significant F-ratios.

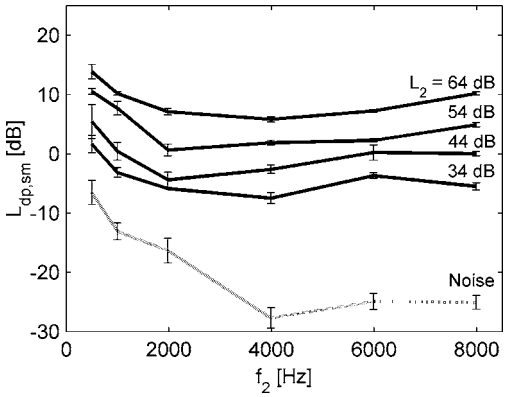


FIG. 2. Mean and standard deviation of the DPOAE level (solid line) and the noise floor (dashed line) across test frequency f_2 from repeated measurements (n=10) in a single subject. DPOAEs were elicited at different L_2 from 34 to 64 dB SPL for evaluating the test performance of DPOAE measurements in sm.

The cosine onset phase φ of the ASSR was measured and converted to latency for easier interpretation by using the "preceding cycles technique" (see John and Picton, 2000). For this, we conducted different ASSR measurements at a fixed carrier frequency f_c by varying the modulation frequency in a small range (f_m=37, 41, and 45 Hz). Absolute latency variation for the different f_m was smallest when two additional cycles were taken into account. Thus, the estimated phase delay φ' was calculated by $\varphi'=\varphi+720°$. The estimated absolute ASSR latency was obtained by the formula $t_{\mathrm{ASSR}}=\varphi'/(f_m 360)$.

C. Measurement protocols

1. Experiment 1

For evaluating the quality of the laboratory-constructed equipment used in this study, in one subject, ten repetitive DPOAE and ASSR measurements in sm were performed [for stimulus paradigm see Figs. 1(a) and 1(b)]. This was done at different primary tone levels from L_2=64 dB SPL to L_2=34 dB SPL (10 dB steps, $L_1=0.4\,L_2+39$) for DPOAEs and, accordingly, from L_m=70 dB SPL to L_m=40 dB SPL (10 dB steps) for ASSRs. For ASSR measurements, a fixed time-out criterion of 5 min was used (without short-cut option) resulting in ten recording sessions lasting 20 h in total. Repetitive DPOAE measurements were performed at the optimal L_1 setting at which the emission was highest (see Sec. II B). The DPOAE measurements lasted 6 h in total. This was done in two recording sessions. For each session, the position of the ear probe in the outer ear canal was unchanged. Results of Experiment 1 are shown in Figs. 2–4.

2. Experiment 2

In order to quantify the loss in amplitude when eliciting DPOAEs and ASSRs simultaneously, DPOAEs and ASSRs were recorded in sm [see Figs. 1(b) and 1(c)] and hm [Fig. 1(a)] in the ten subjects. In hm, the primary tone L_1 was

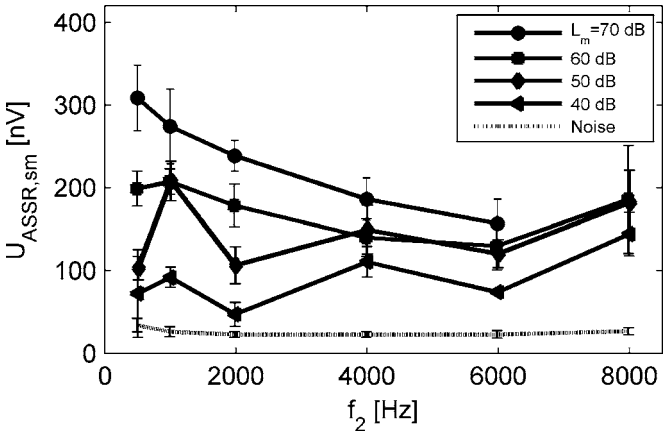


FIG. 3. Mean and standard deviation of the ASSR amplitude (solid line) and the noise floor (dashed line) across test frequency $f_2=f_c$ from repeated measurements ($n=10$) in a single subject. ASSRs were elicited at different L_2 from 40 to 70 dB SPL for evaluating the test performance of ASSR measurements in sm.

switched off after the procedure to search for the optimal L_1 at which the highest DPOAE level was terminated. On average, the primary tone f_1 was present for only 40 s, which is one-fifth of the ASSR recording time. Measurements per subject lasted 2 h and were conducted in one session. Results of Experiment 2 are shown in Figs. 5–7 and in Tables I and II.

3. Experiment 3

For investigating suppressive effects of the primary tone f_1 on the ASSR amplitude, ASSRs were measured in one subject in hm, where (in contrast to Experiment 2) the primary tone was neither changed in level nor switched off but was present during the whole measurement which lasted 3 min [see Fig. 1(d)]. L_1 was varied between 54 and 81 dB in steps of 3 dB. L_m was set to 60 dB SPL. ASSRs were recorded in the presence and absence of the primary tone. Results of Experiment 3 are shown in Fig. 8.

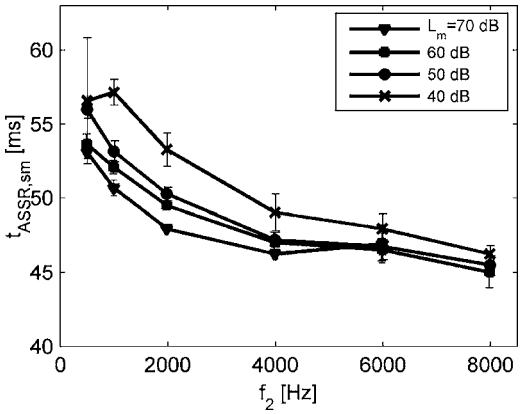


FIG. 4. Mean and standard deviation of the ASSR latency across test frequency $f_2=f_c$ from repeated measurements ($n=10$) in a single subject. ASSRs were elicited at different L_2 from 40 to 70 dB SPL.

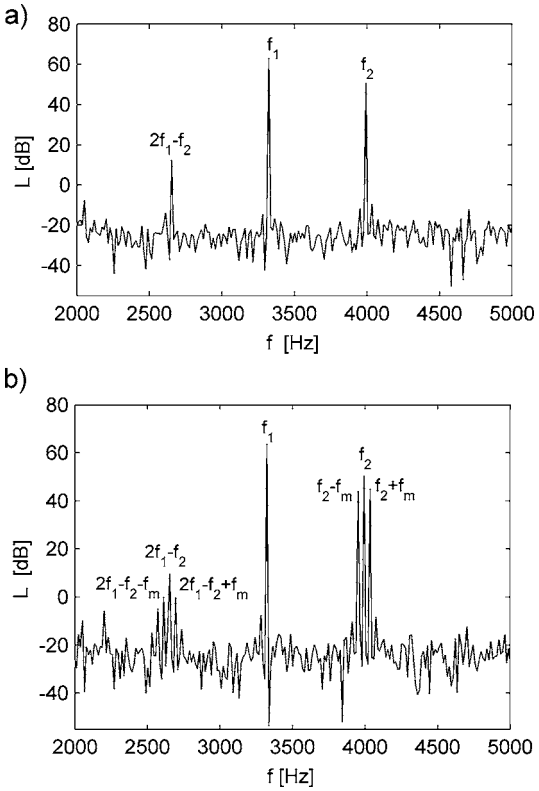


FIG. 5. Frequency spectrum of a DPOAE response from one subject in sm (a) and hm (b). In sm a clear DPOAE response was visible at $2f_1$-f_2. For stimulus setting, see Fig. 1(b). In contrast, in hm additional side-band DPOAEs occurred below and above the $2f_1$-f_2 response due to the side-bands of the amplitude-modulated primary tone f_2; the distance being f_m =41 Hz.

III. RESULTS

A. DPOAE and ASSR reproducibility (Experiment 1)

The mean and standard deviation of the DPOAE level $L_{dp,\mathrm{sm}}$, obtained from repeated DPOAE measurements (n =10) in sm, in one subject are shown in Fig. 2. The mean DPOAE level decreased with decreasing the primary tone

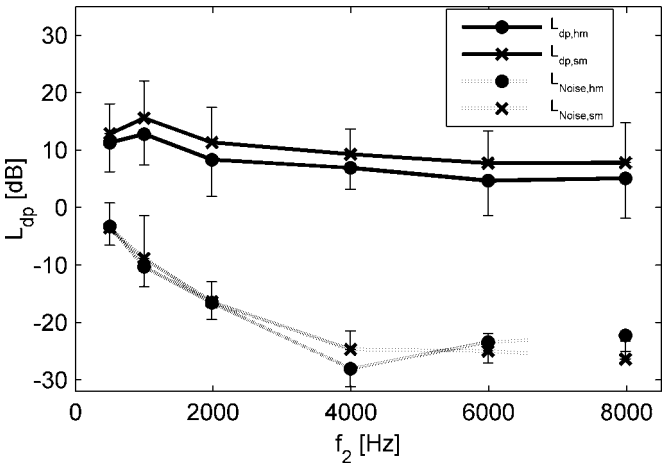


FIG. 6. Mean and standard deviation of the DPOAE level (solid lines) and noise level (dashed lines) across test frequency f_2 obtained in hm and sm in ten subjects. For stimulus setting, see Figs. 1(a) and 1(b), respectively. The decrease in DPOAE level during hm stimulation amounted to about 3 dB, nearly independent of test frequency.

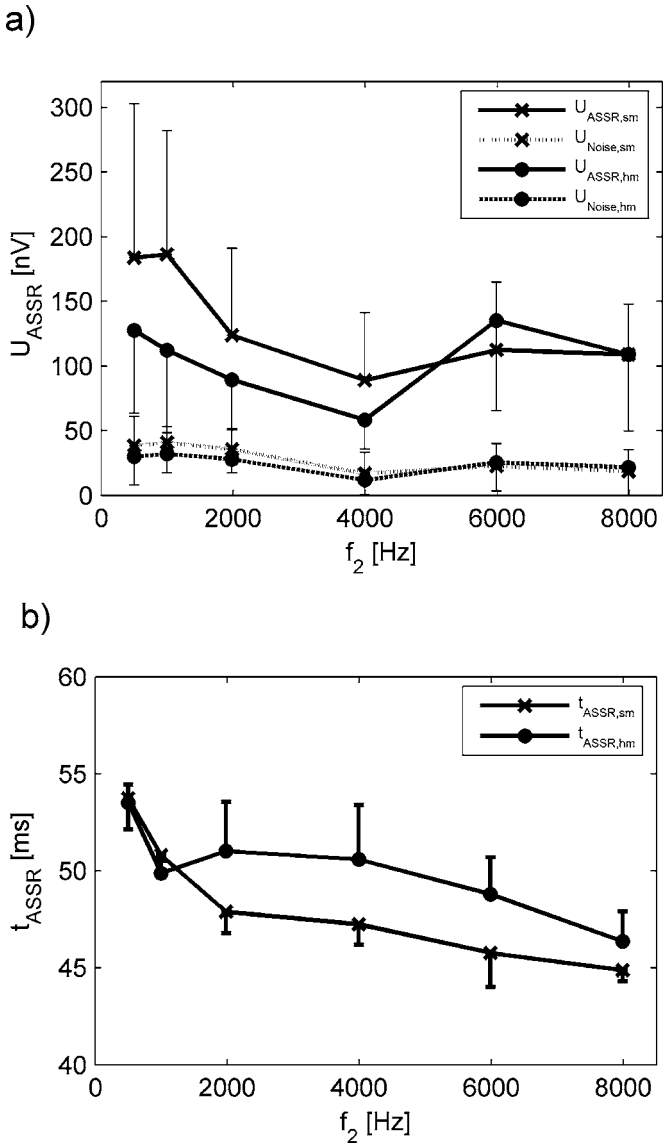


FIG. 7. Mean and standard deviation of the ASSR amplitude (solid lines) and noise level (dashed lines) (a) and the ASSR latency (b) across test frequency $f_2=f_c$ obtained in hm and sm in ten subjects. For stimulus setting, see Figs. 1(a) and 1(c), respectively. Decrease in ASSR amplitude during hm stimulation occurred only at the lower test frequencies. Increase in ASSR latency during hm stimulation occurred only at the higher test frequencies.

level. The standard deviation of the DPOAE level varied between 0.2 dB and 2.2 dB. The standard deviation tended to increase with decreasing stimulus levels and was highest at the low test frequencies. The increase in standard deviation toward lower frequencies can be attributed to the increasing noise level with decreasing frequency. Due to the lower noise-floor level at the higher test frequencies, the highest SNRs could be yielded between 4 and 8 kHz. The low-noise floor and the small standard deviation suggest good DPOAE test performance of the custom-built system.

The mean and standard deviation of the ASSR amplitude $U_{\mathrm{ASSR,sm}}$ and latency $t_{\mathrm{ASSR,sm}}$ obtained from repeated ASSR measurements ($n=10$) in sm in one subject are presented in Figs. 3 and Fig. 4, respectively. Responses were tested valid (F-test) at the 5% significance level in 211 out of 230 recorded ASSRs. At the lowest stimulus level ($L_m=40$ dB), ASSR amplitudes were not discriminable from the noise floor by the F-test in 19 out of 60 cases. The mean ASSR amplitude decreased with decreasing stimulus level (except for $f_c=1$ kHz at $L_m=50$ dB SPL). However, only at $L_m=70$ and 60 dB SPL a significant decrease was present (ANOVA: $F=35.211$; $df=4.50$; $p<0.001$). The highest amplitudes were found at the lower test frequencies. At $L_m=40$ and 50 dB SPL, no decrease in ASSR amplitude with increasing f_c was detected. The variation coefficient (= standard deviation/mean value) of the ASSR amplitude varied between 7% and 35% across stimulus level and test frequency.

The mean ASSR latency $t_{\mathrm{ASSR,sm}}$ exponentially increased with decreasing test frequency f_c (at constant stimulus level L_m) (Fig. 4). The difference in latency between $f_c=500$ Hz and $f_c=8$ kHz amounted to about 7.5 ms, and thus reflects the propagation time of the traveling wave from the base to the apex in the cochlea. The mean ASSR latency $t_{\mathrm{ASSR,sm}}$ also decreased with increasing stimulus level (at constant test frequency, f_c), especially at the lower test frequencies which can in part be attributed to the fact that the area of excitation in the cochlea is expanded from apical to basal with increasing stimulus level. The latencies' variance coefficient varied between 1.3% and 13.2%. The highest standard deviations occurred at the lower stimulus levels independent of test frequency.

B. DPOAE and ASSR in sm and hm (Experiment 2)

Figure 5 compares the amplitude spectra of a typical DPOAE spectrum obtained in sm (a) and hm (b) in one subject. In hm, clear sidebands beside the $2f_1$-f_2 DPOAE signal appeared, their distance corresponding to the modulation frequency f_m [Fig. 5(b)]. Tests in a tube (2 ccm) using the same stimuli did not show any $2f_1$-f_2 DPOAE signal or sidebands. The sidebands around $2f_1$-f_2 are therefore suggested to be additional DPOAEs elicited by the sidebands of the

TABLE I. Mean and standard deviation of the intraindividual difference in DPOAE level (first row) and DPOAE phase (second row) when comparing sm and hm (ten subjects). The ratio of the number of cases in which sm stimulation resulted in a higher DPOAE level than in hm and the respective number of valid data pairs are listed in the third row. The mean and standard deviation of the difference between the optimal stimulus level $L_{1\ \mathrm{opt}}$ for yielding the highest DPOAE level obtained in sm and hm are given in the last row.

f_2 (kHz)	0.5	1.0	2.0	4.0	6.0	8.0
ΔL_{dp} (dB)	1.57 (2.22)	2.77 (1.26)	2.99 (1.19)	2.37 (0.79)	3.05 (1.37)	2.72 (0.88)
$\Delta\phi_{dp}$ (degrees)	−8.5(59)	17.5 (42)	−8.3(20)	−7.9(23)	4.5 (25)	−12.3(41)
$\lvert\{\Delta L_{dp}>0\}\rvert$	8/9	10/10	10/10	10/10	10/10	10/10
$\Delta L_{1\ \mathrm{opt}}$ (dB)	1.4 (1.5)	0.2 (4.7)	−0.8(2.5)	−1.4(3.4)	−1.4(1.9)	−0.2(1.3)

TABLE II. Mean and standard deviation of the intraindividual difference in ASSR amplitude (first row) and ASSR latency (second row) at test frequencies f_c when comparing sm and hm (ten subjects). The last two rows indicate the number of individuals in which sm ASSR had higher amplitudes and longer latencies than in hm. The total number of data pairs in which both measurements were significant is given as second number.

f_c (kHz)	0.5	1.0	2.0	4.0	6.0	8.0
ΔU_{ASSR} (nV)	56.24 (96.48)	73.99(83.01)	34.42(79.41)	30.71 (45.18)	−22.81(86.29)	−0.09(31.78)
ΔU_{ASSR} (dB)	−3.22	−4.73	−3.10	−3.35	1.66	0 dB
Δt_{ASSR} (ms)	0.22 (1.56)	0.92 (1.41)	−3.13(2.57)	−3.34(2.85)	−3.03(1.92)	−1.48(1.76)
$\lvert\{\Delta U_{ASSR}>0\}\rvert$	6/7	7/8	6/8	3/5	3/6	3/5
$\lvert\{\Delta t_{ASSR}>0\}\rvert$	5/7	6/8	0/8	0/5	0/6	1/5

amplitude-modulated f_2 primary tone and the f_1 primary tone. In the following, these additional DPOAE components are not considered.

The mean and standard deviation of the $2f_1$-f_2 DPOAE level obtained in sm and hm for the ten normally hearing subjects are presented in Fig. 6. The mean DPOAE level in sm was higher than that obtained in hm. The difference $L_{dp,\mathrm{sm}}-L_{dp,\mathrm{hm}}$ amounted to 2.6 dB, on average, across the test frequency and showed a significant effect (paired samples t-test: $t=14.482$, $df=58$, $p<0.001$). The standard deviation of the difference was 1.38 dB. The highest decrease in DPOAE level was 6.4 dB, the lowest 0.13 dB. In Table I, the mean and standard deviation of the difference are listed over the test frequencies. Additionally, the number of valid data pairs per test frequency is given. The mean standard deviation of the DPOAE level across the test frequency was 6.27 dB for sm and 6.28 dB for hm. The noise levels in hm and sm were similar, varying between −5 and −27 dB SPL. The mean DPOAE phase difference $\varphi_{dp,\mathrm{sm}}-\varphi_{dp,\mathrm{hm}}$ amounted up to 18° with a standard deviation between 20° and 42°. Only at $f_2=500$ Hz, the standard deviation increased to 60° (see Table I).

In Fig. 7(a), the mean and standard deviation of the ASSR amplitude obtained in sm and hm are plotted across the test frequency f_c for the ten subjects. The ASSR amplitude in hm was significantly lower than that in sm for test frequencies between 0.5 and 4 kHz (paired samples t-test: $t=3.419$; $df=27$; $p<0.05$). At 6 kHz, the ASSR amplitude in hm was higher than that found in sm. At 8 kHz, they were quite similar. At both frequencies, no significant effect was present (paired samples t-test: $t=-0.324$; $df=11$; $p=0.752$). The mean loss in ASSR amplitude across all test frequencies amounted to 1.83 dB (referenced to sm amplitudes) and was significant (paired samples t-test: $t=2.578$; $df=38$; $p<0.05$). The highest decrease in ASSR amplitude was 10.4 dB; the lowest. 0.2 dB, respectively. The highest enhancement was 5.1 dB. The absolute amplitudes, percentages of loss in amplitude, number of valid responses, and number of responses that exhibited higher ASSR amplitude in sm per test frequency are given in Table II.

In Fig. 7(b), the mean and standard deviation of the ASSR latency obtained in sm and hm are plotted across the test frequency f_c for the ten subjects. The latency in sm increased exponentially with decreasing test frequency. The difference in latency between $f_c=500$ Hz and $f_c=8$ kHz amounted to about 8 ms, and thus was similar to that found in the single subject (see Fig. 4). In contrast, the ASSR latency in hm increased with decreasing test frequency only slightly, and did not exhibit an exponential shape. The ASSR latency obtained in hm and sm differed significantly (paired samples t-test: $t=-3.590$; $df=38$; $p<0.001$). The mean variation coefficient across the test frequency and subjects in sm was 8% (standard deviation 3%); and in hm, 10% (standard deviation 3%).

During ASSR measurements, L_1 was varied between 54 and 75 dB SPL to determine the optimal primary tone level setting ($L_2=54$ dB SPL, $L_{1\ \mathrm{opt}}$) at which the DPOAE level is highest. Across the test frequency and subjects, the mean $L_{1\ \mathrm{opt,hm}}$ amounted to 61 dB SPL.

C. Influence of L_1 on ASSR (Experiment 3)

For investigating the impact of L_1 on the ASSR amplitude, L_1 was varied between 54 and 81 dB in 3 dB steps. This was done for all test frequencies. Figures 8(a) and 8(b) show the ASSR amplitude of consecutive measurements in the presence and in the absence of the primary tone f_1 at $f_c=0.5$ kHz and $f_c=4.0$ kHz, respectively. The ASSR amplitude at $f_c=0.5$ kHz was affected for each L_1 and decreased with increasing L_1. For L_1 higher than 66 dB SPL, no valid response could be detected within the recording session of 3 min. At $f_c=4$ kHz, no singificant reduction occured for L_1 between 54 and 69 dB. For L_1 exceeding 66 dB SPL, the ASSR amplitude decreased considerably so that no valid response was detectable. Figure 8(c) shows the mean ASSR amplitudes calculated from ASSRs for all test frequencies revealing a significant reduction in ASSR amplitude for L_1 exceeding 69 dB SPL.

The dashed horizontal line in Fig. 8 represents the ASSR amplitude for hm stimulation (Experiment 2), where the primary tone level L_1 was varied between 54 and 75 dB SPL and was switched off after the procedure to search for the optimal L_1. Since in hm L_1 was switched off after about one-fifth of the ASSR recording time, a smaller reduction in amplitude was present.

IV. DISCUSSION

From the high discrepancy in recording time (seconds for a DPOAE, minutes for an ASSR) arises the question: Why waste time? Why not record DPOAEs during ASSR measurements? For example, ASSR recording time could be used for optimizing DPOAE stimulus settings. The present study therefore addressed the question as to whether ASSRs and DPOAEs can be recorded simultaneously using a modified DPOAE stimulus setting where there is an optimal tone separation for yielding maximum amplitudes of both ASSRs

and DPOAEs independent of test frequency. The present work is a feasibility study and should answer the question about to what extent the additional stimulus (primary tone f_1) and the amplitude modulation (primary tone f_2) influence the respective measure's amplitude or phase.

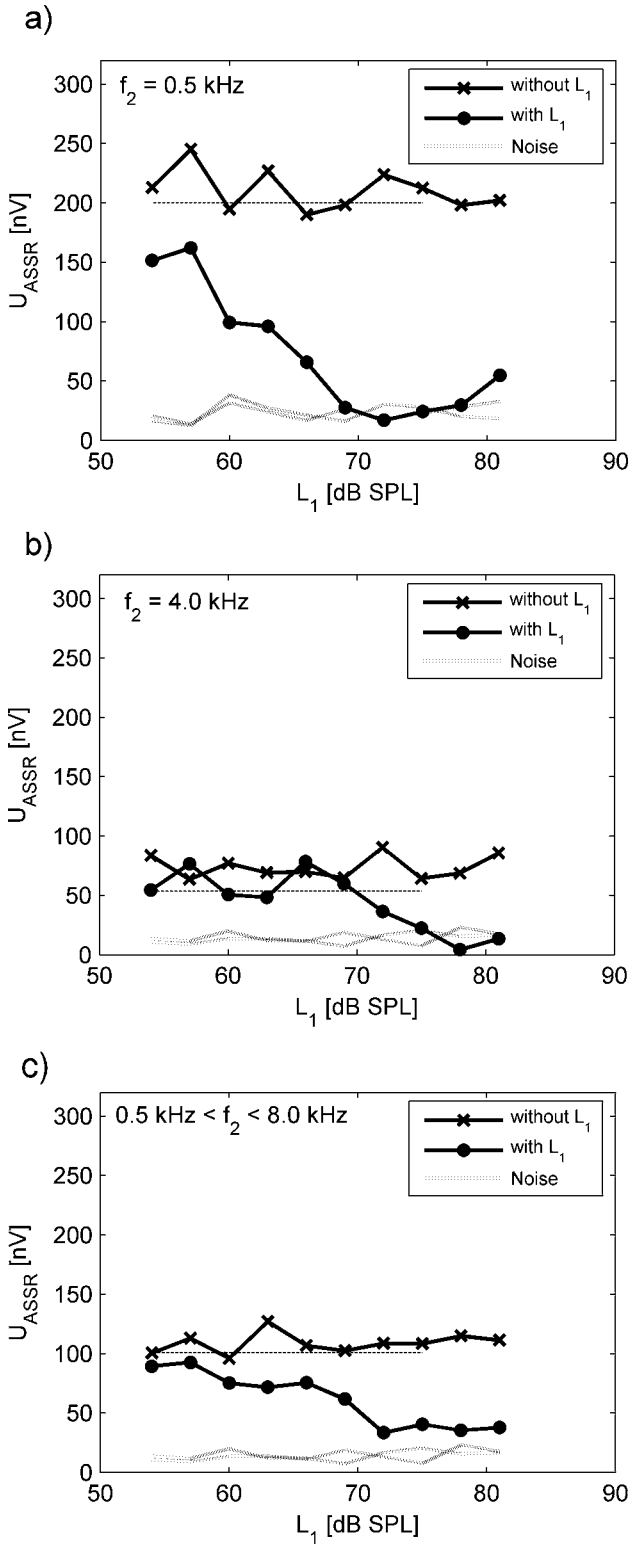


FIG. 8. ASSR amplitude with and without the presence of the primary tone f_1 in one subject at $f_2=f_c=0.5$ kHz (a), $f_2=f_c=4.0$ kHz (b), and mean ASSR amplitude for all test frequencies (c). Primary tone level L_1 was varied between 54 and 81 dB SPL. Primary tone f_1 was present during the whole ASSR measurement (3 min). The level L_m of the amplitude-modulated tone f_2 was 60 dB SPL. For stimulus setting, see Fig. 1(d). The dashed horizontal line represents the hm ASSR amplitude, where L_1 was swiched off after about one-fifth of the total test duration.

A. DPOAE and ASSR test retest stability

As there was no tool available on the market for a simultaneous ASSR and DPOAE recording, we developed our own hardware and software which had to be evaluated with respect to the quality of the test performance. The reproducibility (test-retest stability) of a measure is an important quality characteristic and was therefore investigated in the present study.

The DPOAE level did change only slightly during repetitive recordings, with its standard deviation varying across the test frequency and primary tone level from 0.2 to 2.2 dB (see Fig. 2). In the literature, the variability of the DPOAE level is reported to be in the range of 1.5 dB (e.g., Lasky, 1998a, b). Moreover, the standard deviation of the DPOAE level obtained in repetitive measurements is reported to depend on the SNR amounting to 0.2 dB at SNR =40 dB, 0.7 dB at SNR=20 dB, and 2.5 dB at SNR=6 dB on average across subjects (Janssen *et al.*, 2005). These findings are in accord with the present study. Due to the dependency on vigilance, a 40 Hz ASSR amplitude is reported to differ strongly during repetitive measurements (Pethe *et al.*, 2001). In our study, the variation coefficient of the ASSR amplitude varied from 7.7% (at high SNR) to 35.0% (at low SNR), and thus was similar to that reported in the literature.

Another quality characteristic is the precision with which ASSR latency is able to assess the propagation time of the traveling wave in the cochlea. In our study, the difference in latency obtained for f_c between 0.5 and 8 kHz amounted to about 7.5 ms (see Fig. 4), and thus reflects the propagation time of the traveling wave from the base to the apex. This finding is in accordance with the travel time in the cochlea known from physiological data (Dallos and Cheatham, 1971). Moreover, our absolute ASSR latency, e.g., at f_c =1 kHz and L_m=70 dB SPL, amouted to 50.2 ms and was thus similar to that reported by Herdman *et al.* (2002) (48 ms). However, it differed from that reported by Stapells *et al.* (1987) (41 ms) and Stapells *et al.* (1984) (34 ms). This discrepancy can be due to different electrode placements and different methods for estimating the ASSR latency used in these studies. Nevertheless, the high reproducibility of the ASSR latency suggests the good ASSR test performance of our custom-built system.

B. Stimulus paradigm

In the present study, a DPOAE stimulus setting was used for eliciting the ASSR and DPOAE simultaneously in which one of the primary tones was amplitude modulated. In doing this, optimal stimulus settings could be achieved for both measures independent of test frequency f_2 or f_c, respectively. Thus, the present approach is different from that reported in the literature (Purcell *et al.*, 2002), in which DPOAEs and ASSRs were elicited by two unmodulated tones that were separated corresponding to the modulation frequency f_m be-

ing either around 85 Hz or 180 Hz. Because of that restriction, the optimal frequency ratio ($f_2/f_1=1.2$) for eliciting DPOAEs could not be achieved across the test frequency. Moreover, a modulation frequency of 40 Hz could not be realized in the Purcell *et al.* approach (2002), because frequency separation was too small for eliciting DPOAEs especially at high test frequencies.

Furthermore, in the Purcell *et al.* approach (2002), an equilevel tone pair was used to achieve the highest ASSR amplitudes. This means a further restriction with respect to DPOAE performance, since equilevel primary tone settings do not yield maximum DPOAE levels. However, maximum levels are achieved when using a primary tone setting which accounts for the different compression of the two primaries at the DPOAE generation site in the cochlea at the f_2 place (e.g., $L_1=0.4\ L_2+39$; following Kummer *et al.*, 2000). In our approach, for yielding maximum DPOAE level in each individual ear, the primary tone level L_1 was varied around the initial setting ($L_1=0.4\ L_2+39$), while L_2 was constant. Furthermore, it is known that an amplitude-modulated tone with a 100% envelope modulation yields a higher ASSR amplitude than a tone pair (see Dimitrijevic *et al.*, 2001; John *et al.*, 2001). In our approach, an amplitude-modulated f_2 tone with a modulation depth of 100% was used. This is an additional advantage over the Purcel *et al.* (2002) approach. Nevertheless, when using our protocol for eliciting DPOAE and ASSR simultaneously, two questions arise. First, how does the amplitude modulation of the primary tone f_2 influence the DPOAE amplitude? Second, does the nonmodulated primary tone f_1 which is near f_2 have an impact on the ASSR amplitude? Both questions will be discussed in Secs. IV C and IV D.

C. Simultaneous DPOAE/ASSR recording

When comparing sm and hm measurements, the decrease in the DPOAE level and ASSR amplitude amounted, on average, across subjects and test frequencies to 2.6 dB [standard deviation (SD)=1.38 dB] and 1.8 dB (SD =2.38 dB), respectively. This relatively low reduction in the measures' magnitude suggests simultaneous DPOAE and ASSR measurement to be feasible. However, it should be noted that, in view of a clinical application, a reduction in amplitude can be critical especially at near-to-threshold stimulus levels. It should be emphasized that a longer recording time might be appropriate for compensating for this effect while maintaining the detection criteria.

The loss of DPOAE level is similar to that reported by Kummer *et al.* (1995), who investigated the change in DPOAE level in the presence of a third tone (suppressor tone f_3) while varying its frequency and level. In their protocol, the suppressor tone f_3 was set 50 Hz above or below f_2 and thus behaved similar to our sidebands, whose distance to the f_2 tone was 41 Hz (see Fig. 5). This setting resulted in a loss of DPOAE level of approximately 2 dB and was thus comparable with our findings during hybrid stimulation. It is therefore suggested that the sidebands of the amplitude-modulated primary tone act like suppressor tones. Which sideband of the modulated tone has more of a suppression effect was not investigated in our study. However, since asymmetry in DPOAE suppression tuning curves was found (Kummer *et al.*, 1995; Gorga *et al.*, 2003), one can assume that the lower sideband might have more influence on the DPOAE level than the upper sideband.

The standard deviation of the 40 Hz ASSR amplitude is reported to be high due to the variation in the activity of the EEG, which is known to differ strongly, depending on the vigilance effects (Pethe *et al.*, 2001). In our study, the standard deviation of the ASSR amplitude was also considerably high. Thus, a significant loss of ASSR amplitudes was only present for carrier frequencies between $f_c=0.5$ kHz and 4 kHz. The unexpected result that no change occurred at 6 kHz and 8 kHz might be due to the high variability in the amplitude. Furthermore, since the ASSR amplitude decreased with increasing f_c (see Fig. 3), fewer valid responses were present at 6 kHz and 8 kHz than at lower test frequencies (see Table II). This effect has to be investigated in more detail in future studies.

Ross *et al.* (2003) investigated the effect of a suppressor tone on the 40 Hz ASSR amplitude and found the highest reduction to occur at suppressor frequencies next to the carrier frequency. At a frequency ratio of 1.2 ($f_c=250$ Hz, $f_s=208$ Hz), the reduction in ASSR amplitude amounted to about 4.4 dB which is considerably higher than that found in our study (1.8 dB). However, these findings are not directly comparable, because our lowest f_c was 500 Hz. Since our data showed an increasing reduction in ASSR amplitude with decreasing f_c (see Fig. 8), similar effects could appear at 250 Hz.

Another study by John *et al.* (1998) investigated the effect on 40 Hz ASSR when multiple stimulation was applied (several amplitude modulated tones). In their study, the reduction of the ASSR amplitude amounted to about 4 dB when using a carrier frequency $f_c=1$ kHz and a "suppressor" frequency of $f_c/1.2$. The study also showed that suppressor tones with frequencies below the carrier frequency f_c had less suppression effects on the ASSR amplitude than those above the carrier frequency.

The significant difference in ASSR latency, when comparing sm and hm data [see Fig. 7(b)], indicates that the f_1 tone acts as a suppressor on the ASSR. However, the influence was not the same for different test frequencies.

Whether hybrid ASSR/DPOAE measurements might be more suited than serial measurements of the single ASSR and DPOAE measurements has to be proven in clinical practice. Basically, one could state that the observed loss in the DPOAE level of about 3 dB could be compensated for by doubling the recording time. However, a reduction of 3 dB is only critical at close-to-threshold responses where the SNR is low. Most importantly, in hybrid ASSR/DPOAE measurements, the DPOAE recording time is generally not crucial because a doubling, from, e.g., 4 s to 8 s, in recording time is negligible in view of the much longer ASSR recording time (3 min in our study). The mean loss of 1.8 dB in ASSR amplitude cannot be directly transferred to additional time consumption, as vigilance plays a major role in the 40 Hz ASSR. Future studies have to be performed using repeated measurements with an *a priori* defined significance level in

order to verify the clinical use of hybrid ASSR/DPOAE measurements with respect to time consumption.

D. Influence of L_1 on ASSR amplitude

The purpose of Experiment 3 was to investigate the suppression effect of the primary tone f_1 on the ASSR elicited by the amplitude modulated primary tone f_2. There was an impact of the f_1 tone on the ASSR amplitude at L_1 levels exceeding 66 dB SPL at $f_2=0.5$ kHz and exceeding 69 dB at $f_2=4$ kHz [see Figs. 8(a) and 8(b)]. It should be noted that the mean optimal primary tone level setting for $L_{1\ \mathrm{opt}}$ at $L_2=54$ dB SPL was 61 dB SPL. Thus, the small reduction in ASSR amplitude when using our hm DPOAE/ASSR stimulus setting (Experiment 2) can be attributed to the supressive effect of the primary tone f_1.

It should be emphasized that in Experiment 3, in contrast to Experiment 2, the f_1 tone was present during the whole ASSR recording time. Therefore, in Experiment 2 in which the f_1 tone was only present for approximately one-fifth of the ASSR recording time, the influence of the f_1 tone on the ASSR amplitude is considerably lower.

In view of a clinical application, the impact of the f_1 tone on the ASSR amplitude has to be investigated more thouroughly, especially, at low level stimulation where the level difference between the two primaries has to be increased for optimal DPOAE performance (scissor paradigm: $L_1=0.4\ L_2+39$, Kummer *et al.*, 2000). Also, hybrid measurements should be performed where the primary tone f_1, instead of the primary tone f_2, is amplitude modulated. Due to the fact that L_2 is considerably lower than L_1 when using the scissor paradigm, especially at close-to-threshold stimulation, the suppressive effect (impact of the unmodulated tone f_2 on the ASSR) is expected to be lower.

V. CONCLUSION

Findings of the present study suggest simultaneous ASSR and DPOAE measurements to be feasible without a notable loss in the measures' amplitude (2.6 dB for DPOAE, 1.8 dB for ASSR, on average) when using a modified DPOAE stimulus setting in which one primary tone is amplitude modulated for eliciting ASSRs. In compensating for the loss in amplitude and thus yielding sm test quality, recording time has to be increased. However, a longer recording time is only necessary at close-to-threshold responses where the SNR is low. The advantage of a DPOAE stimulus setting with an amplitude-modulated primary tone over a tone pair is that optimal stimulation and thus maximum response amplitudes can be reached—independent of test frequency for both ASSRs and DPOAEs. For example, in clinical practice, hybrid DPOAE/ASSR measurements can be an advantage in that ASSR measuring time is used for optimizing DPOAE stimulus settings for yielding high DPOAE performance.

ACKNOWLEDGMENTS

The work was supported by Deutsche Forschungsgemeinschaft, DFG Ja 597/6, Ja 597/8. Analog Devices made hard- and software evaluation platforms available. Many thanks to Roland Mühler for suggestions concerning the ASSR part.

Abdala, C. (**1996**). "Distortion product otoacoustic emission ($2f1$-$f2$) amplitude as a function of $f2/f1$ frequency ratio and primary tone level separation in human adults and neonates," J. Acoust. Soc. Am. **100**(6), 3726–3740.

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

Brown, D. K., Bowman, D. M., and Kimberley, B. P. (**2000**). "The effects of maturation and stimulus parameters on the optimal $f2/f1$ ratio of the $2f1$-$f2$ distortion product otoacoustic emission in neonates," Hear. Res. **145**(1–2), 17–24.

Cohen, L. T., Rickards, F. W., and Clark, G. M. (**1991**). "A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans," J. Acoust. Soc. Am. **90**, 2467–2479.

Dallos, P., and Cheatham, M. A. (**1971**). "Travel time in the cochlea and its determination from cochlear microphinic data," J. Acoust. Soc. Am. **49**, 1140–1143.

Dimitrijevic, A., John, M. S., van Roon, P., and Picton, T. W. (**2001**), "Human auditory steady-state responses to tones independently modulated in both frequency and amplitude," Ear Hear. **22**(2), 100–111.

Dobie, R. A., and Wilson, M. J. (**1993**). "Objective response detection in the frequency domain," Electroencephalogr. Clin. Neurophysiol. **88**(6), 516–524.

Galambos, R., Makeig, S., and Talmachoff, P. J. (**1981**). "A 40-Hz auditory potential recorded from the human scalp," Proc. Natl. Acad. Sci. U.S.A. **78**(4), 2643–2647.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, $2f1$-$f2$, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**(2), 821–839.

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.

Herdman, A. T., and Stapells, D. R. (**2001**). "Thresholds determined using the monotic and dichotic multiple auditory steady-state response technique in normal-hearing subjects," Scand. Audiol. **30**(1), 41–49.

Herdman, A. T., Picton, T. W., and Stapells, D. R. (**2002**). "Place specificity of multiple auditory steady-state responses." J. Acoust. Soc. Am. **112**(4), 1569–1582.

Janssen, T., Boege, P., von Mikusch-Buchberg, J., and Raczek, J. (**2005**). "Investigation of potential effects of cellular phones on human auditory function by means of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **117**, 1241–1247.

John, M. S., and Picton, T. W. (**2000**). "Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements," Hear. Res. **141**(1–2), 57–79.

John, M. S., Dimitrijevic, A., van Roon, P., and Picton, T. W. (**2001**). "Multiple auditory steady-state responses to AM and FM stimuli," Audiol. Neuro-Otol. **6**(1), 12–27.

John, M. S., Lins, O. G., Boucher, B. L., and Picton, T. W. (**1998**). "Multiple auditory steady-state responses (MASTER): Stimulus and recording parameters," Audiology **37**(2), 59–82.

Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). "Acoustic emission cochleography—practical aspects," Scand. Audiol. Suppl. **25**, 71–95.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). "Suppression tuning characteristics of the $2f_1$-f_2 distortion-product otoacoustic emission in humans," J. Acoust. Soc. Am. **98**, 197–210.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). "Optimal $L1$-$L2$ primary tone level separation remains independent of test frequency in humans," Hear. Res. **146**, 47–56.

Lasky, R. E. (**1998a**). "Distortion product otoacoustic emissions in human newborns and adults. (I): Frequency effects," J. Acoust. Soc. Am. **103**, 981–991.

Lasky, R. E. (**1998b**). "Distortion product otoacoustic emissions in human newborns and adults. (II): Level effects," J. Acoust. Soc. Am. **103**, 992–1000.

Pethe, J., Mühler, R., and von Specht, H. (**2001**). "Zur Abhängigkeit der amplitde modulation following response von der Vigilanz," HNO **49**, 188–193 (in German).

Picton, T. W., Vajsar, J., Rodriguez, R., and Campbell, K. B. (**1987**). "Reliability estimates for steady state evoked potentials," Electroencephalogr. Clin. Neurophysiol. **68**, 119–131.
Purcell, D. W., John, M. S., and Picton, T. W. (**2002**). "Concurrent measurement of distortion product otoacoustic emissions and auditory steady state evoked potentials," Hear. Res. **176**, 128–141.
Ross, B., Draganova, R., Picton, T. W., and Pantev, C. (**2003**). "Frequency specificity of 40-Hz auditory steady-state respones," Hear. Res. **186**, 57–68.
Stapells, D. R., Linden, D., Suffield, J., Hamel, G., and Picton, T. W. (**1984**). "Human auditory steady-state potentials," Ear Hear. **5**, 105–113.
Stapells, D. R., Makeig, S., and Galambos, R. (**1987**). "Auditory steady-state responses: Threshold prediction using phase coherence," Electroencephalogr. Clin. Neurophysiol. **67**, 260–270.
Valdes, J. L., Perez-Abalo, M. C., Martin, V., Savio, G., Sierra, C., Rodriguez, E., and Lins, O. (**1997**). "Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses," Ear Hear. **18**, 420–429.
Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Effects of ear canal standing waves on measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **98**, 3200–3214.
Zurek, P. M. (**1992**). "Detectability of transient and sinusoidal otoacoustic emissions," Ear Hear. **13**, 307–310.

Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans[a)]

Tiffany A. Johnson,[b)] Stephen T. Neely, Judy G. Kopun, and Michael P. Gorga
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 30 December 2005; revised 30 March 2006; accepted 3 April 2006)

Distortion product otoacoustic emission (DPOAE) fine structure has been attributed to the interaction of two cochlear-source mechanisms (distortion and reflection sources). A suppressor presented near the $2f_1$-f_2 frequency reduces the reflection-source contribution and, therefore, DPOAE fine structure. Optimal relationships between stimulus and suppressor conditions, however, have not been described. In this study, the relationship between suppressor level (L_3) and stimulus level (L_2) was evaluated to determine the L_3 that was most effective at reducing fine structure. Subjects were initially screened to find individuals who produced DPOAE fine structure. A difference in the prevalence of fine structure in two frequency intervals was observed. At 2 kHz, 11 of 12 subjects exhibited fine structure, as compared to 5 of 22 subjects at 4 kHz. Only subjects demonstrating fine structure participated in subsequent measurements. DPOAE responses were evaluated in 1/3-octave intervals centered at 2 or 4 kHz, with 4 subjects contributing data at each interval. Multiple L_3's were evaluated for each L_2, which ranged from 20 to 80 dB SPL. The results indicated that one or more L_3's at each L_2 were roughly equally effective at reducing DPOAE fine structure. However, no single L_3 was effective at all L_2's in every subject. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200048]

PACS number(s): 43.64.Jb, 43.64.Yp [BLM] Pages: 3896–3907

I. INTRODUCTION

There is thought to be more than one mechanism responsible for otoacoustic emission (OAE) generation. These include a coherent-reflection mechanism, a nonlinear-distortion mechanism, or a combination of the two (e.g., Shera and Guinan, 1999; Shera, 2004). Current thinking suggests that, as typically measured in the ear canal, distortion product OAEs (DPOAEs) include contributions from both a distortion source and a reflection source (e.g., Talmadge *et al.*, 1998, 1999; Mauermann *et al.*, 1999; Shera and Guinan, 1999; Stover *et al.*, 1999; Konrad-Martin *et al.*, 2001, 2002). These two sources have phases that rotate with frequency at rates that differ with respect to one another. The phase of the distortion source rotates slowly with frequency, while the phase of the reflection source rotates rapidly with frequency. The difference in rate of phase rotation for the two source mechanisms is believed to explain some of the patterns observed in DPOAE responses (such as fine structure).

When OAEs are recorded in small frequency steps, quasiperiodic patterns of maxima and minima are observed in the response (e.g., Heitmann *et al.*, 1998; Mauermann *et al.*, 1999; Kalluri and Shera, 2001; Goodman *et al.*, 2003). These patterns of alternating maxima and minima are referred to as OAE fine structure or OAE microstructure. In the case of DPOAEs, the fine structure observed in the response is believed to arise from the alternating constructive and destructive interference of the reflection source and the distortion source as a consequence of their different phase-rotation functions. Indeed, it has been shown that DPOAE fine structure is reduced or eliminated if the contribution from the reflection source is reduced or eliminated (Heitmann *et al.*, 1998; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004).

A number of investigators have speculated that the clinical test performance of DPOAEs is affected by the fine structure present in the response when the contributions from these disparate cochlear sources (and locations) interact (e.g., Heitmann *et al.*, 1998; Shera and Guinan 1999; Shaffer *et al.*, 2003; Mauermann and Kollmeier, 2004; Shera, 2004). According to this argument, the presence of maxima and minima in the DPOAE response may have an unknown influence on the overlap of the distributions of responses from normal and impaired ears. Additionally, because the location of DPOAE maxima and minima can shift in frequency as level is changed (He and Schmiedt, 1993, 1997; Heitmann *et al.*, 1996), the slope of a DPOAE input/output (I/O) function may be affected (Mauermann and Kollmeier, 2004). This could, in theory, affect the accuracy with which DPOAE I/O functions predict behavioral threshold using the approach originally described by Boege and Janssen (2002) and extended by others (e.g., Gorga *et al.*, 2003b; Oswald and Janssen, 2003).

There are two general methods for reducing or eliminating the reflection-source contribution to the DPOAE. One method is a signal-processing approach that involves computing an inverse fast Fourier transform (IFFT) of the DPOAE response and mathematically eliminating the long-latency components thought to arise from the reflection source (Kalluri and Shera, 2001; Knight and Kemp, 2001;

[a)]Portions of this work were presented at the 29th Midwinter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, February 2006.

[b)]Electonic mail: johnsonta@boystown.org

 0001-4966/2006/119(6)/3896/12/$22.50

Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004). Although this method, often called time windowing, has been shown to be effective at reducing fine structure in DPOAE responses, it requires making DPOAE measurements for many closely spaced frequencies. As a consequence, the windowing approach is time consuming, which makes it inappropriate for many clinical applications.

A second approach, selective suppression, involves presenting a suppressor tone near the $2f_1$-f_2 frequency in order to suppress the reflection-source contribution to the overall DPOAE. The results obtained with selective suppression and time windowing have been shown to be roughly equivalent, although incomplete reduction of fine structure has been observed with selective suppression (Kalluri and Shera, 2001). The selective-suppression approach is particularly well suited for clinical applications because it involves less test time than is needed to implement the windowing approach for removing fine structure. The suppressor tone is presented simultaneously with the two primaries whereas DPOAE measurements are made in the traditional manner. Unfortunately, the optimal suppressor level for the various stimulus conditions that might be used to make dichotomous decisions or to predict behavioral threshold from a DPOAE I/O function is not known. Several investigators have studied a range of suppressor levels, but have restricted their efforts to a limited range of stimulus levels (e.g., Heitmann *et al.*, 1998; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001; Dhar and Shaffer, 2004). A better understanding of the relationship between stimulus and suppressor levels for a wide range of stimulus conditions is needed if the clinical impact of restricting cochlear-source contribution is to be evaluated.

The purpose of the present study was to comprehensively explore, for a range of stimulus levels, suppressor levels that are effective at reducing DPOAE fine structure in subjects with normal hearing. The goal was to identify a set of optimal suppressor conditions which then will be used in a larger study that evaluates the clinical impact of reducing the reflection-source contribution to the DPOAE. Because the contribution of the reflection source is greatest at low stimulus levels and saturates as stimulus level increases (Stover *et al.*, 1996; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004), we hypothesized that the optimal suppressor level relative to the stimulus level would be higher at low stimulus levels and would grow at a slower rate than the stimulus level.

II. METHODS

A. Subjects

Twenty-two subjects with normal hearing were screened for participation in this project. Normal hearing was defined as behavioral thresholds $\leq$10 dB HL (re: ANSI, 1996) for the octave frequencies 0.25–8 kHz. Normal 226-Hz tympanograms were required just prior to each session in which DPOAE data were collected. The additional criteria for inclusion in the study are described below. Of the 22 subjects screened, complete data were obtained from 7 subjects. Three ears contributed data to the 2-kHz measures, 3 ears contributed data to the 4-kHz measures, and 1 ear contributed data at both 2 and 4 kHz.

B. Stimuli

All data were collected using custom-designed software (EMAV; Neely and Liu, 1993) that controlled a 24-bit soundcard (CardDeluxe, Digital Audio Labs) housed in a PC. This soundcard has a full-scale signal-to-noise ratio $\geq$110 dB and crosstalk (at 1 kHz) of −110 dB. With signal levels as high as 80 dB SPL and noise levels between −20 and −25 dB SPL, the measurements reported here were within the dynamic range of the soundcard. The two primary tones (f_1 and f_2) were generated on separate channels of the soundcard and were mixed acoustically in the ear canal. The suppressor tone (f_3) was generated on the same channel as f_2. Because f_3 and f_2 have a larger frequency separation than f_3 and f_1, generating f_3 on the same channel as f_2 ensured that intermodulation distortion products arising from the interaction of f_3 and f_2 did not occur in the same frequency bin as the DPOAE response ($2f_1$-f_2).

An ER-10C (Etymotic Research) probe microphone was used to present stimuli and record responses. This system had been modified to remove 20 dB of attenuation from each channel. This modification made it possible to achieve primary and suppressor levels as high as 80 dB SPL in each subject's ear canal.

DPOAE responses were recorded with f_2 varying in 1/64-octave steps over 1/3-octave intervals centered at either 2 or 4 kHz. The level of f_2 (L_2) varied from 20 to 80 dB SPL in 10-dB steps. The level of f_1 (L_1) and f_2/f_1 were set according to the following:

$$L_1 = 80 + 0.137 \log_2(18/f_2)(L_2 - 80),$$

$$f_2/f_1 = 1.22 + \log_2(9.6/f_2)(L_2/415)^2.$$

These equations have been shown to produce, on average, the largest DPOAE levels in normal-hearing human ears (Johnson *et al.*, 2006).

At each L_2, DPOAE responses were obtained for a control condition in which no attempt was made to suppress the reflection-source contribution (i.e., the suppressor was not presented). DPOAE responses were then recorded in the presence of a suppressor tone (f_3) whose frequency was 16 Hz below the $2f_1$-f_2 frequency. Presenting a suppressor tone close to the $2f_1$-f_2 frequency reduces or eliminates the contribution of the reflection source to the DPOAE (e.g., Heitmann *et al.*, 1998; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001). However, because the level of f_3 required to suppress the reflection component across a range of L_2 is not known, a variety of suppressor levels (L_3) were evaluated.

In all cases, L_3 was set according to the following:

$$L_3 = 0.75L_2 + C.$$

Setting L_3 according to this equation with C=20 dB results in at least 15 dB of DPOAE suppression when a suppressor is presented within 100 Hz of f_2 (Gorga *et al.* 2002,

2003a). When L_2=20 or 30 dB SPL, C values ranged from 0 to 45 dB. For higher L_2 levels, C values ranged from 0 to 40 dB, with the constraint that L_3 never exceeded 80 dB SPL. In both cases, C was incremented in 5-dB steps. By evaluating this set of L_3, we expected to cover conditions ranging from incomplete suppression of the reflection component to over suppression, meaning that the distortion component was also suppressed.

C. Procedures

Prior to data collection, stimuli were calibrated in sound pressure level at the plane of the probe. Although concerns have been raised regarding in-the-ear pressure calibration (e.g., Siegel, 1994; Siegel and Hirohata, 1994; Neely and Gorga, 1998), similar concerns exist for other methods of calibrating stimulus level. Many of these concerns are eliminated if stimuli are calibrated in acoustic intensity (Neely and Gorga, 1998). Unfortunately, acoustic-intensity calibration has not been widely used and *in situ* pressure calibration is the current standard of practice for DPOAE measures. However, we recognize that this calibration approach may introduce errors for frequencies at and above 4 kHz.

During data collection, measurement-based stopping rules were used such that averaging continued until either the noise floor reached −25 dB SPL or 32 s of artifact-free averaging had elapsed. Both DPOAE and noise levels were estimated from the $2f_1$-f_2 frequency bin. This was accomplished during data collection by alternately storing 0.25 -sec samples of data in one of two buffers. DPOAE level was estimated by summing the contents of the two buffers. Noise level was estimated by subtracting their contents. The use of these stopping rules resulted in noise levels below −20 dB SPL for the majority of subjects and conditions tested.

All data were collected in a sound-treated room. Subjects were seated comfortably in a reclining chair during data collection and either read quietly or slept. Data collection for the 2-kHz conditions took between 7 and 12 two-hour sessions. Data collection for the 4-kHz conditions took between 3 and 7 two-hour sessions. Data-collection time was shorter for the 4-kHz conditions as a result of the lower noise levels in this frequency interval. More averaging was required to produce similar noise levels for the 2-kHz conditions

D. Analyses

DPOAE level and phase at each point in the fine-structure measure were converted to complex values. Each of these values was then placed in the appropriate frequency position within an 8192-point array of zeros. An IFFT was computed on the resulting array of real and imaginary numbers. The time window for the analysis is dependent on the step size used to collect the fine-structure measures. The step size for the measurements around 2 kHz was approximately 21 Hz, and a 20-ms window was used for the analysis. The step size for the measurements around 4 kHz was approximately 43 Hz, and a time window of 10 ms was used for the analysis at this frequency. The resulting amplitude values as a function of time were then used to estimate the magnitude of the early (distortion source) and late (reflection source) components in the response. See Konrad-Martin *et al.* (2001) for additional details regarding this procedure.

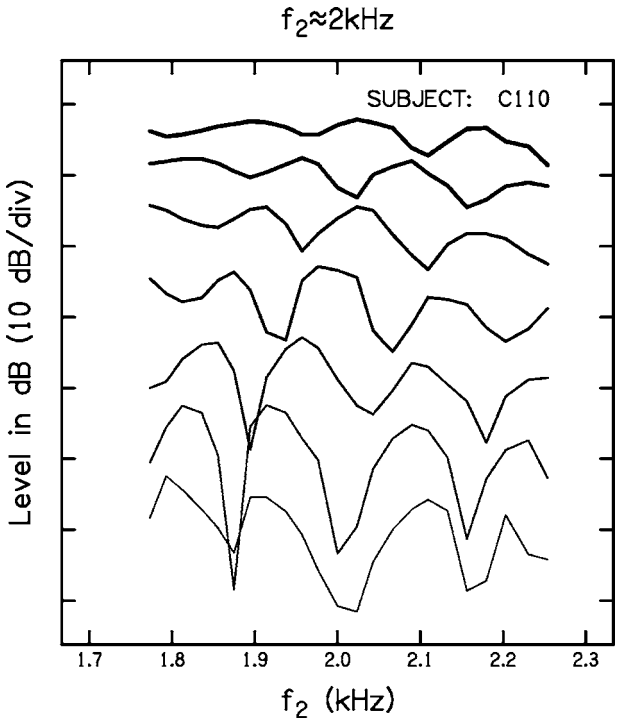


FIG. 1. DPOAE level (dB SPL) as a function of f_2 (kHz) for a single subject for the 1/3-octave interval surrounding f_2=2 kHz. Increasing line thickness indicates increasing L_2, which ranged from 20 to 80 dB SPL in 10-dB increments. Each trace has been offset by successive 10-dB increments to aid in visualization.

In addition to the IFFT analysis, a discrete cosine transform (DCT) (Rao and Yip, 1990) was computed on the DPOAE fine-structure measures. The DCT, which is similar in concept to the FFT, was applied to the log-magnitude (dB) data and had the advantage of closer agreement with visual assessment of fine structure. It provides a quantification of the overall level of the response (the DC component), as well as the levels of higher "frequency" components (measured in cycles/octave) that ride on the DC component. The coefficients of these higher frequency components provide a means to quantify the magnitude of the fine structure.

III. RESULTS

A. Fine structure in DPOAE responses

Figure 1 plots an example of fine structure in DPOAE responses for one subject (C110) when measured in a 1/3-octave interval surrounding f_2=2 kHz. DPOAE responses with L_2 ranging from 20 to 80 dB SPL in 10-dB steps are shown with increasing line thickness. Maxima-to-minima depths exceeding 15–20 dB are observed for low levels. As has been reported previously (e.g., He and Schmiedt, 1993, 1997; Mauermann and Kollmeier, 2004), the depth of the fine structure decreases as L_2 increases, and the frequency at which a given minima occurs shifts with L_2. Data from other subjects were similar to those shown in Fig. 1, at least at 2 kHz.

When screening subjects for participation, DPOAE fine-structure measures were recorded in 1/3-octave intervals

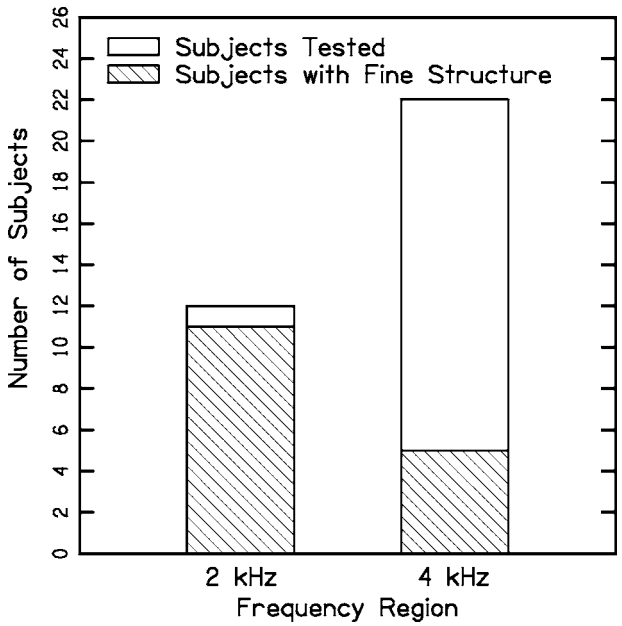


FIG. 2. The number of subjects screened in each frequency interval who produced DPOAE fine structure, which was defined as a maxima-to-minima ratio $\geqslant$10 dB in one cycle of fine structure when L_2=30 dB SPL. The open bars represent the number of subjects evaluated and the cross-hatched region indicates the number who had DPOAE fine structure.

surrounding f_2=2 and/or 4 kHz with L_2=30 dB SPL. In order to participate, subjects were required to have at least a 10-dB difference between maxima and minima in one cycle of the fine structure within the 1/3-octave interval at this L_2. Figure 2 summarizes the prevalence of fine structure in the 2- and 4-kHz intervals. The open bars represent the number of subjects evaluated in each frequency interval and the cross-hatched region indicates the number of subjects meeting the above inclusion criterion. Of 12 subjects screened in the 2-kHz interval, 11 (92%) met the 10-dB maxima-to-minima requirement. In contrast, only 5 of 22 subjects (23%) met the requirement at 4 kHz.

Each of the 12 people evaluated at 2 kHz were also evaluated at 4 kHz. Data demonstrating the difference in microstructure patterns for these subjects in the intervals around 2 and 4 kHz are shown in Fig. 3. Results for the 2-kHz interval are shown in the first and third columns, with results for the 4-kHz interval shown in the second and fourth columns. Adjacent panels within a row represent data from the same subject, as indicated by the subject code in the upper-right-hand corner of each panel. All DPOAE data shown in Fig. 3 are for L_2=30 dB SPL. At 2 kHz, all but one subject (C106) had fine structure with a maxima-to-minima ratio of 10 dB or greater. In these same subjects, only three (C103, C109, and C111) met this definition for the presence of fine structure at 4 kHz.

The data shown in Fig. 3 were transformed to the time domain by computing an IFFT of the DPOAE responses. This technique has been used previously (e.g., Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001, 2002; Mauermann and Kollmeier, 2004) as a method for exploring cochlear-source contribution to the DPOAE recorded in the human ear canal. Figure 4 plots an IFFT of the data shown in Fig. 3, using the same general layout. Columns 1 and 3 represent data for the 2-kHz interval and columns 2 and 4 represent data for the 4-kHz interval. Adjacent panels within a row represent results at 2 and 4 kHz for the same subject. The early components (those occurring near time=0 ms) are believed to arise from the distortion-source mechanism (near the f_2 place), with the later components arising from the reflection-source mechanism (at the $2f_1$-f_2 place).

In general, at 4 kHz, the amplitude of the later components (reflection-source components) in Fig. 4 is reduced relative to the amplitude of the early components (distortion-source components) when compared to 2 kHz. This is consistent with the observation in Figs. 2 and 3 that the prevalence of fine structure at 4 kHz is less than the prevalence of fine structure at 2 kHz. Recall that in the case of DPOAEs, fine structure in the response is believed to arise from the constructive and destructive interference of the two source components (e.g., Heitmann *et al.*, 1998; Kalluri and Shera, 2001; Konrad-Martin *et al.*, 2001; Mauerman and Kollmeier, 2004). The reduced amplitude of the reflection-source component at 4 kHz would, therefore, be expected to produce less fine structure at 4 kHz.

B. Effect of a suppressor on DPOAE fine structure in individual ears

Figure 5 shows the effects of a suppressor on DPOAE fine structure in a single subject. The left column plots DPOAE responses for the control condition (no f_3) and for several suppressor conditions. The DPOAE levels for the control (no f_3) and two suppressor conditions are shown in each panel, and suppressor level increases as one moves from the top to bottom panel. The suppression condition is indicated by the value of C used for the given recording. Recall that L_3 was set according to $L_3=0.75L_2+C$, so that L_3 increases as C increases. L_2=30 dB SPL for all data shown in Fig. 5. As can be seen in Fig. 5, the lowest suppressor levels have little influence on the depth of the fine structure. However, as suppressor level is increased, the difference between maxima and minima decreases. Fine-structure depth reaches a minimum when C is between approximately 20 and 35 dB. Additionally, for these suppressor levels (C =20–35 dB), the overall level of the DPOAE has not been reduced, i.e., the mean DPOAE level across all f_2 conditions is relatively constant, despite the apparent reduction in the differences between minima and maxima. Evidence for this conclusion is provided by the constancy (and, therefore, overlap) of the level of the DC component in the top four panels in the left column. This suggests that these suppressor conditions are effective for eliminating the reflection-source component, but do not suppress the distortion-source component. In contrast, the highest suppressor levels (C=40 or 45 dB) result in a reduction of the overall level of the DPOAE, suggesting that for these conditions the distortion-source component has been suppressed in addition to the reflection-source component.

In order to quantify the magnitude of the overall DPOAE level and the fine structure in control and suppressor conditions, a DCT of the data shown in the left column of

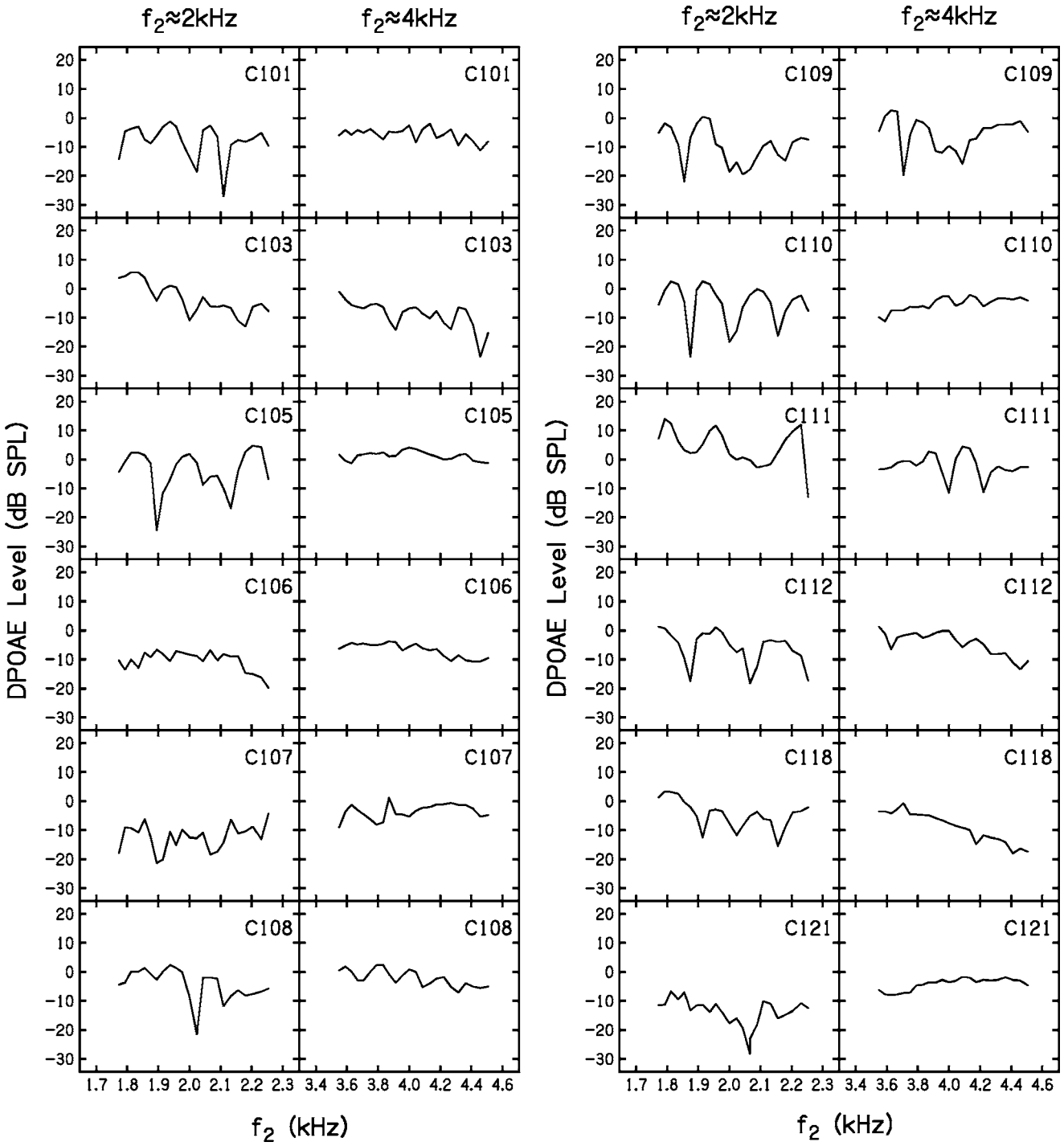


FIG. 3. DPOAE level (dB SPL) as a function of f_2 (kHz) for all subjects screened in both the 2- and 4-kHz intervals. Data for $f_2 \approx 2$ kHz are shown in the first and third columns. Data for $f_2 \approx 4$ kHz are shown in the second and fourth columns. Adjacent panels within a row represent data from the same subject, as identified in the upper-right corner of each panel. All data were collected with $L_2 = 30$ dB SPL in the absence of a suppressor tone.

Fig. 5 was computed. The DCT provides a way to quantify the overall level of the DPOAE response, which is the DC component shown as the unattached symbols on the far right side of each panel in the left column. The DCT also quantifies the level of the coefficients corresponding to higher frequency constituents of the DPOAE response, which correspond to the fine structure that rides on the overall or DC response in the DPgram. The level of these higher frequency components for each suppressor condition is shown in the middle column in Fig. 5. As expected, the level of these DCT coefficients is greatest when the fine structure is the largest, which occurs in the control condition and/or for the lowest suppressor levels. The level of these DCT coefficients is reduced to a minimum for suppressor conditions corresponding to $C = 20-35$ dB, with little or no reduction in the DC component level (as stated earlier). Stated another way, these conditions resulted in maximum suppression of fine structure with no effect on overall level. This illustrates how DCT analyses can be used to assess contributions from distortion and reflection sources. Suppressor conditions corresponding

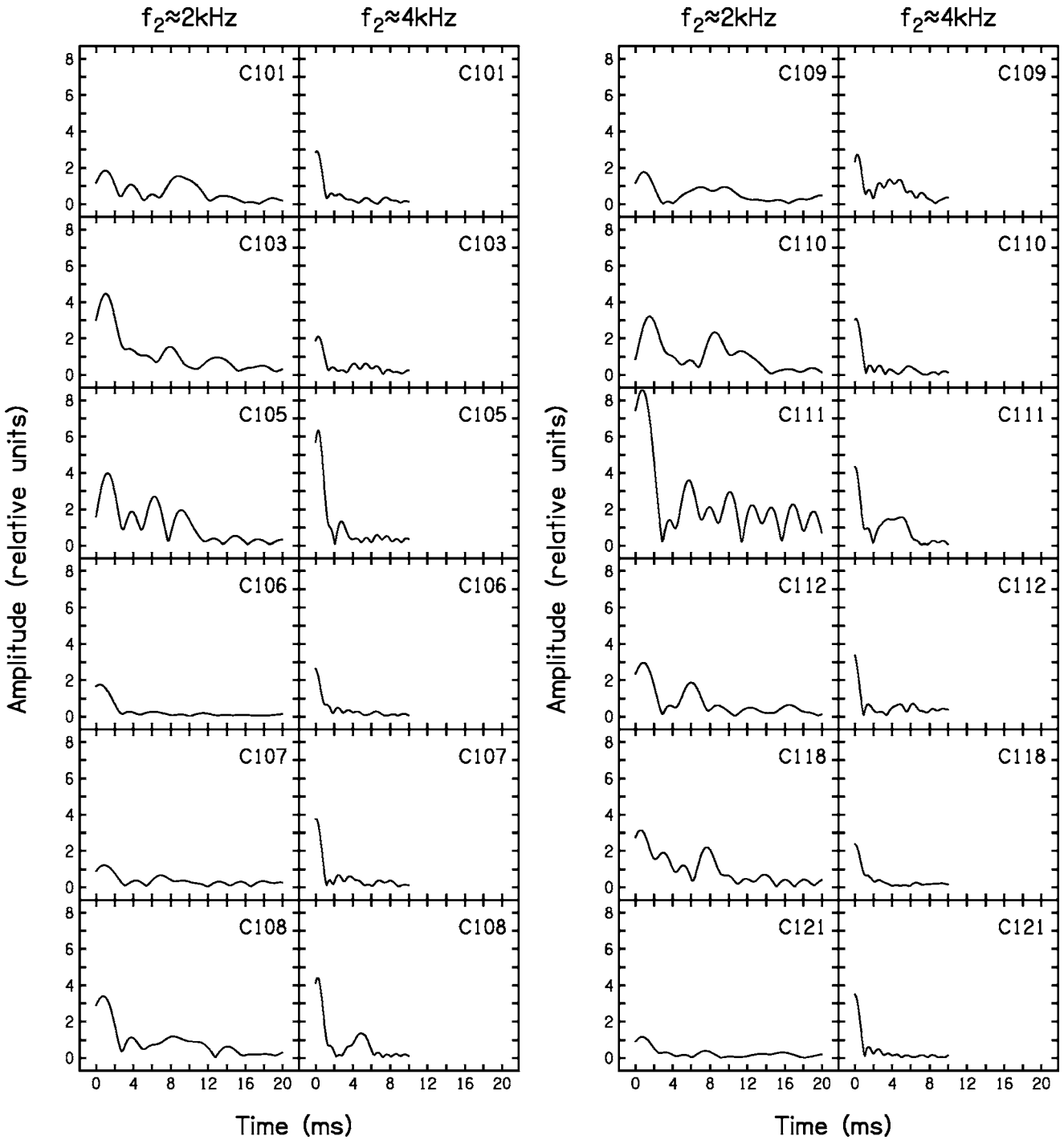


FIG. 4. Inverse fast Fourier transforms (IFFTs) of the data shown in Fig. 3. The data are plotted as amplitude (relative units) as a function of time (ms). The layout is the same as in Fig. 3. Data for $f_2 \approx 2$ kHz are shown in the first and third columns, $f_2 \approx 4$ kHz are shown in the second and fourth columns, adjacent panels within a row are for the same subject.

to $C=40$ and 45 dB are associated with an apparent increase in the higher frequency DCT coefficients, but this is more a consequence of a reduction in the DC level as the DPOAE is suppressed toward the noise floor and fluctuations in the noise exert a larger influence on the measured response. In Fig. 5, the noise floor was typically less than −20 dB SPL.

The data shown as a DPgram in the left column of Fig. 5 are shown as an IFFT in the right column. This analysis provides an alternate means for evaluating the effect of the suppressor on the DPOAE. As expected from the results displayed in the left and middle columns of Fig. 5, the lowest suppressor levels have little effect on either the early or late components in the DPOAE response. Suppressors corresponding to $C=20-35$ dB have relatively little suppressive effect on the early components of the DPOAE, while suppressing the late components to a level close to the noise floor. Suppressors corresponding to $C=40$ and 45 dB have the undesirable effect of suppressing not only the late components (the reflection-source components) but also the early components (the distortion-source components). The similar

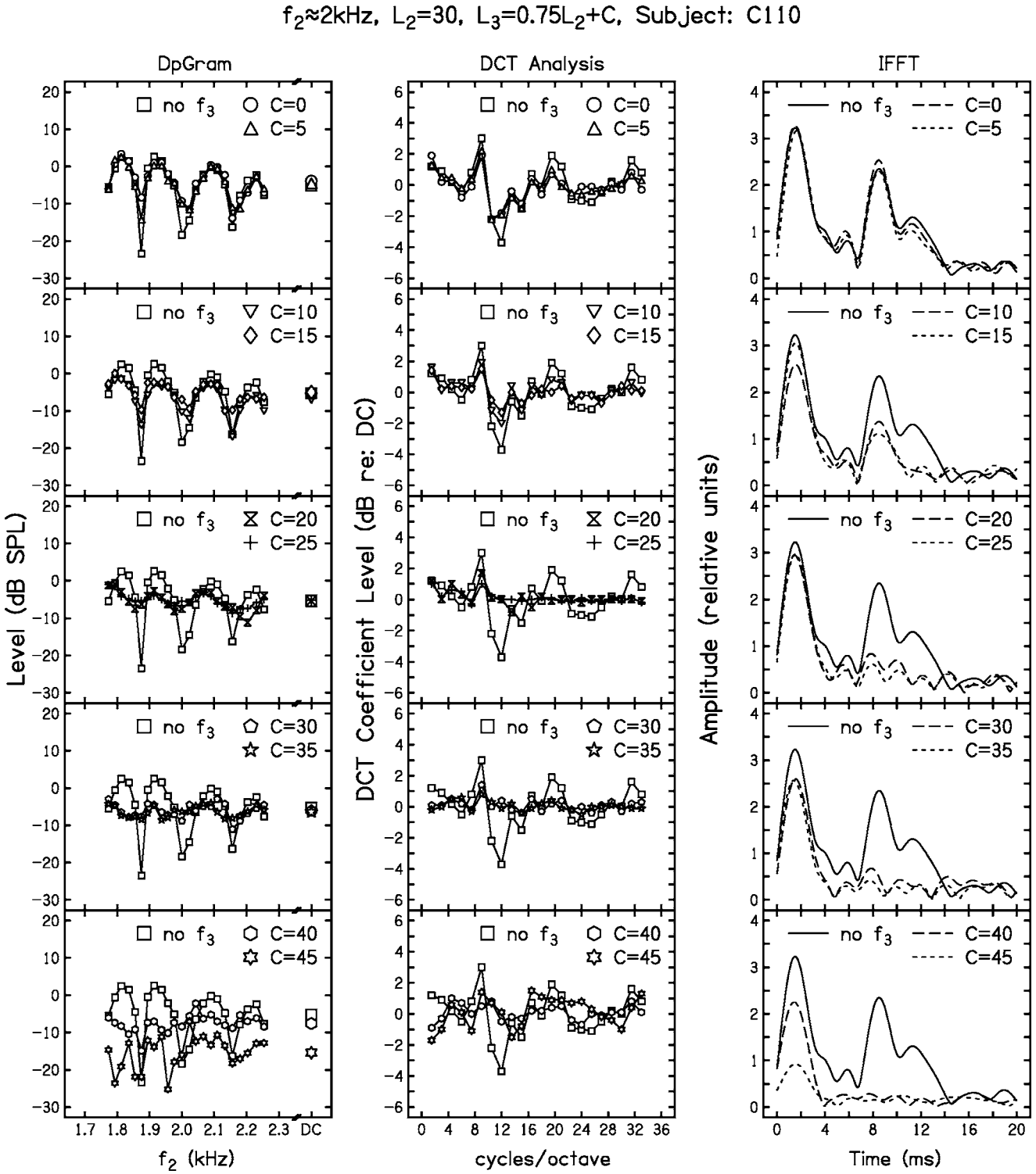


FIG. 5. Summary of the analyses completed on the data. The data shown are for a single subject (C110), with L_2=30 dB SPL, and $f_2 \approx 2$ kHz. DPOAE responses are shown in the left column along with the DC component (overall DPOAE level) from the discrete cosine transform (DCT) of the data (see the text for details regarding the DCT). The level of the higher "frequency" components of the DCT for the same data are shown in the middle column, and an IFFT of the data in the left column is shown in the right column. In each row, the response in the control condition (no f_3) along with results for two suppressor levels (indicated by C value) is shown. Suppressor level increases from top to bottom.

conclusions drawn from both analyses shown in Fig. 5 suggest that the new DCT analysis can be used to quantify the influence of the suppressor on the fine structure in DPOAE responses.

A summary of the data shown in Fig. 5 (subject C110, L_2=30 dB SPL and f_2 in the interval surrounding 2 kHz) is shown in the left column of Fig. 6. Similar data from a different subject (C117) for the interval surrounding 4 kHz when L_2=30 dB SPL are shown in the right column. The upper row of Fig. 6 shows the level of the DC component (corresponding to the overall DPOAE level) as a function of the suppressor condition, which ranges from the control con-

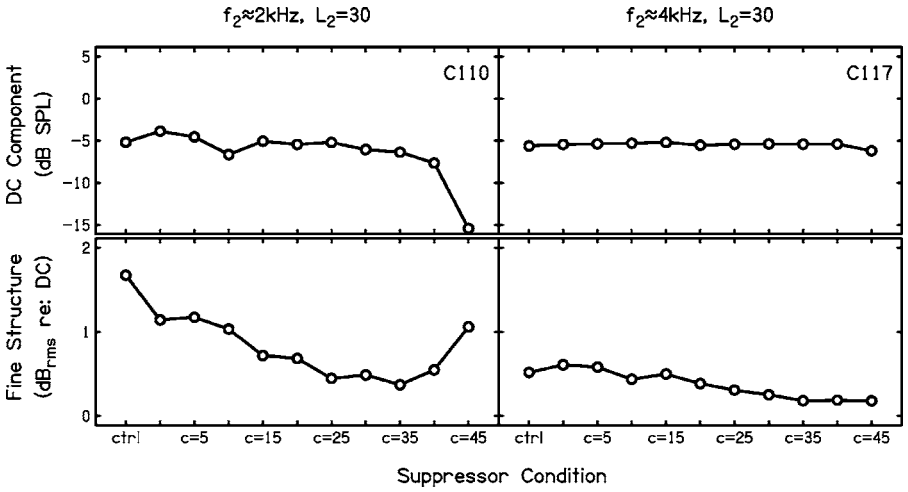


FIG. 6. The left column plots a summary of the discrete cosine transform (DCT) of the data shown in Fig. 5. The right column plots similar data from a different subject (C117) for the interval surrounding 4 kHz when L_2 =30 dB SPL. The top row plots the DC component level, which corresponds to the overall DPOAE level, as a function of suppressor condition (C). The bottom row represents the rms level of the higher "frequency" components of the DCT (see middle column, Fig. 5) and represents the magnitude of the fine structure in the DPOAE response for the various suppressor conditions. See the text for additional information regarding the derivation of these data.

dition (no suppressor) to the highest suppressor level (L_3 =67.5 dB SPL when L_2=30 dB SPL and C=45 dB). The bottom row plots the level of the fine structure relative to the DC component level for each suppressor condition. This measure for the fine structure was obtained by computing the rms level of the coefficients between 1.5 and 22.5 cycles per octave. These components were chosen because pilot data showed them to be the primary components contributing to the fine-structure pattern. The two cases shown in this figure were chosen to demonstrate extreme conditions, one in which it was easy to select an optimal suppressor and one in which the optimal suppressor condition was not obvious.

As expected from the data shown in Fig. 5, the data shown in the left column of Fig. 6 indicate that the DC level is relatively constant through approximately C=35 dB, while the fine structure reaches a minimum at this suppressor level. This suggests that C=35 dB may be an optimum suppressor condition for this subject and this set of stimulus conditions because it minimizes the fine structure but does not reduce the overall level of the response. The data shown in the right column describe results in which the choice of an optimal suppressor level was more ambiguous. For this subject (C117), the depth of the fine structure is less than was observed for the subject shown in the left column (compare the dB_{rms} level of the fine structure in the control condition for the two subjects). As a result, the effect of the suppressor on the fine structure is less apparent. In the right column, the DC or overall level of the DPOAE is essentially constant for all suppressor levels, with only a small reduction when C =45 dB. The magnitude of the fine structure reaches a minimum between C=35 and 45 dB. These data suggest that suppressor levels corresponding to C values between 35 and 45 dB are roughly equivalent and any in this range could be used to reduce fine structure. Additionally, the suppressors had less of an effect in this subject than for the subject whose data are shown in the left column, primarily because this subject (C117) has less fine structure in the control condition. The data from the two subjects in Fig. 6 demonstrate the variability across subjects in terms of the depth of fine structure and the effect of the suppressor on the fine structure.

C. Effect of a suppressor on DPOAE fine structure in a group of subjects

Figure 6 illustrated the effect of a suppressor in two individual ears for a limited set of stimulus conditions. However, the goal of the present study was to identify a set of optimal suppressor conditions across a range of stimulus conditions that could be applied to future studies that evaluate the effect of reducing the reflection-source contribution to the DPOAE in a larger group of subjects. Data addressing this question are shown in Fig. 7.

Figure 7 summarizes the mean DC and fine-structure levels across the four subjects from whom detailed data were collected. Error bars correspond to ±1 standard deviation. Results for the 2- and 4-kHz intervals are shown in the left- and right-hand columns, respectively. Data for L_2=20, 40, 60, and 80 dB SPL are shown (although not shown, data for intermediate L_2 levels were similar). For each L_2, two panels are shown. The upper panel corresponds to the DC level and is labeled "dB SPL." The lower panel in each pair corresponds to the rms level of the fine structure (as described for Fig. 6) and is labeled "dB_{rms}."

Data such as these could be used to choose suppressor levels for use in evaluating the effect of reducing the reflection-source contribution on DPOAE test performance, including dichotomous decisions about auditory status and predictions of behavioral thresholds. Optimal suppressor level for this purpose is defined as the highest suppressor level at which there was little or no reduction in the DC (overall) level, but there was a corresponding minimum in the rms level of the fine structure. According to this definition of optimal suppressor level, when f_2=4 kHz and L_2 =40 dB SPL, a suppressor corresponding to C=30 dB would be identified as optimal. In general, for both the 2- and 4-kHz intervals, the optimal suppressor level (L_3=0.75L_2 +C) decreases as L_2 increases. This approach to identifying optimal suppressor levels for use in future studies has the advantage of demonstrating the effect the suppressor had on the DC and rms levels of the response. It does not, however, provide information regarding how much the optimal sup-

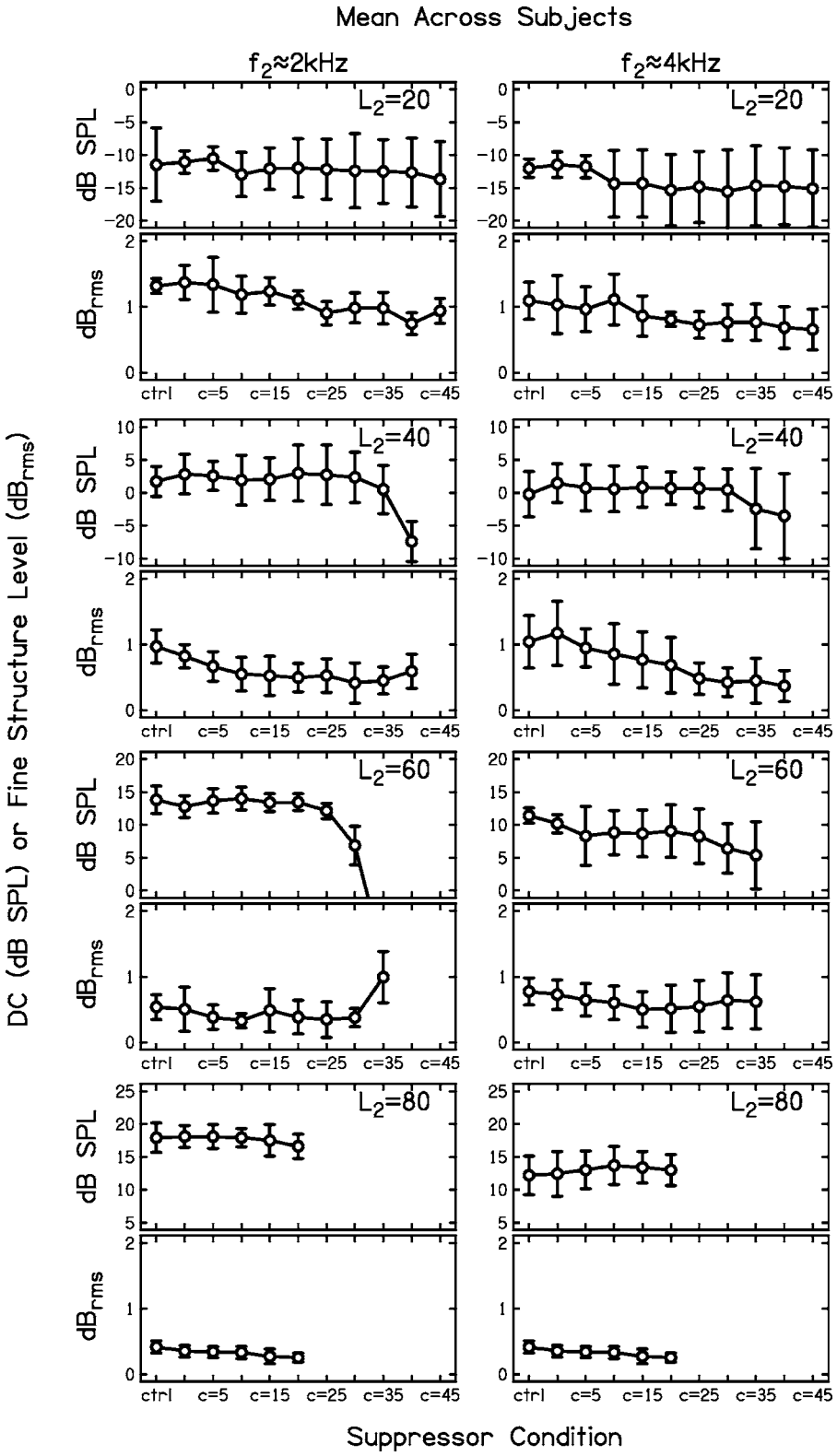


FIG. 7. Summary of the mean (±1 standard deviation) DCT results across the 4 subjects. The left column plots data for the 2-kHz interval, with data for the 4-kHz interval shown in the right column. Data for L_2=20, 40, 60, and 80 dB SPL are shown, as indicated. For each L_2, the upper panel plots DC level (dB SPL) as a function of suppressor condition and the lower panel plots the rms level (dB_{rms}) of the fine structure. See the text for additional information regarding the derivation of these data.

pressor level varied across individual subjects. For example, it blurs the differences in patterns of response across subjects, such as those shown in Fig. 6.

Figure 8 describes intersubject variability in optimal-suppressor condition as well as the relationship between optimal-suppressor level and stimulus level. Instead of using data plotted in the form shown in Figs. 6 and 7, these plots could be used as an alternate method for identifying optimal suppressor levels for future use. In Fig. 8, the optimal suppressor (as defined previously) is plotted for each subject as a function of L_2. Data for $f_2 \approx 2$ kHz are shown in the left column, with data corresponding to $f_2 \approx 4$ kHz shown in the right column. The top row plots C value as a function of L_2, the middle row plots suppressor level ($L_3 = 0.75L_2 + C$) as a function of L_2, and the bottom row plots suppressor level relative to L_2. The sloping lines represent linear fits to these

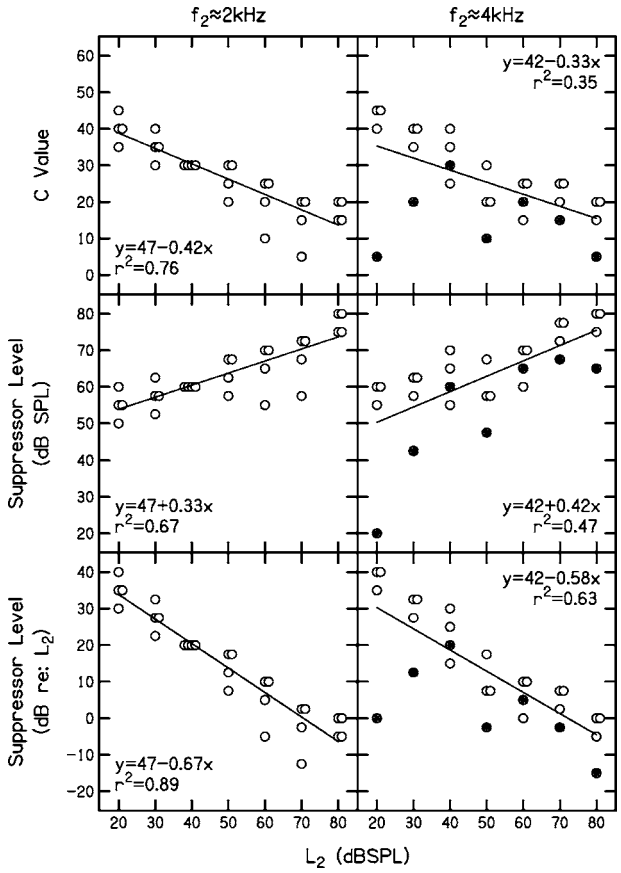


FIG. 8. The relationship between suppressor level and L_2 in individual subjects. Data for the 2-kHz interval are shown in the left column and data for the 4-kHz interval are shown in the right column. Each row represents an alternate view of suppressor level (L_3). The top row plots C ($L_3=0.75L_2+C$) as a function of L_2. The middle row plots absolute suppressor level (dB SPL) as a function of L_2, with the bottom row plotting suppressor level re: L_2 as a function of L_2. The solid slanting lines represent linear fits to these data; the fits are shown as insets in each panel. The r^2 values associated with these fits are also shown in each panel. The filled circles in the right column highlight data from an individual subject, as indicated in the text.

data, with the fits shown as insets in each panel. The r^2 values associated with these fits are also shown in each panel. The filled circles in the right column indicate results for an individual subject.

Consistent with the data shown in Fig. 7, the optimal value of C (upper panel) decreases with L_2 and optimal suppressor level (middle panel) grows at a rate slower than L_2 such that the optimal suppressor is much higher than L_2 for low L_2 levels and is nearly equal to L_2 for high levels (bottom panel). Although there is variability across subjects in the optimal suppressor levels, most subjects showed the same general trend of decreasing suppressor level relative to L_2 as L_2 increased. The optimal-suppressor levels across subjects were typically within 10–15 dB. An exception to this is the subject whose data are highlighted (filled circles) in the right column of Fig. 8. This subject typically required suppressor levels lower than the other subjects and did not show a consistent pattern of decreasing C value as L_2 increased.

The r^2 values associated with the linear fits to these data indicate that between 35% and 89% of the variance is accounted for, with higher r^2 values at 2 kHz, as compared to 4 kHz. Although these r^2 values indicate reasonable fits to the data, the variability in optimal suppressor level across individual subjects suggests that it may be difficult to choose a single suppressor level that is optimal across all subjects. Even after excluding the data from the subject whose response patterns differed from the pattern seen in other subjects, the optimal suppressor levels covered a range of 10–15 dB at nearly every L_2 (the exception being the suppressor level when $f_2 \approx 2$ kHz and $L_2=40$ dB SPL).

IV. DISCUSSION

The results described above summarize an exploration of the suppressor level that is effective at reducing fine structure in DPOAE responses for a range of stimulus levels. As reported by others, these results suggest that it is possible to reduce or eliminate fine structure in DPOAE responses by playing a suppressor tone close to $2f_1$-f_2. A general trend was observed, such that the suppressor level (relative to stimulus level) needed to eliminate fine structure decreases as stimulus level increases. There is, however, variability across subjects in the suppressor level that is most effective at reducing fine structure. Additionally, the prevalence of fine structure was frequency dependent, occurring more frequently around 2 kHz, as compared to 4 kHz.

A. Differences in fine structure across frequency

The mechanism responsible for the relative lack of fine structure at 4 kHz, as compared to 2 kHz, is unclear, although this observation has been reported previously. Dhar and Shaffer (2004) observed that, in a group of 20 subjects, 60% showed little or no fine structure above 3 kHz. This percentage is slightly lower than observed for the present study where 77% of subjects evaluated at 4 kHz, failed to meet our inclusion criteria for depth of fine structure.

The data shown in Fig. 4 indicate that there is a relatively smaller contribution to the DPOAE from later (reflection-source) components at 4 kHz as compared to 2 kHz, which, in turn, would be expected to produce less fine structure at 4 kHz, as compared to 2 kHz. A similar observation was reported in a more quantitative fashion by Konrad-Martin *et al.* (2001). Although the measurement paradigm used by Konrad-Martin *et al.* was different than the present study (Konrad-Martin *et al.* used a fixed f_2 paradigm), they conducted a similar IFFT analysis and quantified the magnitude of the early and late components in the response. The magnitude of the early components, which are believed to correspond to the distortion-source components, was greater than the magnitude of the late components, which are believed to correspond to the reflection-source components.

Even if the previous observation regarding the relative magnitude of early and late components is correct, an explanation for why the reflection source might contribute less to the DPOAE at 4 kHz as compared to 2 kHz is lacking. The modeling work of Talmadge *et al.* (1998) suggests that reflections within the cochlea at the stapes footplate (termed basal reflectance in the model) play a primary role in producing the fine structure observed in OAE responses, including DPOAEs. The magnitude of the basal reflectance varies

with frequency and could potentially play a role in producing differences across frequency. In addition to basal reflectance, the Talmadge *et al.* model suggests two other terms that play a role in producing DPOAE fine structure. These include apical reflectance, which essentially corresponds to the reflection source, and the wave-number ratio. The latter term does not have a simple physical interpretation. However, if the Talmadge *et al.* model is correct, the basal-reflectance term and the wave-number-ratio term have phases that vary slowly with respect to one another. In certain frequency regions, they would be expected to cancel out, with the result being reduced or absent fine structure in that frequency region. It is unclear whether this prediction of the model is correct, but it is consistent with the observation of limited fine structure in the vicinity of 4 kHz in present study and the Dhar and Shaffer (2004) results.

It is also possible that our choice of stimulus parameters played a role in the relative lack of fine structure at 4 kHz. Konrad-Martin *et al.* (2001) showed that changes in stimulus parameters could affect the relative energy of early and late components in the DPOAE response. When DPOAE responses were recorded using the L_1, L_2 relationship recommended by Kummer *et al.* (1998), which was thought to optimize the DPOAE level (as compared to a constant 10-dB difference between L_1 and L_2), Konrad-Martin *et al.* showed that the late components remained essentially constant in level whereas the early components increased in level. The stimulus parameters in the present study produce larger DPOAE levels in normal-hearing subjects (Neely *et al.*, 2005; Johnson *et al.*, 2006) than those recorded using the stimulus parameters described by Kummer *et al.* (1998). It is possible that our stimulus parameters helped to reduce fine structure by increasing the level of the early components relative to the level of the later components. This hypothesis, however, was not tested. Additionally, it is unclear why changing stimulus parameters would have a differential effect on fine structure at 4 kHz as compared to 2 kHz, particularly since the two regions are close together in relation to the entire length of the basilar membrane.

B. Suppressor levels relative to L_2

Consistent with our hypotheses at the outset, the data shown in Figs. 7 and 8 suggest that for low L_2 levels, the suppressor needs to be much higher than L_2 in order to reduce DPOAE fine structure. However, the optimal suppressor level increases at a slower rate than L_2 so that when $L_2 = 80$ dB SPL, suppressors equal to or less than L_2 result in maximum suppression of fine structure with little or no effect on overall DPOAE level. This trend of higher suppressor levels (re: L_2) for low L_2 levels, as compared to high L_2 levels, is consistent with expected differences in the relative contributions from reflection and distortion sources as level is increased. The results from a number of investigators (e.g., Stover *et al.*, 1996; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004) suggest that the reflection-source contribution, relative to the contribution from the distortion source, is larger at low stimulus levels and saturates as level is increased, such that the contribution from the reflection source is reduced for high stimulus levels.

C. Suppressing DPOAE fine structure and clinical test performance

A number of investigators have speculated that the multiple sources contributing to DPOAEs and the resulting fine structure may have a negative impact on the clinical test performance of DPOAEs (e.g., Heitmann *et al.*, 1998; Shera and Guinan 1999; Shaffer *et al.*, 2003; Mauermann and Kollmeier, 2004; Shera, 2004). However, no data have been reported that directly test this hypothesis. Dhar and Shaffer (2004) reported a general lack of statistically significant correlations between DPOAE level (for moderate-level stimuli) and behavioral threshold in a small group of normal-hearing subjects, even when the reflection source was eliminated. Although the Dhar and Shaffer data cast doubt on the view that controlling source contribution will improve DPOAE test performance, additional work with a larger group of subjects with both normal and impaired hearing is needed. However, before such studies can be completed, the relationship between stimulus level and optimal suppressor level needs to be understood. The data from the present study demonstrate that it is possible to reduce or eliminate fine structure from DPOAE measurements without also affecting overall DPOAE level. In general, the trend of decreasing optimal suppressor level (re: L_2) as L_2 increased was consistent across subjects. Although there was variability in the optimal suppressor level across subjects (typically with a range of 10–15 dB at each L_2), for many subjects, several suppressor levels were nearly equivalent in their ability to reduce fine structure (see, e.g., Figs. 5 and 6). Although this result is encouraging, one subject (see highlighted points at 4 kHz in Fig. 8) responded differently to the suppressor tone than the other subjects. In this subject, lower and more variable suppressor levels were deemed optimal. Presenting higher level suppressors in this subject's ear typically resulted in reduction in the overall level of the DPOAE. If a large number of ears react in this manner, then it might actually be the case that test performance will be poorer if one attempted to suppress the reflection source. Furthermore, the data shown in Figs. 2–4 call into question the extent to which uncontrolled source contribution limits test performance when $f_2 \approx 4$ kHz. For this frequency interval, fine structure was not present in the majority of ears and, when it was present, the reflection-source contribution was reduced compared to the 2-kHz interval. This latter observation suggests that even if fine structure was present, its impact would be minimal, especially for stimulus levels in routine clinical use.

In summary, by evaluating a range of suppressor levels, it was possible to reduce fine structure in DPOAE responses. The time involved in collecting data such as those in the present study, however, make it impossible to identify an optimal suppressor level for each individual ear under clinical conditions. Under clinical conditions, it is necessary to have *a priori* knowledge of the optimal suppressor level for a given stimulus condition. The data shown in Fig. 7 or the linear equations shown in Fig. 8 could be used to choose

suppressor conditions when evaluating DPOAE test performance when the reflection-source contribution is reduced or eliminated. Although similar trends were observed across most subjects, the intersubject variability in optimal suppressor level suggests that a single suppressor level may not be appropriate in all cases.

ACKNOWLEDGMENTS

Work supported by the NIH NIDCD F32-DC007536, R01-DC02251, and P30-DC04662. The authors would like to thank Darcia Dierking and Cassie Garner for their assistance with data collection. Helpful comments from John Rosowski and an anonymous reviewer on a previous version of the manuscript are also gratefully acknowledged.

ANSI (1996). "Specifications for audiometers," ANSI 3.6-1996, American National Standards Institute, New York.

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.

Dhar, S., and Shaffer, L. A. (**2004**). "Effects of a suppressor tone on distortion-product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-Grams," Ear Hear. **25**, 573–585.

Goodman, S. S., Withnell, R. H., and Shera, C. A. (**2003**). "The origin of SFOAE microstructure in the guinea pig," Hear. Res. **183**, 7–17.

Gorga, M. P., Neely, S. T., Dierking, D. M., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. F. (**2003a**). "Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **114**, 263–278.

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

Gorga, M. P., Neely, S. T., Dorn, P. A., and Konrad-Martin, D. (**2002**). "The use of distortion product otoacoustic emission suppression as an estimate of response growth," J. Acoust. Soc. Am. **111**, 271–284.

He, N-j., and Schmiedt, R. A. (**1993**). "Fine structure of the $2f_1$-f_2 acoustic distortion product: Changes with primary level," J. Acoust. Soc. Am. **94**, 2659–2669.

He, N-j., and Schmiedt, R. A. (**1997**). "Fine structure of the 2f1-f2 acoustic distortion products: Effects of primary level and frequency ratios," J. Acoust. Soc. Am. **101**, 3554–3565.

Heitmann, J., Waldmann, B., and Plinkert, P. K. (**1996**). "Limitations in the use of distortion product otoacoustic emissions in objective audiometry as the result of fine structure," J. Acoust. Soc. Am. **253**, 167–171.

Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1$-f_2 removes DP-gram fine structure—Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (**2006**). "Influence of primary-level and primary-frequency ratio on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 418–428.

Kalluri, R., and Shera, C. A. (**2001**). "Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," J. Acoust. Soc. Am. **111**, 1800–1809.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). "The level and growth behavior of the $2f_1$-f_2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss," J. Acoust. Soc. Am. **103**, 3431–3444.

Mauermann, M., and Kollmeier, B. (**2004**). "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source," J. Acoust. Soc. Am. **116**, 2199–2212.

Mauermann, M., Uppenkamp, S., van Hengel, P. W. J., and Kollmeier, B. (**1999**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1," J. Acoust. Soc. Am. **106**, 3473–3483.

Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.

Neely, S. T., Johnson, T. A., and Gorga, M. P. (**2005**). "Distortion-product otoacoustic emission measured with continuously varying stimulus level," J. Acoust. Soc. Am. **117**, 1248–1259.

Neely, S. T., and Liu, Z. (1993). "EMAV: Otoacoustic emission averager," Tech. Memo. 17, Boys Town National Research Hospital, Omaha, NE.

Oswald, J. A., and Janssen, T. (**2003**). "Weighted DPOAE input/output functions: A tool for automatic assessment of hearing loss in clinical application," Z. Med. Phys. **13**, 93–98.

Rao, K. R., and Yip, P. (**1990**). *Discrete Cosine Transform: Algorithms, Advantages, Applications* (Academic, Boston).

Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). "Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.

Shera, C. A. (**2004**). "Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions," Ear Hear. **25**, 86–97.

Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.

Siegel, J. H., and Hirohata, E. T. (**1994**). "Sound calibration and distortion product otoacoustic emissions at high frequencies," Hear. Res. **80**, 146–152.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**). "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1999**). "Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears," J. Acoust. Soc. Am. **106**, 2669–2678.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions." J. Acoust. Soc. Am. **105**, 275–292.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction

Alain de Cheveigné and Daniel Pressnitzer
Equipe Audition, FRE 2929, CNRS, Université Paris 5, ENS, 29 Rue d'Ulm, F-75230 Paris cedex 05, France

(Received 6 February 2005; revised 18 March 2006; accepted 22 March 2006)

Temporal models of pitch and harmonic segregation call for delays of up to 30 ms to cover the full range of existence of musical pitch. To date there is little anatomical or physiological evidence for delays that long. We propose a mechanism by which delays may be synthesized from cross-channel phase interaction. Phases of adjacent cochlear filter channels are shifted by an amount proportional to frequency and then combined as a weighted sum to approximate a delay. Synthetic delays may be used by pitch perception models such as autocorrelation, segregation models such as harmonic cancellation, and binaural processing models to explain sensitivity to large interaural delays. The maximum duration of synthetic delays is limited by the duration of the impulse responses of cochlear filters, itself inversely proportional to cochlear filter bandwidth. Maximum delay is thus frequency dependent. This may explain the fact, puzzling for temporal pitch models such as autocorrelation, that pitch is more salient and easy to discriminate for complex tones that contain resolved partials. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195291]

PACS number(s): 43.66.Ba, 43.64.Bt, 43.66.Hg [AK] Pages: 3908–3918

I. INTRODUCTION

The autocorrelation (AC) model is a popular account for pitch perception (for a review of pitch models see de Cheveigné, 2005). Initially proposed by Licklider (1951) and later refined by Meddis and Hewitt (1991a, b) and others, the AC model accounts well for a wide range of pitch phenomena and is consistent with the electrophysiological recordings of Cariani and Delgutte (1996a, b). It has however at least two weaknesses. First, in the implementation proposed by Licklider it requires arrays of internal *delays* of up to about 30 ms to account for the full range of musical pitch (Pressnitzer *et al.*, 2001a). Conclusive anatomical and physiological evidence for such delays has not been found to date. Second, it fails to predict the greater pitch salience and discrimination accuracy of stimuli with resolved harmonics, as compared to stimuli that contain only unresolved harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Carlyon and Shackleton, 1994; Oxenham *et al.*, 2004), although this question is still in debate (Meddis and O'Mard, 1997; Carlyon, 1998).[1]

Another popular account of pitch is based on the concept of pattern matching (de Boer, 1956; Goldstein, 1973; Wightman, 1973; Terhardt, 1974): periodicity pitch (also called low, residue, or virtual pitch) is derived from the pattern formed by the frequencies of individual partials. For that, these partials must be *resolved* by peripheral filtering, and thus pattern matching cannot account for the pitch of complexes with only unresolved partials. The pitch evoked by those stimuli is weak but nevertheless musical (Pressnitzer *et al.*, 2001a; Kaernbach and Bering, 2001). To account for pitch over a full range of stimulus conditions, pattern matching must be associated with another mechanism such as AC. This "dual mechanism" hypothesis (Shackleton and Carlyon, 1994) is less parsimonious than a "unitary model" (e.g., Meddis and O'Mard, 1997), but has the appeal that it accounts for a difference in pitch salience between resolved and unresolved complexes.

In addition to pitch, internal delays are invoked for other aspects of auditory processing. The cancellation model of harmonic sound segregation uses a delay equal to the period of an unwanted sound (de Cheveigné, 1993, 1997, 1999). Delays of about 10 ms are needed to account for typical concurrent vowel identification results. Stimuli with interaural delays of up to 10 ms (Mossop and Culling, 1999) or 20 ms (Blodgett *et al.*, 1956) can be discriminated on the basis of laterality. To explain this using a crosscorrelation display model (e.g., Stern and Shear, 1996), this display would need to extend to delays of a similar size. There are however other accounts of laterality that involve shorter delays (Saberi *et al.*, 2001). Other pitch models in addition to autocorrelation require delays, for example the STI (strobed temporal integration) model of Patterson *et al.* (1995), or the cancellation model of de Cheveigné (1998). Long delays are also required by models that attempt to account for the accuracy of pure tone pitch discrimination on the basis of intervals that extend over multiples of the period (e.g., de Cheveigné, 1989, 2000; Slaney, 1990). Delay is an ingredient of many auditory models, and the lack of evidence for neural delays over the range needed is puzzling.

This paper examines a mechanism by which delays can be synthesized by phase interaction between cochlear channels. Interestingly, this mechanism might also account for the resolvability limits of pitch. It does, however, not absolve from the need of a second mechanism involving long physical delays for the pitch elicited by unresolved partials. The next section presents the idea, Sec. III discusses issues that must be addressed before the idea is considered plausible for auditory modeling, and Secs. IV and V, respectively, review relevant psychophysical and physiological data.

 0001-4966/2006/119(6)/3908/11/$22.50

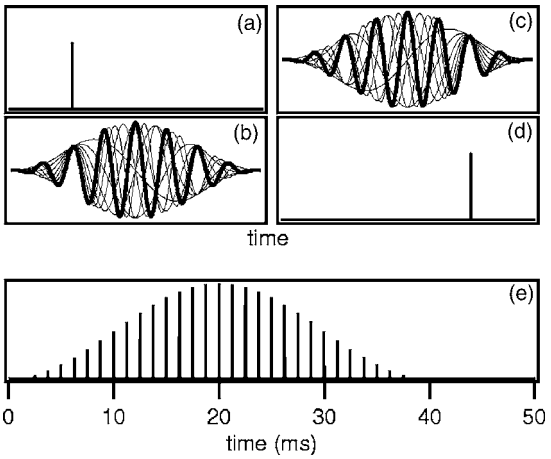


FIG. 1. Illustration of a delay implemented with a short-term Fourier transform. (a) Pulse. (b) Decomposition of the pulse over a basis of windowed sinusoids (first ten terms). (c) Same sinusoids after a phase shift proportional to frequency. (d) Sum of phase-shifted windowed sinusoids. The resulting pulse is delayed with respect to the input pulse. (e) Illustration of the range of pulses that can be synthesized by adding the windowed sinusoids of (b) supposing a 40 ms raised-cosine window.

II. DELAY AS PHASE SHIFT

A delay in the time domain is equivalent to a phase shift proportional to frequency in the Fourier domain. For a delay of D, the phase shift is

$$\phi = \omega D, \tag{1}$$

where $\omega = 2\pi f$. Indeed, if the spectrum of $x(t)$ is $X(\omega)$, the delayed signal is

$$x(t+D) = \int_{-\infty}^{+\infty} [X(\omega)e^{-j\omega D}]e^{-j\omega t}d\omega. \tag{2}$$

Any delay can be obtained in this manner with the Fourier transform. With the *short-term* Fourier transform, in which the signal is decomposed over a basis of windowed sinusoids, a restricted set of delays can be synthesized. Figure 1 illustrates the process: The pulse (a) is analyzed into a sum of windowed sinusoids (b). Sinusoids are phase shifted (c) and then added to produce the delayed pulse (d). Interference between the sinusoids is everywhere destructive except at one time point. By manipulating the phases, the position of this point can be moved within the range covered by the analysis window. Figure 1(e) illustrates the range of pulses that can be synthesized for a raised-cosine window of 40 ms.

The cochlea has been likened to a Fourier transformer (von Helmholz, 1877). Could it be used to support a similar operation? Several differences between auditory peripheral frequency analysis and a short-term Fourier transform need to be considered. Cochlear filter bank parameters (bandwidth, sampling) are not uniform, there are nonlinearities at several stages (cochlear filtering, transduction, and neural processing), and the neural analogue of summation in Eq. (2) is not immediately obvious. We will show that a delaylike operation can nevertheless be approximated using a model of cochlear filtering instead of Fourier analysis.

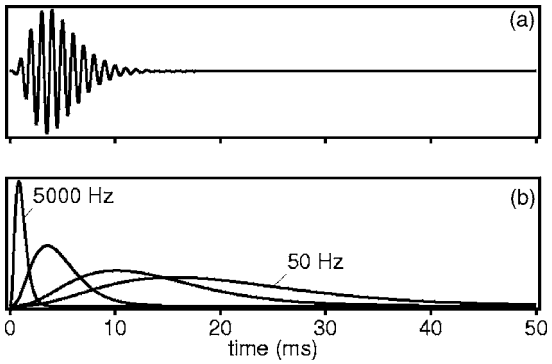


FIG. 2. Top: impulse response of a 1000 Hz gammatone filter. Bottom: envelopes of impulse responses of gammatone filters centered at 50, 200, 1000, and 5000 Hz.

A. Delays produced with a gammatone filterbank

An approximation of cochlear filtering is the gammatone filter (Carney and Yin, 1988). Its impulse response is the product of a sinusoid and an envelope function

$$h(t) = g(t)\cos(2\pi ft + \phi), \tag{3}$$

where f is the center frequency (CF) of the filter, ϕ determines the phase, and

$$g(t) = at^{(n-1)}e^{-2\pi bt} \tag{4}$$

defines the envelope. Parameter n is the order of the filter ($n=4$ in this paper) and b determines the filter bandwidth. For a fourth-order gammatone, b and the equivalent rectangular bandwidth B (ERB, Moore and Glasberg, 1983) are related by $b=1.018B$. Bandwidths of auditory filters follow approximately the formula $B=24.7+0.108f$ where f is the filter center frequency, so that high-frequency channels are wider than low-frequency channels. Figure 2(a) illustrates the impulse response of a gammatone filter centered on 1 kHz, and Fig. 2(b) shows temporal envelopes of impulse responses of selected filters with CFs between 50 and 5000 Hz. Note that low-CF responses peak later and last longer than high-CF responses. In contrast, the basis functions of the short-term Fourier transform illustrated in Fig. 1(b) had a common envelope determined by the windowing function.

We now make two important assumptions. The first is that the phase of each channel of the gammatone filterbank [ϕ in Eq. (3)] may be adjusted arbitrarily. Specifically, we need to set the phase of each channel k to $\phi_k = -2\pi Df_k$, where f_k is the CF and D is the desired delay. Doing so produces a peak in the fine structure of all impulse responses at D. Possible physiological sources of phase shift are discussed in Sec. III. The second assumption is that, after phase shifts, the channels can be summed. The entire operation (from input to sum) can be seen as a filter with impulse response

$$H_D(t) = \sum_k h_{k,D}(t), \tag{5}$$

where $h_{k,D}(t)$ is the phase-shifted impulse response of channel k. Figure 3(a) illustrates $H_D(t)$ for values of D ranging from 1 to 40 ms in 1 ms steps. Each response consists of a

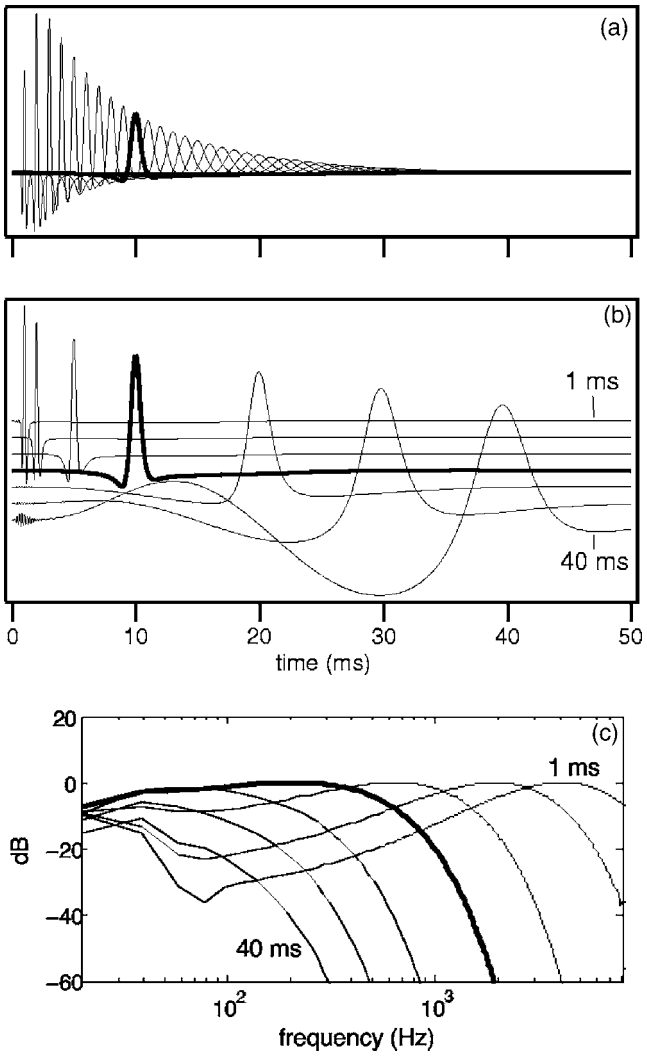


FIG. 3. (a) Impulse responses of gammatone-based synthetic delays of 1 to 40 ms in 1 ms steps. Bold: 10 ms. (b) Impulse responses for delays of 1, 2, 5, 10, 20, 30, and 40 ms, normalized for equal peak amplitude and offset vertically for clarity. Note the reduced amplitude and wider shape of impulse responses at longer delays. (c) Magnitude transfer functions of the same delays. Note the reduced amplitude and sharp low-pass characteristic of larger delays, and the more gradual high-pass characteristic of shorter delays.

pulse that appears to be delayed by D. Such delays will be termed *synthetic* to distinguish them from physical delays (e.g., neural transmission delays). Responses are indeed delayed, but two observations must be made. First, pulse amplitudes are smaller for larger D. Second, pulse shapes are not all the same, being wider for larger D. This is obvious in Fig. 3(b) where selected responses have been normalized for equal peak amplitude, and offset vertically for clarity. Figure 3(c) shows magnitude transfer functions of these selected synthetic delays. Longer delays have a relatively severe low-pass characteristic, and shorter delays a more gradual high-pass characteristic. This can be understood by noting that synthetic delays are restricted to the extent of impulse responses of the filter bank [Fig. 1(e)]. Responses within high-frequency channels fade quickly [Fig. 2(b)] so long delays are devoid of high frequencies. Conversely, the late onset of low frequency channels causes short delays to lack low frequencies. Details about simulations are given in the Appendix .

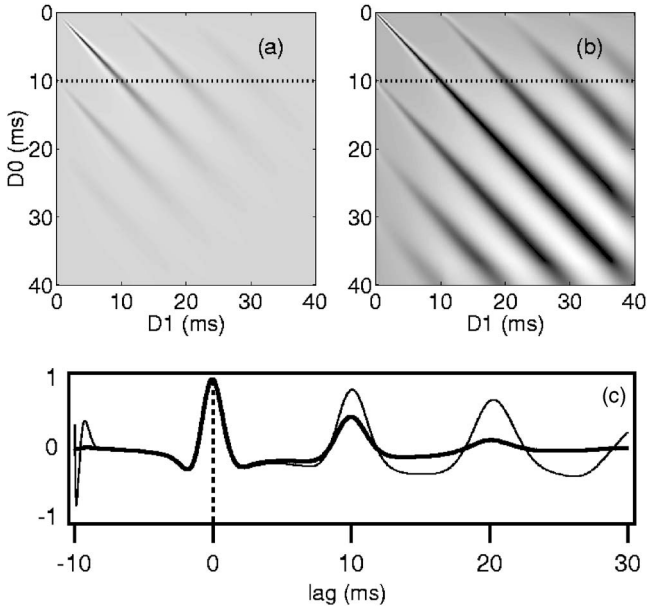


FIG. 4. (a) Cross product between wave forms delayed by $D0$ and $D1$, in response to a 100 Hz pulse train. (b) Same, normalized (see text), (c) Thick line: cross product plotted as a function of $D=D1-D0$, for $D0=10$ ms [dotted line in (a)]. This approximates the ACF. Thin line: same, normalized [dotted line in (b)].

B. Using synthetic delays to produce autocorrelation

Synthetic delays may be used to approximate an autocorrelation function (ACF). The running ACF of a wave form may be defined as the cross product between the delayed wave form $x(t+D)$ and the undelayed $x(t)$

$$r_t(D) = (1/W) \sum_{j=t+1}^{t+W} x(j)x(j+D), \tag{6}$$

where W is the size of the integration window (supposed square for simplicity) and the subscript t indicates the time at which the calculation is made. We wish to implement Eq. (6) using the synthetic delay mechanism to obtain the delayed and undelayed terms. A complication is that we cannot set $D=0$ because all gammatone impulse responses are initially zero, and we do not have access to the original acoustic wave form. What we can do instead is calculate the cross product between two wave forms synthetically delayed by D_0 and D_1 such that $D_1-D_0=D$

$$r_t(D_0,D_1) = (1/W) \sum_{j=t+1}^{t+W} x_{D_0}(j)x_{D_1}(j), \tag{7}$$

where $x_{D_0}(t)$ and $x_{D_1}(t)$ are obtained by convolving $x(t)$ with $H_{D_0}(t)$ and $H_{D_1}(t)$, respectively. Figure 4(a) displays this cross product as a function of D_0 and D_1 for a stimulus consisting of a 100 Hz pulse train. Large values map to dark. The pattern consists of diagonal "stripes" with a spacing that reflects the stimulus period, 10 ms. An approximation of the ACF can be obtained by taking a section at any nonzero value of D_0 [Fig. 4(c), thick line].

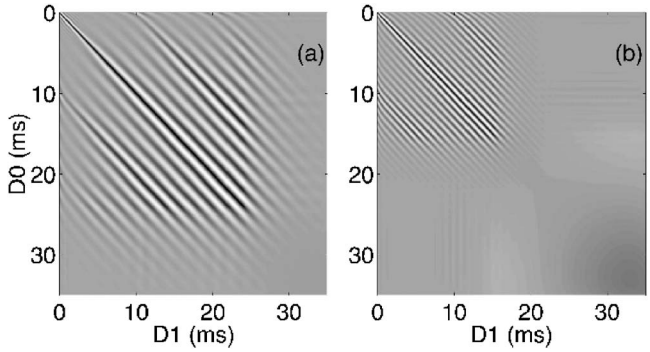


FIG. 5. Same as Figs. 4(a) and 4(b) for 100 Hz pulse trains high-pass filtered from 500 Hz (a) or 1000 Hz (b). The upper left-hand corner of each graph reflects stimulus periodicities (mainly the fundamental and component nearest cutoff).

The pattern fades for large values of D_0 and D_1 as a result of two factors: greater attenuation at longer delays [Fig. 3(a)] and delay-dependent spectral distortion [Fig. 3(c)] that reduces the similarity between $x_{D_0}(t)$ and $x_{D_1}(t)$. Effects of the first factor may be countered by normalizing the cross product

$$r'_t(D_0,D_1) = r_t(D_0,D_1) \times \left[\sum_{j=t+1}^{t+W} x_{D_0}(j)^2\right]^{-1/2} \left[\sum_{j=t+1}^{t+W} x_{D_1}(j)^2\right]^{-1/2}. \quad (8)$$

The normalized cross product is shown in Fig. 4(b) for the same wave form as in Fig. 4(a). Normalization requires a large dynamic range to be effective; if it were limited (as might be the case for a physiological implementation), the pattern would be intermediate between Figs. 4(a) and 4(b).

The thin line of Fig. 4(c) shows the approximation of the ACF obtained after amplitude normalization: The amplitude decrease is less fast than without normalization (thick line), but higher-order peaks are still somewhat broader than low-order peaks.

Figure 5 shows the normalized cross-correlation pattern in response to a 100 Hz complex tone high-pass filtered at 500 Hz (a) or 1000 Hz (b). Two things may be noted. First, periodicity-related structure is restricted to the upper left corner and does not extend beyond about 30 ms (a) or 15 ms (b). Second, the 10 ms fundamental pattern is overlayed by components near the cutoff: 2 ms in (a) and 1 ms in (b). These effects can be understood from the low-pass characteristic of most delays [Fig. 2(c)]. This illustrates once again an important property of synthetic delays: the range of delays that they offer is *frequency dependent*.

To summarize, relatively large "synthetic" delays may be produced by cross-channel phase interaction at the output of a cochlear model. Their size is limited by the CF-dependent duration of impulse responses: the longest delays are available only at low frequencies. The auditory system might use such delays to implement various processing mechanisms including autocorrelation. The rest of this paper discusses this idea in greater detail.

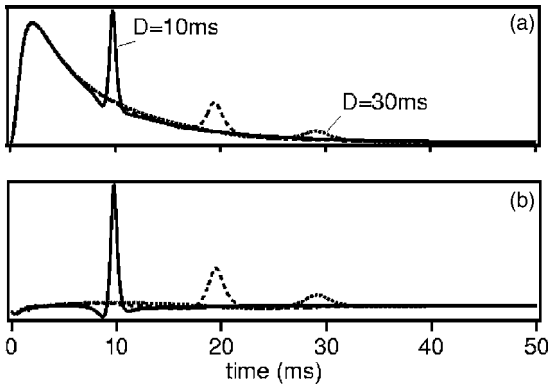


FIG. 6. Effects of non linear transduction. (a) Impulse responses for synthetic delays of 10, 20, and 30 ms in the case that filter outputs are processed by half-wave rectification. (b) Same, but half-wave rectification is followed by a high-pass "unsmoothing" filter.

III. ISSUES

This section examines several issues raised by the synthetic delay hypothesis. The purpose is to identify important issues but not necessarily resolve them.

A. Nonlinearities

Linear processing was assumed up to this point, but several stages of auditory processing are nonlinear. Hair cell transduction is corresponds roughly to half-wave rectification. Figure 6(a) illustrates the effect of introducing half-wave rectification before summation. Impulse responses are now dominated by a prominent initial portion that is common to all delays. However, Fig. 6(b) shows the result of high-pass filtering the half-wave rectified wave form in each channel (by subtracting the same wave form smoothed by convolution with a triangular window with a span of twice the characteristic period). Differences with Fig. 3 are now minimal. Such high-pass filtering might be obtained by a combination of fast excitatory and smoothed inhibitory inputs to a neuron, or by particular membrane properties (Kalluri and Delgutte, 2001).

Another source of nonlinearity is cochlear mechanics. Frequency and phase characteristics are known to change with level (Robles and Ruggero, 2001). If this were to result in level-dependent synthetic delays, the delays would be less useful. However measurements suggest that level affects mainly the relative *amplitudes* of the early and later segments of a chirp-shaped impulse response, leaving the temporal structure (position of peaks and zero crossings) invariant (Carney *et al.*, 1999; Robles and Ruggero, 2001; Shera, 2001). This in turn insures level invariance of synthetic delays. Shera (2001) commented that this property puts strong constraints on the mechanics and biophysical properties of the cochlea. We may note that it is a prerequisite to produce level-independent synthetic delays.

B. Loss of synchrony

The synthetic delay mechanism depends on precise temporal coding, which is known to degrade rapidly beyond

2.5–5 kHz in cat (Johnson, 1980). In man the limit is unknown. Stimulus components above that limit cannot be delayed.

C. Phase shifts

In our simulations, we implemented phase shifts by adjusting the phase term of the gammatone impulse response [Eq. (3)]. In the auditory system, a phase shift within $[0,2\pi]$ could be approximated by a physical delay within $[0,1/f]$.[2] This might seem to defeat the purpose of the model; the point is that these physical delays are shorter than the synthetic delays that they produce when combined across channels. The required range is further reduced by a factor of 2 if subtraction is allowed (e.g., inhibitory interaction). Phase shift may also be implemented as a weighted sum of two signals in quadrature phase, as suggested for binaural interaction by McAlpine *et al.* (2001). The appeal of this hypothesis is that the same set of signals would serve multiple purposes, and that weights may be easier to tune than physical delays. Also, only one physical delay $(1/4f)$ is needed to produce the quadrature signal, rather than a range of delays. A third conceivable source of phase shifts is basilar membrane propagation (Shamma, 1985a; van der Heijden and Joris, 2005). Our model requires *independent* control of phase and CF whereas these factors presumably covary along the basilar membrane (BM), but the redundancy of overlapping channels might nevertheless allow BM phase shifts to be exploited.

D. Cross-channel summation

The neural equivalent of summation [Eq. (5)] would involve cross-frequency convergence of phase-locked inputs. This may occur at several stages within the auditory system, as reviewed further in Sec. V. The delayed pattern is supposed to be temporally accurate, and this narrows the options down to stages that have a synchronized *output* in addition to input. A possible way to relax this requirement is discussed below.

E. Two possible implementations of the ACF

A first implementation is schematized in Fig. 7(a). It approximates the ACF for lag $D=D_1-D_0$. Each branch requires a cascade of two neurons, the first with additive properties and the second with coincidence-counting properties. These two steps might instead be combined within the dendritic field of a single neuron with appropriate properties. Agmon-Snir *et al.* (1998) give an example of complex dendrite-based computations. Calculating the ACF over a range of lags requires either an array of such phase-shift-and-sum circuits, or else a single circuit but with a tuning mechanism.

A second possible implementation is schematized in Fig. 7(b). To understand how it works, note that the right hand side of Eq. (7) can be expanded and rearranged

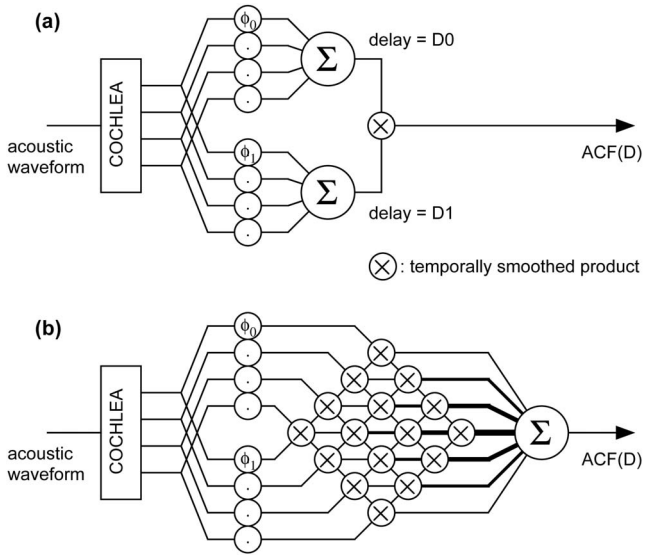


FIG. 7. (a) Operations required to produce one delay of the autocorrelation function. (b) Functionally equivalent circuit. The first circuit requires two stages of accurate temporal processing (the product is applied to fast-varying sums). The second needs only one stage (the sum is applied to temporally smoothed products).

$$r_t(D_0,D_1)=\sum_{k,k'} r_{k,k',t}(D_0,D_1), \tag{9}$$

where

$$r_{k,k',t}(D_0,D_1)=(1/W)\sum_{j=t+1}^{t+W} x_{k,D_0}(j)x_{k',D_1}(j) \tag{10}$$

is the cross product between channel k phase shifted according to the formula for D_0, and channel k' phase shifted according to the formula for D_1. Equation (9) replaces the temporally smoothed product of fast-varying sums of Eq. (7) by a slow-varying sum of temporally smoothed products. The result is the same.

Several things make this second formulation attractive. First, a physiological implementation needs to maintain synchrony only up to the input of the initial (cross-product) stage, and not over two synapses as in Fig. 7(a). Second, cross coincidence between channels of different CF has been proposed repeatedly to account for a range of tasks including pitch and loudness perception (Schroeder, 1977; Carney, 1990; Shamma, 1985b; Shamma *et al.*, 1988; Deng and Geisler, 1987; Loeb *et al.*, 1983; Shamma and Klein, 2000; Heinz, 2001b; Carney *et al.*, 2002). This second formulation fits with the notion that the auditory system uses multipurpose processing, rather than specialized modules for each task such as pitch perception, and it adds to the range of possible incarnations of a useful functional model such as autocorrelation.

F. Cancellation

Delays are also required by hypothetical cancellation mechanisms involved in binaural or periodicity-based processing and sound segregation (e.g., Durlach, 1963, de Cheveigné, 1993). It has been argued (de Cheveigné, 2001) that cancellation and correlation-based statistics can be derived

one from the other. On the basis of that argument we may assume that the same mechanism that produces the ACF can also produce a cancellation-based statistic, and that both statistics could be based in part on inhibitory as well as excitatory interaction. In other words, synthetic delays can fulfill the needs of cancellation models.

G. Parameters

Each synthetic delay involves K phase parameters ϕ_k, where K is the number of cochlear channels. All these parameters are functions of a single parameter via Eq. (1), but one may wonder how the auditory system applies that equation, or at least discovers which parameter sets correspond to delays. One possibility is that learning is involved in the tradition of learning-based models (Licklider, 1959; Terhardt, 1974; Shamma and Klein, 2000; see de Cheveigné, 2005 for a review). This issue is too complex to be further addressed within the scope of this paper.

To summarize, the synthetic delay model raises several issues that need to addressed before the model is deemed plausible. The next section examines how a model based on synthetic delays might account for pitch perception.

IV. SYNTHETIC DELAYS FOR PITCH PERCEPTION

In his model of pitch, Licklider (1951) proposed that spike trains from the cochlea were processed by a neural network consisting of a series of synaptic delays and coincidence neurons. Later models, such as Meddis and Hewitt (1991a,b), are less specific but assume nevertheless that delays are produced within the auditory system (by synaptic transmission, axonal conduction, or rebound from inhibition) with durations sufficient to calculate an ACF over the range necessary to for pitch perception. ACFs calculated within each CF channel are then summed to produce a *summary autocorrelation function* (SACF) from which the period is derived (Meddis and Hewitt, 1991a, b). Two major objections have been made to the AC model. The first is the lack of evidence for neural delays, at least at stages where phase locking is present. The second is that it works *too* well: It can handle any periodic wave form, and thus predicts that any periodic sound should evoke pitch. This argues against it, as pitch has limited regions of existence or salience. The synthetic delay model may help to address the second objection as well as the first.

Synthetic delays have two major limitations: (1) they cannot be applied to components for which there is no phase locking, and (2) they cannot exceed the impulse response duration of cochlear filters activated by the signal to be delayed. The first limit fits with the observation that periodicity pitch exists only for stimuli with components that allow phase locking (Moore, 2003, p. 215).[3] The second limit implies that period measurements are restricted to periods shorter than the impulse response of filters that respond to components of the tone. Such is the case if components are more widely spaced than filter bandwidths, that is, if they are *resolved*. Synthetic delays thus give the autocorrelation model a property that is usually associated with pattern-matching models. Data that show that resolvability is important for pitch (or other tasks) can potentially be explained by the use of synthetic delays.

Figure 8(a), dotted line, shows the latency of the gammatone envelope peak as a function of CF, in units of characteristic period (CP, inverse of CF). Latencies of the skirts at −20 dB are plotted as thick lines, and at −40 and −60 dB as thinner lines. If the dynamic range of processing were 20 dB, the useful duration would be the interval between thick lines, replotted as a thick line in Fig. 8(b). Superimposed upon that plot are data from several studies that measured pitch discrimination. Symbols indicate the rank and frequency of the lowest component of stimuli used in those studies. Large triangles are the resolved (full) and unresolved (open) conditions of Carlyon and Shackleton (1994). Small symbols are for other studies that measured pitch discrimination thresholds (Houtsma and Smurzynski, 1990; Krumbholz, Patterson, and Pressnitzer, 2000; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003). For each study, full symbols represent conditions for which pitch thresholds were low (below the geometric mean over conditions) and open symbols conditions for which they were high.

The 20 dB line separates the two populations quite neatly: good performance below, and poor above. This line is a contour of constant resolvability, and the data thus illustrate the well-known observation that pitch discrimination is better for stimuli with at least some resolved components. The line also marks the maximum delay that can be synthesized by the model, and the data thus equally support the hypothesis that synthetic delays are required for accurate pitch. This conclusion is based on the 20 dB value chosen for the dynamic range, and also on the gammatone model chosen to parametrize auditory filters.

Note that maximum delay *covaries* with resolvability but does not depend upon it. Bernstein and Oxenham (2003) found that discrimination thresholds were not improved when odd partials were sent to one ear and even partials to the other. That manipulation increases interpartial spacing at each ear, and should improve discrimination if resolvability were the determining factor. Their data fit the hypothesis, put forward by Moore (2003) and recently explored by Bernstein and Oxenham (2005), that internal delays are limited to some *ad hoc* value that depends on the spectral range occupied by the stimulus. For the synthetic delay model this limit is an emergent property. The model is consistent with other aspects of pitch that are hard to explain on the basis of autocorrelation. Bernstein and Oxenham (2005) found that the transition from good to poor pitch discrimination occurs at a lower harmonic rank at high stimulus levels, for which cochlear filters are wider and impulse responses presumably shorter. Supposing that pure tone discrimination involves higher order peaks of the ACF (de Cheveigné, 1989; Slaney, 1990), the same explanation might account for elevated pure tone thresholds at high levels (Bernstein and Oxenham, 2005). Synthetic delays require intact cochlear filters. Absence of filters in cochlear implantees might explain their lack of accurate pitch perception, despite the precise temporal patterns of nerve discharge produced by electrical stimulation.

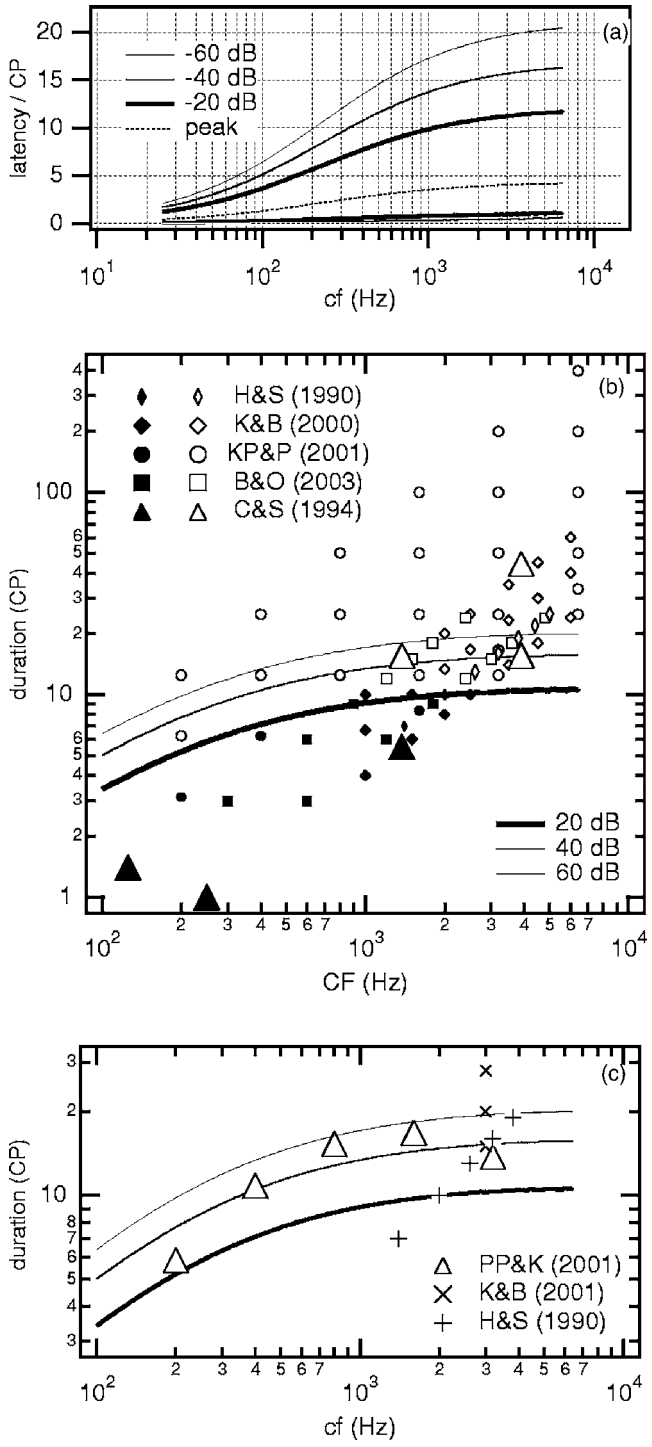


FIG. 8. (a) Latencies of the peak (dotted line) and onset and offset (full lines) of gammatone filter impulse responses in units of characteristic period (1/CF). The useful duration is the interval between onset and offset. (b) Thick line: useful duration of the gammatone impulse response given a dynamic range of 20 dB. Thin lines: same for 40 and 60 dB. Symbols represent the rank of the lowest component of stimuli used in pitch discrimination experiments (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994; Krumbholz, Patterson, and Pressnitzer, 2000; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003). For each study, open (respectively, full) symbols represent conditions for which threshold was above (respectively, below) the geometric mean over conditions. The transition from good to poor discrimination in most of these studies was typically abrupt and large (up to an order of magnitude). (c) Lines are as in (b). Crosses represent the rank of the lowest component of stimuli for which musical tasks were performed at better than chance level (Houtsma and Smurzynski, 1990; Kaernbach and Bering, 2001). The triangles are the lower limit of melodic pitch as determined by Pressnitzer *et al.* (2001b).

A difficulty, shared with pattern matching models, is that stimuli with unresolved components may still evoke a pitch, albeit with weak salience and poor accuracy. Indeed, the studies shown in Fig. 8(c) report threshold measurements for ranks of up to 400. The pitch of such stimuli is weak but may still be musical. Houtsma and Smurzynski (1990) and Kaernbach and Bering (2001) found better-than-chance interval recognition for stimuli with ranks of up to 19 and 28, respectively. Pressnitzer *et al.* (2001a) likewise found a lower limit of melodic pitch that was well outside the limits of resolvability [Fig. 8(c)].

It is worth noting, however, that the limit plotted in Fig. 8 is conditional on the gamma-tone model chosen to represent cochlear filters and on its parameters. From human otoacoustic emission and behavioral data on nonsimultaneous masking, Shera *et al.* (2002) and Oxenham and Shera (2003) argued that filters might be twice as narrow as found by Moore and Glasberg (1983), and thus their impulse responses twice as long. If that conclusion can be extended to the context of complex tones, the limit of Fig. 8 might shift to higher ranks. Measurements of basilar membrane motion or auditory nerve responses reveal complex, multiple-spindle-shaped click or revcor responses (Recio *et al.*, 1998; Recio and Rhode, 2000; Lin and Guinan, 2000, 2004) that do not conform to the gammatone model. In at least one case the response extended in time as far as the 90th cycle (Recio *et al.*, 1998, Fig. 4). Thus, it is possible that the synthetic mechanism could operate beyond the limit shown in Fig. 8.

However, this explanation cannot account for pitch perception with electrical stimulation (Pilj and Scharz, 1995). For that, one would need to postulate some other source of delay, for example, rebound from inhibition, or axonal conduction. This "dual-delay" hypothesis, which could also account for the pitch of unresolved stimuli, resembles the "dual pitch mechanism" hypothesis (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994). It differs in that the hypothetical pitch mechanism that follows the delays could be unitary, avoiding the need for a "translation" mechanism to convert outputs of different mechanisms to a common ground. Evidence for "translation noise" was found by Carlyon and Shackleton (1994), but their interpretation has recently been challenged (Micheyl and Oxenham, 2004; Gockel *et al.*, 2004; Oxenham *et al.*, 2005).

To summarize, the synthetic delay mechanism can be used to implement a delay-based pitch model such as autocorrelation. The availability of period-size synthetic delays coincides with component resolvability. Data that show the importance of resolvability for pitch are consistent with the hypothesis that synthetic delays are involved. However pitch is evoked also by stimuli that are not resolvable, and in subjects that lack peripheral filters. Another source of delay is needed to handle cases for which the synthetic delay mechanism is unavailable.

V. PHYSIOLOGICAL EVIDENCE

A way to evaluate the synthetic delay hypothesis is to find anatomical or physiological evidence for or against the ingredients that it assumes, or the activity that it would im-

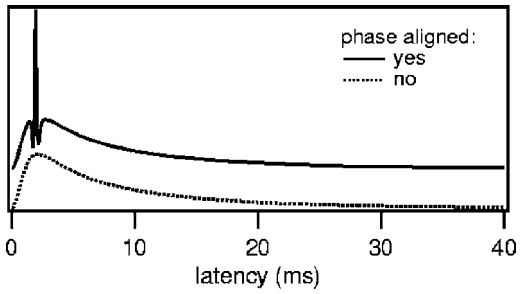


FIG. 9. Sum across auditory fibers of modeled firing probabilities in response to a click, with (full) and without (dashed) phase alignment. Plots are shifted vertically for clarity.

ply. These vary according to whether we assume the standard implementation of Fig. 7(a), or the alternative implementation of Fig. 7(b).

The implementation of Fig. 7(a) leads us to expect to observe phase-locked, delayed responses somewhere within the auditory system. Lack of evidence for them is as damning as for other hypothetical sources of delay such as axonal conduction, although it is possible that they exist but in a form that is hard to observe (for example, within thin axones or dendrites). In contrast, the implementation of Fig. 7(b) does not imply observable delays. In that implementation, the ACF is "assembled" from between-channel cross-correlation terms. Both implementations lead us to expect responses tuned to periodicity pitch. The lack of period-selective responses is a problem for any model of pitch, including pattern matching (but see Bendor and Wang, 2005).

Both implementations require cross-frequency convergence of phase-locked inputs. This is known to occur in the cochlear nucleus, for example, in octopus (onset-I) and D-stellate (chopper-C) cells (Jiang *et al.*, 1996; Oertel *et al.*, 2000). It could also, occur at higher stages that receive accurate temporal patterns relayed by primarylike cells, such as the superior olivary complex, nuclei of the lateral lemniscus, or inferior colliculus. Both implementations assume linear summation, of fast patterns in Fig. 7(a) or slow in Fig. 7(b). Accurate linear summation has been observed in cultured pyramidal cells (Cash and Yuste, 1998). The implementation of Fig. 7(b) assumes cross coincidence between channels with different CF, an assumption made by many models (e.g., Carney, 1990).

Both implementations assume that the *phase* of each cochlear channel can be manipulated. Onset-I and chopper-C cells of the cochlear nucleus receive input from AN fibers over a wide range of CFs (Jiang *et al.*, 1996; Oertel *et al.*, 2000). They respond to clicks with little temporal jitter, despite the fact that activity within fibers that feed them is presumably staggered in time due to phase dispersion along the basilar membrane. Response to a complex in random phase is less good than sine or cosine phase, implying that cross-frequency phase alignment is critical (Palmer and Winter, 1996; Evans and Zhao, 1997; Shofner, 1999; Winter and Palmer, 1995). Figure 9 (dashed line) shows an estimate of nerve firing probability pooled across the entire AN, as might drive one of these onset cells (based on a gammatone model of peripheral filtering). With no phase alignment the pattern is relatively wide, but aligning the phases makes it temporally much sharper (full line). Octopus cells require multiple excitatory postsynaptic potentials (EPSPs) to occur within a very short window (about 1 ms) to fire (Oertel *et al.*, 2000), a condition obviously easier to meet if inputs are phase aligned. In the onset cell model of Kalluri and Delgutte (2001), realistic entrainment required that inputs span less than 1.5 kHz, an unrealistically small range (Jiang *et al.*, 1996). That constraint could have been relaxed if phases were aligned. Observation of a *range* of phases for a given CF would favor the model. Indeed, McAlpine *et al.* (2001) report responses in quadrature phase from which other phase relations may be synthesized.

Winter and Palmer (1995) reported first spike latencies for onset cells between 0.5 and 5 ms, but these were level dependent. Behrend *et al.* (2002) found cells in the gerbil superior paraolivary nucleus (SPN) with temporally accurate onset responses (jitter smaller than 100 μs for a subset of cells) and a range of latencies (0.9 to 11.2 ms). Some were phase locked to amplitude modulation up to 1 kHz, and most had CFs below 6 kHz and relatively broad tuning. SPN receives inhibitory input from medial nucleus of trapezoid body (MNTB) and excitatory input from multipolar and octopus cells in CN, that both integrate inputs over a range of CFs, and the dendritic fields of its neurons span a wide range (Dehmel *et al.*, 2002). However such evidence for systematic delays, as might be produced by the implementation of Fig. 7(a) (or other forms of neural delay) remains fragmentary. To summarize, ingredients required by the synthetic delay model are available even if explicit evidence for the model is lacking.

VI. DELAY, PHASE AND COCHLEAR FILTERING

The concept of the ear as a Fourier transformer dates to von Helmholtz (1877), who noted that we are deaf to the relative phases of low-order partials. It was subsequently taken for granted that phase information is discarded, but Huggins and Licklider (1951; Huggins, 1952) pointed out that frequency analysis can be enhanced by phase interaction between channels of a filterbank. Similar ideas are embodied in the lateral inhibitory network (LIN) of Shamma (1985b) which involves subtraction between adjacent channels to sharpen selectivity, or the model of Deng and Geisler (1987) involving multiplicative interaction. Carney (1990) searched for evidence of such interaction in the cochlear nucleus, and Heinz *et al.* (2001a, b) and Carney *et al.* (2002) suggested that it can explain loudness perception and tone detection in noise. In another twist on pitch perception, Shamma and Klein (2000) used cross coincidence to produce harmonic templates as required by pattern-matching models. Together, these phase-sensitive mechanisms form a richer set than those based only on magnitude spectral patterns. In the models of Jeffress (1948) and Licklider (1951), coincidence detectors are preceded by neural delays, that are also required by equalization-cancellation (Durlach, 1963), harmonic cancellation (de Cheveigné, 1993) or strobed temporal integration (Patterson *et al.*, 1992). Delays were initially assumed to be neural, but basilar membrane propagation delay has also

been proposed for binaural (Schroeder, 1977; Shamma *et al.*, 1988, 1989; Shamma, 2001; Joris *et al.*, 2005) and monaural processing (Hurst, 1895; Loeb, 1983). For the narrow-band signal that exists within a cochlear channel, delay is equivalent to *phase shift* and can be implemented as a weighted sum of two terms in quadrature phase. McAlpine *et al.* (2001) argued that such a mechanism might underly localization in the guinea pig, in place of Jeffress's delay-based mechanism. The synthetic delay model draws upon all these earlier ideas and models.

Our model involves peripheral filtering and phase manipulation. It could conceivably be reformulated in terms of pattern matching on a complex spectral representation (involving fine structure) as in Shamma and Klein (2000). Our formulation in terms of delay would nevertheless remain useful, at least for pedagogical reasons. It is not equivalent to a pattern-matching model based on a magnitude spectrum (e.g., a rate-based tonotopic pattern).

VII. CONCLUSION

1. Cross-channel phase interaction between cochlear channels can produce "synthetic" delays of up to about 30 ms, as a result of interference between phase-shifted impulse responses of cochlear filters. Synthetic delays offer an alternative to neural delays in models of auditory processing.
2. Synthetic delays are limited by the finite duration of cochlear filter impulse responses, that tend to vary inversely with bandwidth: filters are wider at high CFs, so maximum delay is shorter in high-CF than in low-CF channels. The *ratio* of maximum delay to characteristic period (1/CF) is however greater in high-CF than in low-CF channels.
3. Synthetic delays are contingent on phase locking, and ineffective for stimuli with components beyond about 2.5–5 kHz (if this happens to be the limit for humans).
4. Synthetic delays may be used in models of pitch based on autocorrelation, or models of segregation based on cancellation. Limits on delay duration and synchrony impose performance limits for these models. Limits for pitch are roughly consistent with those observed behaviorally, in particular (a) a lower limit of pitch of about 30 Hz for wide-band stimuli, (b) the increase of this limit when stimuli are high-pass filtered, (c) poor pitch discrimination for stimuli containing only harmonics with ranks >10, and more generally (d) the apparent dependency of accurate pitch on *resolvability* of stimulus components. The model is also consistent with the absence of accurate pitch perception for electrical stimulation.
5. The model can be formulated so as to use an initial stage of processing based on between-channel cross correlation that might be shared with other recent models of auditory processing that postulate such interaction.
6. In order to explain the weak pitch of stimuli with "unresolved" harmonics, it may be necessary to postulate an additional, presumably less accurate, source of delay such as produced by axonal conduction or rebound from inhibition.

ACKNOWLEDGMENTS

The authors wish to thank Maria Chait, Christian Kaernbach, Shihab Shamma, Mounya Elhilali, Andrew Oxenham, and Armin Kohlrausch for insightful comments on this manuscript. Katharine Earp provided last minute inspiration.

APPENDIX. SIMULATION DETAILS

Simulations of synthetic delays were performed with a filterbank comprising 2000 channels uniformly spaced on an ERB-rate scale from 25 to 10 000 Hz. Gammatone filters were implemented in the time domain using the "Auditory Toolbox" of Slaney (1993), or in the frequency domain based on the formula

$$Gf = (n-1)\,!\,(1/2)([\cos(\phi) - i\sin(\phi)]\{b/2\pi[i(f+f_c) + b]\}^n + [\cos(\phi) + i\sin(\phi)]\{b/(2\pi[i(f-f_c)+b])\}^n) \quad \text{(A1)}$$

Roll off of synchrony at high frequencies was simulated by applying to the stimulus wave form a low-pass filter consisting of seven cascaded first-order low-pass filters with cutoffs at 4800 Hz. This filter produces a −3 dB cutoff at 2500 Hz followed by a 100 dB/decade rolloff that simulates loss of AN synchrony (Heinz *et al.*, 2001a).

[1]Other shortcomings of autocorrelation models have been put forward (Kaernbach and Demany, 1998; Pressnitzer *et al.*, 2001b, 2004) but they do not relate to the issue of delays, so they will not be discussed in this paper.
[2]Delay shifts both the envelope and the fine structure, whereas phase shift affects only the latter. Simulations (not shown) show that the synthetic delay mechanism is nevertheless effective when phase shifts are produced in this way.
[3]Missing-fundamental stimuli have a low pitch up to about 1400 Hz, at which point the third harmonic, 4200 Hz, falls close to the phase locking limit.

Agmon-Snir, H., Carr, C., and Rinzel, J. (**1998**). "The role of dendrites in coincidence detection," Nature (London) **393**, 268–272.
Behrend, O., Brand, A., Kapfer, C., and Grothe, B. (**2002**). "Auditory response properties in the superior paraolivary nucleus of the gerbil," J. Neurophysiol. **87**, 2915–2928.
Bendor, D., and Wang, X. (**2005**). "The neuronal representation of pitch in the auditory cortex," Nature (London) **436**, 1161–1165.
Bernstein, J. G. W., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability of harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.
Bernstein, J. G. W., and Oxenham, A. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.
Blodgett, H. C., Wilbanks, W. A., and Jeffress, L. A. (**1956**). "Effect of large interaural time differences upon the judgment of sidedness," J. Acoust. Soc. Am. **28**, 639–643.
Cariani, P. A., and Delgutte, B. (**1996a**). "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," J. Neurophysiol. **76**, 1698–1716.
Cariani, P. A., and Delgutte, B. (**1996b**). "Neural correlates of the pitch of complex tones, II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch and the dominance region for pitch," J. Neurophysiol. **76**, 1717–1734.
Carlyon, R. P. (**1998**). "Comments on "A unitary model of pitch perception"

[J. Acoust. Soc. Am. **102**, 1811–1820 (1997)]," 1998). "Comments on "A unitary model of pitch perception" J. Acoust. Soc. Am. **104**, 1118–1121.
Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.
Carney, L. (**1990**). "Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents," J. Neurophysiol. **64**, 437–456.
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.
Carney, L., McDuffy, M. J., and Shekhter, I. (**1999**). "Frequency glides in the impulse responses of auditory-nerve fibers," J. Acoust. Soc. Am. **105**, 2384–2391.
Carney, H., and Yin, T. C. T. (**1988**). "Temporal coding of resonances by low-frequency auditory nerve fibers: Single fiber responses and a population model," J. Neurophysiol. **60**, 1653–1677.
Cash, S., and Yuste, R. (**1998**). "Input summation by cultured pyramidal neuron is linear and position independent," J. Neurosci. **18**, 10–15.
de Boer, E. (**1956**). "On the "residue" in hearing," University of Amsterdam, unpublished doctoral dissertation.
de Cheveigné, A. (**1989**). "Pitch and the narrowed autocoincidence histogram," *Proc. ICMPC*, Kyoto, pp. 67–70 (The Japanese Society of Music Perception and Cognition).
de Cheveigné, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.
de Cheveigné, A. (**1997**). "Concurrent vowel identification III: A neural model of harmonic interference cancellation," J. Acoust. Soc. Am. **101**, 2857–2865.
de Cheveigné, A. (**1998**). "Cancellation model of pitch perception," J. Acoust. Soc. Am. **103**, 1261–1271.
de Cheveigné, A. (**1999**). "Pitch shifts of mistuned partials: A time-domain model," J. Acoust. Soc. Am. **106**, 887–897.
de Cheveigné, A. (**2000**). "A model of the perceptual asymmetry between peaks and troughs of frequency modulation," J. Acoust. Soc. Am. **107**, 2645–2656.
de Cheveigné, A. (**2001**). "Correlation Network model of auditory processing," *Proc. Workshop on Consistent & Reliable Acoustic Cues for Sound Analysis* (Aalborg, Danmark).
de Cheveigné, A. (**2005**). "Pitch perception models," in *Pitch*, edited by C. Plack and A. Oxenham Springe-Verlag, New York, pp. 169–233.
Dehmel, S., Kopp-Scheinpflug, C., Dörrscheidt, G. J., and Rübsamen, R. (**2002**). "Electrophysiological characterization of the superior paraolivary nucleus in the mongolian gerbil," Hear. Res. **172**, 18–36.
Deng, L., and Geisler, C. D. (**1987**). "A composite auditory model for processing speech sounds," J. Acoust. Soc. Am. **81**, 2001–2012.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
Evans, E. F., and Zhao, W. (**1997**). "Periodicity coding of the fundamental frequency of harmonic complexes: Physiological and pharmacological study of onset units in the ventral cochlear nucleus," in *Psychophysical and Physiological Advances in Hearing*, edited by A. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (London, Whurr), pp. 186–194.
Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.
Goldstein, J. L. (**1973**). "An optimum processor theory for the central formation of the pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). "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.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.
Huggins, W. H. (**1952**). "A phase principle for complex-frequency analysis and its implications in auditory theory," J. Acoust. Soc. Am. **4**, 582–589.
Huggins, W. H., and Licklider, J. C. R. (**1951**). "Place mechanisms of auditory frequency analysis," J. Acoust. Soc. Am. **23**, 290–299.
Hurst, C. H. (**1895**). "A new theory of hearing," Proc. Trans. Liverpool Biol. Soc. **9**, 321–353 (and plate XX).
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Jiang, D., Palmer, A., and Winter, I. M. (**1996**). "Frequency extent of two-tone facilitation in onset units in the ventral cochlear nucleus," J. Neurophysiol. **75**, 380–395.
Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. **68**, 1115–1122.
Joris, P. X., van der Heijden, M., Louage, D., Van de Sande, B., and Van Kerckhoven, C. (**2005**). "Dependence of binaural and cochlear "best delay" on characteristic frequency," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*," edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 478–484.
Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.
Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of pitch perception." J. Acoust. Soc. Am. **104**, 2298–2306.
Kalluri, S., and Delgutte, B. (**2001**). "Characteristics of cochlear nucleus onset units studied with a model," "*Computational Models of Auditory Function*," edited by S. Greenberg and M. Slaney (IOS Press, Amsterdam), pp. 29–48.
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.
Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.
Licklider, J. C. R. (**1959**). "Three auditory theories," "*Psychology, a Study of a Science*," edited by S. Koch (McGraw-Hill, New York), **I**, pp. 41–144.
Lin, T., and Guinan, J. J. (**2000**). "Auditory-nerve-fiber responses to high-level clicks: interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.
Lin, T., and Guinan, J. J., Jr (**2004**). "Time-frequency analysis of auditory-nerve-fiber and basilar-membrane click responses reveal glide irregularities and noncharacteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.
Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). "Spatial cross-correlation - A proposed mechanism for acoustic pitch perception," Biol. Cybern. **47**, 149–163.
McAlpine, D., Jiang, D., and Palmer, A. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.
Meddis, R., and Hewitt, M. J. (**1991a**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. J. (**1991b**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2883–2894.
Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., and Oxenham, A. (**2004**). "Sequential F0 comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms," J. Acoust. Soc. Am. **116**, 3038–3050.
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic Press, London).
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Mossop, J. E., and Culling, J. F. (**1999**). "Lateralization of large interaural delays," J. Acoust. Soc. Am. **104**, 1574–1579.
Oertel, D., Bal, R., Gardner, S. M., Smith, P. H., and Joris, P. X. (**2000**). "Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus," Proc. Natl. Acad. Sci. U.S.A., pp. 11773–11779.
Oxenham, A., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.
Oxenham, A., Bernstein, J. G., and Micheyl, C. (**2005**). "Pitch perception of complex tones within and across ears and frequency regions," *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 126–135.

Oxenham, A., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.
Palmer, A. R., and Winter, I. M. (**1996**). "The temporal window of two-tone facilitation in onset units of the ventral cochlear nucleus," J. Nurs. Meas **1**, 12–30.
Patterson, R. D., Allerhand, M., and Guiguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: Amodular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon Press, Oxford), pp. 429–446.
Pilj, S., and Scharz, D. W. F. (**1995**). "Melody recognition and musical interval perception by deaf subjects stimulated with electric pulse trains through single implant electrodes," J. Acoust. Soc. Am. **98**, 886–895.
Pressnitzer, D., de Cheveigné, A., and Winter, I. M. (**2004**). "Physiological correlates of the perceptual pitch shift of sounds with similar waveform autocorrelation," ARLO **5**, 1–6.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001a**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
Pressnitzer, D., Winter, I. M., and de Cheveigné, A. (**2001b**). "Perceptual pitch shift for sounds with similar waveform autocorrelation," ARLO **3**, 1–6.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Saberi, K., Takahashi, Y., Egnor, R., Farahbod, H., and Mazer, J. (**2001**). "Detection of large interaural delays and its implication for models of binaural interaction," J. Assoc. Res. Otolaryngol. **3**, 80–88.
Schroeder, M. R. (**1977**). "New viewpoints in binaural interaction," in *Psychophysics and Physiology of Hearing*," edited by E. F. Evans and J. P. Wilson (Academic Press, London), pp. 455–467.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Shamma, S. A. (**1985a**). "Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve," J. Acoust. Soc. Am. **78**, 1612–1621.
Shamma, S. A. (**1985b**). "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.
Shamma, S. (**2001**). "On the role of space and time in auditory processing," Trends in Cognitive Science **5**, 340–348.
Shamma, S., N., S., and Gopalaswamy, P. (**1988**). "Binaural processing without neural delays," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Wit (Academic Press, London), pp. 135–143.
Shamma, S., and Klein, D. (**2000**). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," J. Acoust. Soc. Am. **107**, 2631–2644.
Shamma, S. A., Shen, N., and Gopalaswamy, P. (**1989**). "Stereausis: Binaural processing without neural delays," J. Acoust. Soc. Am. **86**, 989–1006.
Shera, C. A. (**2001**). "Intensity-invariance of fine time structure in basilar-membrane click responses: Implications for cochlear mechanics," J. Acoust. Soc. Am. **110**, 332–348.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shofner, W. P. (**1999**). "Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms," J. Neurophysiol. **81**, 2662–2674.
Slaney, M. (**1990**). "A perceptual pitch detector," *Proc. International Conference on Acoustics and Speech and Signal Processing (IEEE)*, pp. 357–360.
Slaney, M. (**1993**). "An efficient implementation of the Patterson-Holdsworth auditory filter bank," Apple Computer Technical Report, 35.
Stern, R. M., and Shear, G. D. (**1996**). "Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay," J. Acoust. Soc. Am. **100**, 2278–2288.
Terhardt, E. (**1974**). "Pitch, consonance and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
van der Heijden, M., and Joris, P. X. (**2005**). "Reconstructing the traveling wave from auditory nerve responses," in *Auditory Signal Processing: Psychophysics, Physiology and Modeling*," edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collect (Springer-Verlag, New York), pp. 7–13.
von Helmholtz, H. L. F. (**1877**). *On the Sensations of Tone* (English translation A. J. Ellis, 1885, 1954), (Dover, New York).
Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.
Winter, I. M., and Palmer, A. (**1995**). "Level dependence of cochlear nucleus onset unit response and facilitation by second tones or broadband noise," J. Neurophysiol. **73**, 141–159.

Amplitude modulation sensitivity as a mechanism for increment detection[a)]

Frederick J. Gallun[b)] and Ervin R. Hafter
Department of Psychology, University of California, Berkeley, California 94720-1650

(Received 12 April 2005; revised 4 April 2006; accepted 4 April 2006)

Detectability of a tonal signal added to a tonal masker increases with increasing duration ("temporal integration"), up to some maximum duration. Initially assumed to be some form of energy integration over time, this phenomenon is now often described as the result of a statistical "multiple looks" process. For continuous maskers, listeners may also use a mechanism sensitive to changes in stimulus intensity, possibly a result of inherent sensitivity to amplitude modulation (AM). In order to examine this hypothesis, change detection was investigated in the presence of AM maskers presented at either the same carrier frequency as the target signal or at a distant frequency. The results are compatible with the hypothesis that listeners detect intensity increments by using change-detection mechanisms (modeled here as the outputs of a bank of modulation filters) sensitive to envelope modulation at both low (4–16 Hz) and high (around 100 Hz) rates. AM masking occurred even when the masker was at a carrier frequency more than two octaves above that of the signal to be detected. This finding is also compatible with the hypothesis that similar mechanisms underlie sensitivity to AM (where across-frequency masking is commonly shown) and detection of intensity increments. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200136]

PACS number(s): 43.66.Ba, 43.66.Dc, 43.66.Mk [AJO] Pages: 3919–3930

I. INTRODUCTION

This study examines the detection of a tonal signal added to an ongoing tonal pedestal in terms of two cues. One is the energy added by presentation of the signal. The other is amplitude modulation of the pedestal plus signal, i.e., change in the stimulus envelope. The cue based on energy is commonly thought to reflect the integration of some neural correlate of signal amplitude over time ("temporal integration"). Hughes (1946) was one of the first to report that as the duration of an auditory signal increases, the peak intensity required for detecting its presence decreases, to some maximum duration [for modern reviews, see Gerken *et al.* (1990) and Moore *et al.* (1999)]. This relationship can be demonstrated in animals [e.g., Dooling (1979) and Tougaard (1999)] and is also found for vision (Bartlett, 1965). One common method of expressing the relationship between intensity and detectability across various conditions is by defining "threshold" as that change in intensity required to reach a given level of performance, usually a d' of 1. For the purposes of this discussion, threshold will always be expressed in decibels as $10 \log \Delta I/I$, where ΔI is the change in peak intensity relative to I, the peak intensity of a no-signal trial. For a discussion of alternative measures, see Green (1993). In backgrounds of wideband noise, it is generally found that, for durations between 10 and 250 ms, the threshold for detecting a tonal signal is halved for every doubling of signal duration (Hughes, 1946; Garner and Miller, 1947; Green *et al.*, 1957). Such a rate of decrease is also found for tonal signals presented in quiet (Garner and Miller 1947; Dallos and Olsen, 1964). This relationship can be expressed by plotting threshold versus the logarithm of duration. This provides a slope of −10 dB/decade. The mechanism implied by this temporal integration is one in which all of the input that falls within some fixed time window is summed. As the duration of the signal increases, the signal-to-noise ratio increases and performance improves. Once the signal duration exceeds the window duration, increases in signal duration no longer improve performance. For a tone in quiet, the noise is assumed to be internal.

An alternative to the fixed integration-time model is the proposal that the listener is able to use a temporal window *matched* to the signal duration (Green and Swets, 1966; Viemeister, 1988; Dau *et al.* 1997a, b). In this case, increasing signal duration results in the availability of additional samples of information, thus decreasing sampling error. With an adjustable window matched to the signal duration, performance should be linearly related to the square-root of duration and thus the slope of the function should be −5 dB/decade. This relationship describes those data that show temporal integration when the signal is a tone added to a tonal pedestal (Leshowitz and Wightman, 1971; Green *et al.*, 1979; Viemeister, 1988; Oxenham 1997; 1998; Moore *et al.*, 1999). A further elaboration of this model (Viemeister and Wakefield, 1991) suggests that listeners are not obligated to use a single window to integrate information, but instead are free to take samples from any portion of the stimulus in which the signal is present. This "multiple-looks" method also predicts a −5 dB/decade slope due to reduction of variance.

[a)]Parts of this research were presented at the 26th meeting of the Association for Research in Otolaryngology, the 143rd meeting of the Acoustical Society of America and the 13th International Symposium on Hearing. The data were also presented in F.J.G.'s doctoral dissertation.

[b)]Present address: Hearing Research Center, 635 Commonwealth Ave., Boston University, Boston, MA 02215. Electronic mail: gallun@bu.edu

Either form of the energy model (fixed or adjustable window) runs into difficulties, however, when *detection* of increments is compared with *identification* of changes—distinguishing an increment from a decrement (Macmillan, 1971; Bonnel and Hafter, 1998; Hafter *et al.*, 1998a). Using a tonal signal that is either added to or subtracted from a tonal pedestal, listeners are asked to either detect the increment when a signal is added to a pedestal or, in a separate condition, to identify signals as either added or subtracted. The prediction, if listeners base their decision on stimulus energy alone, is that performance on the identification task should be better than on the detection task. That is because the difference in energy between an increment and a decrement (the steady-state portions of the stimuli in identification) is greater than the difference between an increment and the pedestal alone (the stimuli in detection). For short signals (less than 100 ms), the results are the opposite (Macmillan, 1971; Bonnel and Hafter, 1998; Hafter *et al.*, 1998a), with performance in detection exceeding that in identification. These results are consistent with the idea that listeners detect changes in the envelope of the pedestal rather than the energy *per se*.

One frequently discussed cue to changes in envelope is the presence of off-frequency energy ("spectral splatter"). It is impossible to change the intensity of a tone without generating *some* off-frequency energy (Leshowitz and Wightman, 1971), but the amount of energy depends on the rate at which the intensity changes rather than on signal duration. If the change is sufficiently rapid, a listener using spectral splatter as the sole cue to the presence of a signal might show no change in performance with increasing duration—a slope of 0 dB/decade. Remarkably, this result was obtained by Leshowitz and Wightman (1971) when using "rectangularly gated" signals. From this perspective, the presence of background noise reduces the influence of spectral splatter by decreasing the signal-to-noise ratio in auditory filters adjacent to that containing the signal, thus restoring the improvement in performance expected with increasing duration, whether from energy or multiple looks. If this interpretation of the results is correct, then the influence of spectral splatter should also be reduced by band-pass filtering the stimuli to be detected prior to presentation—and this is exactly what Leshowitz and Wightman (1971) report. Another method of reducing spectral splatter as a cue involves raising the carrier frequency of the signal such that the signal falls in an auditory filter that is wide enough that the energy spilling into adjacent filters is minimized. This method is similar to adding background noise, for then listening in those adjacent filters becomes less informative. Perhaps the most common method used to reduce spectral splatter is gating the signals with gradual onsets and offsets. For signals with carrier frequencies of 400 Hz or greater, 5 ms onset and offset ramps are sufficient to reduce splatter (Moore *et al.*, 1999), whereas a 3 kHz or higher carrier-frequency permits onsets and offsets as rapid as 1 ms without creating a spectral-splatter cue (Oxenham, 1998).

As the studies of detection versus identification (Macmillan, 1971; 1973; Bonnel and Hafter, 1998; Hafter *et al.*, 1998a) all used ramps of at least 10 ms, spectral splatter is unlikely to be the cue that listeners were using. Macmillan (1973) and Hafter *et al.* (1998b) showed essentially the same result using a noise-signal added to a noise carrier. Oxenham (1998), having explored the detection of brief increments and decrements, presented modeling suggesting that listeners are making use of "amplitude modulation energy" in detecting rapid changes in the intensity of a pedestal. Amplitude modulation (AM) refers to fluctuations in the amplitude envelope of the signal and the modulation energy is measured by performing a frequency analysis on the envelope. The work of Houtgast (1989) and Bacon and Grantham (1989) suggest that modulation energy is a quantity to which listeners are sensitive and that it is detected by band-limited filters tuned in modulation frequency. A model based on such a bank of filters has been shown to predict modulation-detection sensitivity quite successfully (Dau *et al.*, 1997a, b; Dau and Verhey, 1999). Using a variation on Dau's model consisting of a single modulation filter tuned to detect the presence of energy in the modulation frequencies between 80 and 150 Hz, Oxenham (1998) successfully predicted performance not predicted by a model based on signal energy alone. Since Oxenham's stimuli were at frequencies of 4 kHz or greater and onsets and offsets were never less than 1 ms, it is unlikely that this modulation energy was detectable as spectral splatter. In addition, Oxenham presented a low-level background noise in order to mask any small changes in energy outside the critical band. Wojtczak and Viemeister (1999) found that listeners' thresholds in a modulation-detection experiment can be predicted from their thresholds in an increment detection experiment, supporting Oxenham's (1998) suggestion that listeners are using both modulation sensitivity and signal energy to detect increments.

This study is a further examination of the possibility that, for ongoing tones, listeners are able to detect changes in intensity by using a cue based on the output of a mechanism that can be modeled as a bank of filters acting in the envelope-frequency domain. By assessing the basic temporal integration performance of a set of listeners and comparing this performance to changes in the output of a bank of modulation filters, the first experiment examined whether the output of a bank of modulation filters can predict the relationship observed between the detectability of intensity increments and the duration of those increments.

The second and third experiments were designed to test the hypothesis that envelope modulations provide a cue to increment detection from a masking perspective. Thus, modulated and unmodulated maskers were presented along with the signals to be detected. By varying the rates of modulation, the importance of energy in different envelope frequency regions could be assessed. In addition, the carrier frequency of the masker was either identical to that of the signal to be detected (experiment 2) or was more than two octaves above that of the signal (experiment 3). The rationale for varying the carrier frequency of the masker is that while energy detection is a phenomenon that should be limited by the energy falling within a critical band (Green and Swets, 1966), there is extensive evidence that interference in the modulation-frequency domain occurs whether or not the tar-

get and masker fall in the same critical band (Yost and Sheft, 1989, 1990; Bacon and Konrad, 1993; Bacon and Moore, 1993; Oxenham and Dau, 2001; Gockel *et al.* 2002).

II. EXPERIMENT 1: TEMPORAL INTEGRATION

Listeners were asked to detect brief increments ("signals") in the intensity of ongoing 477 Hz tones ("pedestals"). The duration of the increment was varied and changes in threshold were measured in the manner described in the following. Predictions of the changes in threshold that should accompany changes in duration were generated based on the output of a bank of modulation filters as well as two versions of an energy-detector model. Model predictions were compared to the performance of human listeners.

A. General Procedures

The following procedures were followed for all experiments described in this paper. The listener's task was to detect a signal that was present on 50% of the trials. The specific signals to be detected and maskers (if any) are described at the appropriate points below. Sounds were generated digitally in a PC and transformed through locally constructed 16-bit digital-to-analog converters with a sampling rate of 50 kHz and a low-pass reconstruction filter set to 20 kHz. Stimuli were presented monaurally over the left channel of Stax SR-5 electrostatic headphones to subjects seated in a sound-attenuated booth. Individual conditions were run in blocks of 50 trials that lasted approximately 3 min, with 25 trials presenting the pedestal alone and 25 the pedestal plus signal. Listeners responded by using a mouse to place the cursor on a computer screen either on the word "change" or on "no change" and pressing a button. Trial-by-trial feedback showed the correct response and then the listener started the next trial, also with a mouse click. Extensive training preceded each experiment and a listener's performance was judged to be stable before collection commenced on the data presented below. Post-training data from each subject for each condition were included in separate calculations of the detection index, d' (Green and Swets, 1966).

B. Methods

Five undergraduate students aged 18–24 were listeners in this experiment. Before testing, all completed a hearing test for detection of tones between 125 Hz and 4 kHz. All listeners were found to have hearing thresholds within 10 dB of published norms in the test ear. Listeners were paid for their participation.

Pedestals were 1000 ms tones with a carrier-frequency of 477 Hz, gated on and off in sine phase with 10-ms onsets and 10-ms offsets in the form of raised-cosine ramps. Pedestals were presented at a peak level of 60 dB SPL. Signals were also tones with a carrier-frequency of 477 Hz, gated on and off with 10-ms, raised-cosine ramps and added to the pedestal in phase. Three signal durations (ramps plus steady portions) were tested:20, 50, and 85 ms. For ease of comparison with experiments 2 and 3, duration is expressed as the half-amplitude duration, which is the duration of the signal calculated between the half-amplitude points of the onset and offset ramps. This yields values of 10, 40, and 75 ms for the three durations tested. The amplitude envelopes of the pedestal-alone stimulus and the three signal-plus-pedestal stimuli are represented in column A of Fig. 1. Column B shows the output produced when each of these envelopes is transformed by a second-order band-pass modulation filter with a center frequency of 4 Hz. Column C shows the changes in output between the pedestal-alone stimulus and each of the three signal-plus-pedestal stimuli for a bank of modulation filters with center frequencies of 4, 8, 16, 32, 64, and 128 Hz. The modulation filters and the method by which changes were calculated are discussed in more detail in Sec. II D.

Signals were delayed such that the temporal midpoints of pedestal and signal coincided. Individual signal levels were set such that adding signal and pedestal in phase resulted in effective signals that, when expressed as $10\log_{10}(\Delta I/I)$, ranged from −9.16 to −3.85 dB. These values correspond to pedestal-plus-signal sound pressure levels ranging from 60.5 to 61.5 dB SPL (i.e., values of ΔL between 0.5 and 1.5 dB). Only one signal level was presented in a block and a minimum of four blocks were obtained with that level. Five signal levels were presented to each listener for each of the three durations and linear fits were made to psychometric functions plotting $10\log_{10}(d')$ against $10\log_{10}(\Delta I/I)$. If the average d' value was below 0.2 or above 3.5, the value was not included in the estimate of the psychometric function. This restricts the fitting procedure to the portion of the data that is actually linear. A threshold for each listener was defined as the level associated with a d' of 1.00 on that listener's psychometric function.[1] The slope of the temporal integration function was calculated by computing the change in threshold as a function of $10\log_{10}(D)$, where D is the half-amplitude duration.

C. Results

Performance was analyzed by comparing threshold for each subject in each condition. Average thresholds and standard deviations across listeners are plotted in Fig. 2. The main effect of duration was reliable ($F_{2,14}=20.360$, $p<0.001$). A planned-comparison between the three durations found no reliable difference between the 40- and 75-ms conditions, but the 10-ms condition was different from both at a level of $p<0.001$. The slope of the temporal integration function plotted against half-amplitude duration is well fitted by a value of −3.9 dB/decade. This slope value is less than the −5 dB/decade reported in the past, but if the slope is calculated versus total duration (i.e., from onset to offset), the result is −5.6 dB/decade, which is similar to what previous investigators have found (Leshowitz and Wightman, 1971; Green *et al.*, 1979; Viemeister, 1988; Oxenham 1997; 1998; Moore *et al.*, 1999). Half-amplitude duration was used in order to allow comparison with the results of experiments 2 and 3, in which total duration is a less useful metric.

D. Modulation-based modeling

Ewert and Dau (2000) showed that differences in the output of filters tuned in envelope frequency can be used to

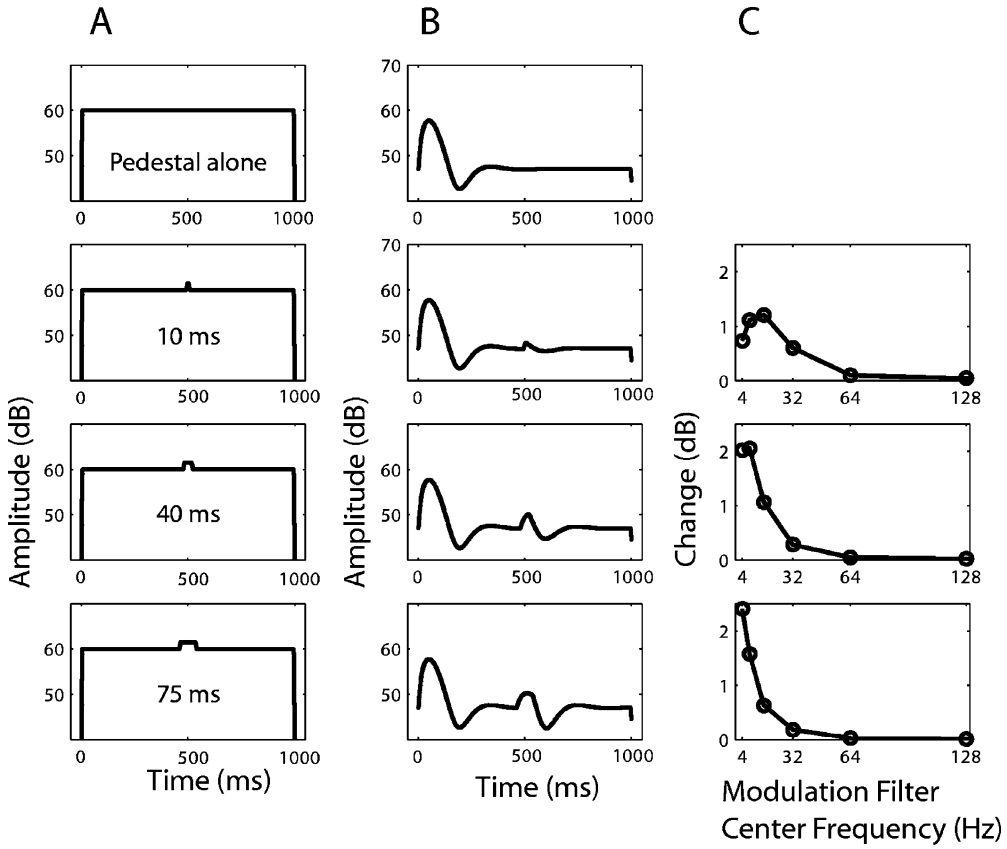


FIG. 1. The amplitude envelopes of the signals and pedestals presented in experiment 1. (A) Plots of the envelopes as a function of time. (B) Plots of the output of a 4-Hz modulation filter for the stimuli in (A). (C) The amount of change (dB) that occurs in the output of modulation filters of various center frequencies when each of the signals is added to the pedestal (see Sec. II D for details on the filter shapes and the temporal windows across which the changes were measured).

accurately predict performance in an AM detection task. In that model, performance is predicted to be based on the output of a modulation filter tuned to the AM rate to be detected. When a signal is added, the output of this filter increases relative to when no signal is present. Threshold is predicted to occur at the signal level that leads to a 1 dB change in output. A similar analysis can be applied to the duration data of experiment 1. Unlike with AM detection, however, adding a brief signal introduces variations in the output of a range of modulation filters (see Fig. 1, column C). In addition, the filter output varies over time in accordance with the duration of the signal (see Fig. 1, column B). In order to capture this variation over time and across modulation filters, several modifications of the model were required. In general, however, the model was kept as close to that of Ewert and Dau (2000) as possible. In particular, the filters are second-order band-pass filters with a Q value of 1. The center frequencies are 4, 8, 16, 32, 64, and 128 Hz.

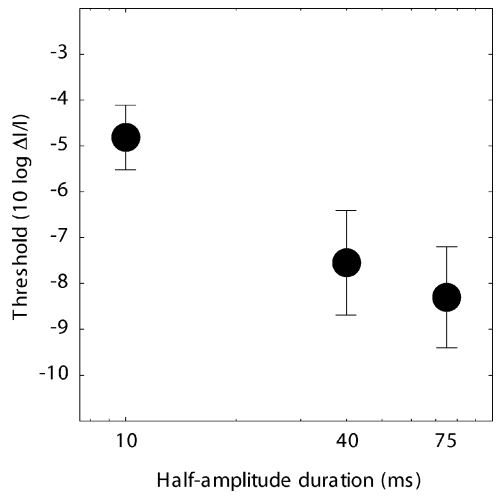


FIG. 2. Average threshold values (see the text) for five listeners as a function of increment duration. Error bars indicate ±1 standard deviation.

The output of each modulation filter was obtained for each of the envelopes of the signals used in experiment 1 by using Matlab code made available by Ewert and Dau and corresponding to the filters used in Ewert and Dau (2000). Column B of Fig. 1 represents the output of a 4-Hz modulation filter to the signal envelopes shown in column A. In order to predict listener performance from this output, a temporal window was used to integrate the output over time. In order to simulate internal noise, a constant small amplitude value (0.04 relative to a maximum of 1) was added to each sample (of which there were 20,000 in each one-second stimulus). Various window sizes were examined and the best fit to the data (considering all three experiments) was obtained with a variable-duration window. The temporal center of the window was aligned with the temporal center of the signal and the duration of the integration window was adjusted to be that of the signal plus 10 ms. The change in dB

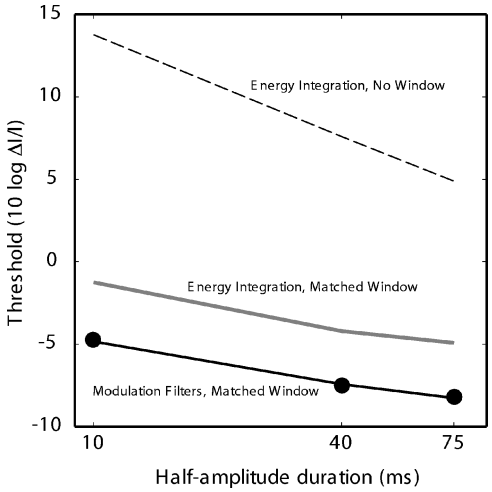


FIG. 3. Model predictions (lines) for the threshold data (black circles) from experiment 1. Solid black line: modulation filter-bank model. Gray line: unfiltered envelope model with a temporal integration window matched to signal duration. Dashed black line: unfiltered envelope model with no window limiting temporal integration.

was calculated as the ratio of the integrated values of the signal-plus-pedestal and pedestal-alone outputs. Column C of Fig. 1 shows the dB change associated with 4, 8, 16, 32, 64, and 128 Hz filters for the three signal durations used in the first experiment (all input signals had a ΔL of 1.5 dB). The output of the model for a given signal was the maximum difference in dB that occurred at the output of any filter. As can be seen in column C of Fig. 1, the maximum change was at lower frequencies for longer signals.

Ewert and Dau (2000) used an adaptive tracking simulation to obtain the signal level that resulted in a 1 dB change in the output. A second modification made for the modeling reported here is that, since the listeners in experiment 1 performed the task at five different levels from which a threshold was estimated by linear interpolation, linear interpolation was used in the modeling as well. In order to match the method used in experiment 1, signal level was varied across the same five values (ΔL of 0.5–1.5 dB), resulting in a set of output values for each signal duration. The output values were then integrated across the window duration and transformed into changes in dB on the basis of the ratio of the signal-present output to the signal-absent output. These changes in dB were then plotted against the signal level expressed as the logarithm of $\Delta I/I$. Linear regression was used to obtain predictions of which $\Delta I/I$ value resulted in a 1 dB change in the integrated output. These $10 \log_{10}(\Delta I/I)$ values appear as the solid black line plotted in Fig. 3. For comparison, the thresholds predicted by using the same matched integration window directly on the stimulus envelopes with no modulation filtering (i.e., matched-window energy integration) is plotted as a solid grey line. Threshold predicted on the basis of the entire duration of the stimuli with no filtering are plotted as a dashed black line (energy integration, no window). As can be seen, the slopes of the two matched-window functions are similar but the function based on the outputs of the modulation filters gives a much more accurate fit to the data. Energy integration based on the entire duration of the stimulus gives the traditional 10 dB/decade slope but the values are much higher than were obtained in this study.

E. Discussion

The fit to the data shown in Fig. 3 is quite good for the modulation filters with a matched window. Given that Oxenham (1998) achieved similar success by using modulation frequencies between 80 and 150 Hz it is perhaps surprising that the filters with the greatest change in output for these stimuli were between 4 and 16 Hz (see Fig. 1, column C). One potential reason for the difference between the modeling results is that it was the short duration (2–20 ms) signals in Oxenham's study that were well fit by a modulation-based model. Detectability of signals longer than 20 ms required resorting to an "energy" mechanism. As the duration of the signal is inversely related to the spread of energy to distant frequencies, both in the carrier frequency domain ("spectral splatter") and in the envelope frequency domain, it is not surprising that very short signals were detected on the basis of modulation-energy at higher frequencies. In addition, it is significant that the signal values needed by Oxenham's (1998) listeners to reach threshold performance at the longer durations were well above those found in this study (ΔL of 3–5 dB). The reason for this is may be the introduction of low-level background noise in all conditions. The distribution of the modulation-energy in broadband noise stretches from 0 to Δf, where Δf is the bandwidth, with an energy distribution that decreases with increasing modulation frequency [c.f., Ewert and Dau (2000)]. For this reason, the degree of modulation masking from broadband noise is greatest at the lowest frequencies. As the distribution of modulation energy in signals of varying durations is dependent on the duration of the signal (see Fig. 1) it is possible that the longer signals were not detectable on the basis of modulation energy. Similar threshold levels to those of Oxenham (1998) have been reported by Leshowitz and Raab (1967) who also used a low-level background noise. On the other hand, Viemeister (1988), Leshowitz and Wightman (1971), and Jeffress (1975), none of whom used background noise, reported similar threshold values to those found in this study (ΔL of 0.2–2 dB for durations between 10 and 200 ms). In accordance with this analysis, Oxenham (1997) reported elevated thresholds for a variety of on and off-frequency noise maskers relative to thresholds in quiet. Those results were not well described either by the masking of spectral splatter or upward spread of masking.

The main result of this modeling is to suggest that if listeners are provided with a sufficiently long-duration pedestal with very little modulation energy and with no background noise, then the output of a bank of modulation filters could provide an effective cue to the presence of increments. Furthermore, it seems to be the case that, for signals with durations greater than 10–20 ms, the output of modulation filters tuned to low-frequencies provide a more effective cue than the output of filters tuned to high-frequency modulation. Experiment 2 was designed to further examine the relative importance of modulation energy in different frequency regions by introducing modulation masking.

III. EXPERIMENT 2: MODULATION MASKING, SAME CARRIER FREQUENCY

Experiment 2 examined the relative masking caused by the presence of AM chosen to selectively interfere with the signal-to-masker ratios present in the modulation filters used in the modeling of the results of experiment 1. To investigate whether rise-fall time interacts with the rate of AM, two different ramps were used. The main difference between the first and second experiments was the introduction of a masker at the same carrier frequency as the pedestal (477 Hz). The masker was either unmodulated ("0 Hz") or was modulated at 4, 48, or 96 Hz. The peak level of the unmodulated masker was set to be the same as the peak levels of the modulated maskers so that masker energy was always decreased by the addition of modulation. Thus, a strictly energy-based model would predict *better* performance in the presence of modulation whereas an AM-based model would predict worse.

A. Methods

All five listeners from experiment 1 participated. The peak level of the pedestal was set to the 60 dB level that was used for the pedestals in the first experiment. Pedestals and signals were tones with 477-Hz carrier frequencies. Pedestals were 1000-ms in duration and had peak levels of 60 dB SPL. The increase in the levels of the stimulus when the signal was present (relative to the peak level of the pedestal alone) were $10\log_{10}(\Delta I/I) = -2.33$, -1.09, or 0 dB (overall levels of 62, 62.5, or 63 dB, or ΔL's of 2–3 dB). Listeners completed all conditions at two of the three signal levels. The signal with a half-amplitude duration of 10 ms had a 20 ms total duration with onset and offset ramps that were each 10 ms. The signal with a half-amplitude duration of 45 ms had a total duration of 85 ms and ramps that were each 40 ms. Signals were delayed such that the temporal centers of the pedestal and signal envelopes coincided. Signals were added in phase.

Maskers were also 477-Hz tones with the same duration and phase as the pedestals. All maskers (modulated or unmodulated) had a peak level of 60 dB SPL. Modulated maskers were sinusoidally amplitude-modulated at rates of 4, 48 or 96 Hz. Modulation was at a depth of 80% (corresponding to minima of 41 dB SPL and maxima of 60 dB SPL). AM rate was always a multiple of two so that there was a maximum at the temporal center of both pedestal and signal. The equation for the AM maskers (Y_m) is the following, where $X(t)$ is the function that defines the temporal envelope of the pedestal, $m=0.08$, $\phi=0$, $f=477$ Hz, and f_m takes on one of the following values: 0 (unmodulated), 4, 48, or 96 Hz:

$$Y_m(t) = [X(t)\sin 2\pi ft]\{1 + m[\cos(2\pi f_m t + \phi)]\}. \quad (1)$$

The envelopes of a 4-Hz modulated masker added to the pedestal alone and a 4-Hz modulated masker added to a pedestal with each of the two signal types are shown in column A of Fig. 4. The output of a 4-Hz modulation filter for those envelopes are shown in column B. Column C shows the outputs for a 16-Hz modulation filter.

B. Results

The results of this experiment, averaged across listeners and signal levels, are shown in Fig. 5 and the full data set appears in Table I. Because each listener had completed all conditions, a repeated-measures analysis of variance was conducted.[2] The signal levels were treated as a covariate for the analysis. A main effect of AM rate was found for the 10-ms increment ($F_{3,21}=13.869$, $p<0.001$) and there was no interaction with signal level. Planned-comparison t-tests showed that for all 4, 64, and 96 Hz AM, levels of performance were significantly different from that found with the unmodulated masker ($p<0.001$). A main effect of AM rate ($F_{3,24}=25.022$, $p<0.001$) was also found for the 45-ms increment. In this case, however, planned -comparison t-tests showed that although 4 Hz is significantly different from the unmodulated condition ($p<0.001$), the higher rates are not ($p>0.05$). The results indicate that (1) envelope modulation is an important cue for increment detection and (2) the duration (or perhaps the onset/offset ramps) of the signal can effect which envelope frequencies are weighted the most heavily in the listener's decision process.

C. Model predictions

As in experiment 1, the envelopes of the signals presented to the listeners were processed by a bank of second-order band-pass modulation filters. Signal levels were chosen to be representative of those used in the experiment (signal-plus-pedestal level of 62.5 dB for the 10-ms signal and 62 dB for the 45-ms signal). As in the modeling for experiment 1, the changes in the outputs of the modulation filters were used to generate model predictions. In this case, the maximum change in dB across filters was used to directly predict listener performance expressed as d'. This is similar to the assumption made in experiment 1 that threshold ($d'=1$) corresponds to a change in the filter output of 1 dB. It is also similar to the assumption made in Ewert and Dau (2000) that the threshold signal level is that which results in a 1 dB change in the output of a filter tuned to the modulation frequency being detected. In all cases, the same temporal integration windows were used as in experiment 1. For comparison, the unfiltered envelope (with the same temporally matched filter) was also used to predict the listener data.

Figure 6 contains the data from Fig. 5 as well as predictions from the no-filter condition (an energy model) and the predictions from the modulation filter-bank model. The abscissa plots the modulation rate and the ordinate plots d′ for the data and the largest change in dB for the modeling. The most obvious result is that although the energy model fails to predict the appropriate changes in performance with modulation rate, the modulation filter-bank does quite well at capturing the relative levels of performance. The absolute levels predicted for the 10-ms signal are too high, however.

D. Discussion

From the perspective of purely energetic masking, the greater energy in the unmodulated masker would be expected to produce more masking, and yet it did not. These results are in accord with previous reports (Macmillan, 1971; 1973;

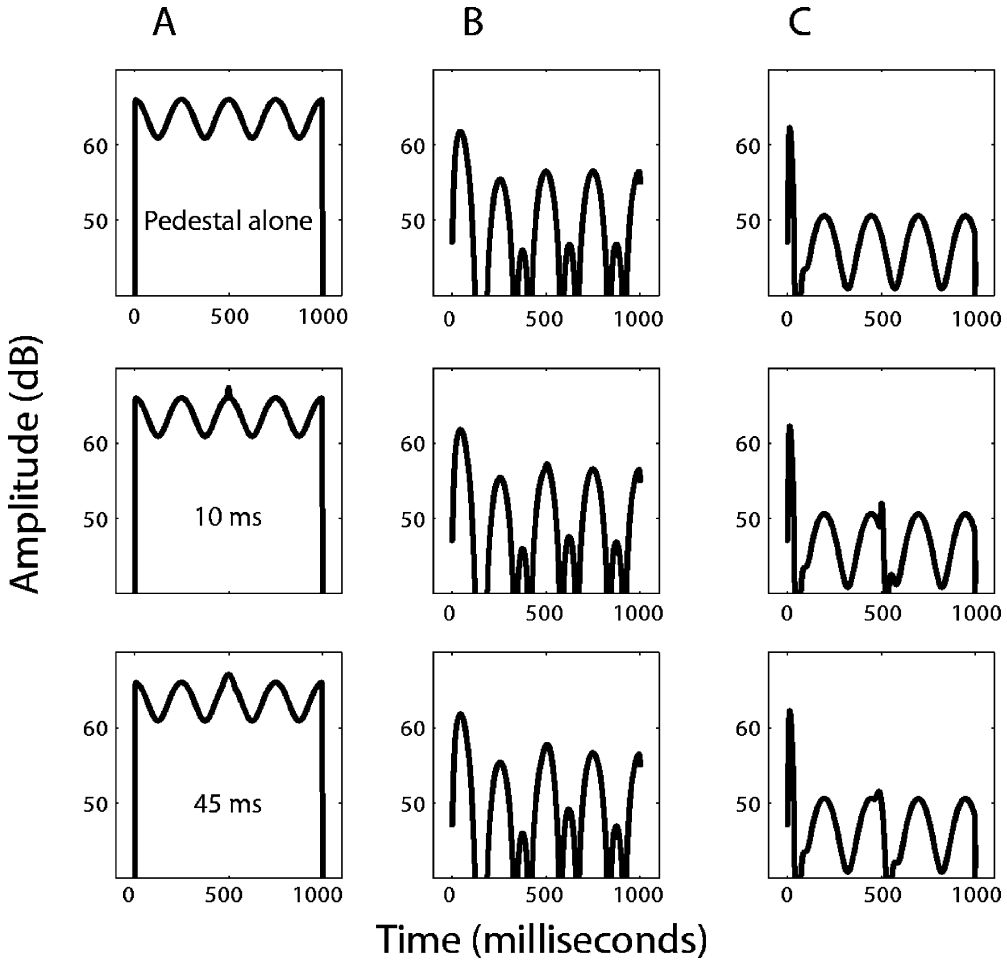


FIG. 4. (A) The amplitude envelopes of the pedestal alone (amplitude modulated at 4 Hz) and the 4-Hz modulated pedestal plus the two signals presented in experiment 2. (B) Plots of the output of a modulation filter centered at 4 Hz for the stimuli in (A). (C) Similar plots for the output of a 16-Hz modulation filter.

Bonnel and Hafter, 1998; Hafter *et al.*, 1998a; Oxenham, 1998), suggesting that listeners detect changes in the intensity of an ongoing pedestal by using information beyond that obtained by estimating stimulus energy at various points in time. These results are consistent with the hypothesis that listeners are using sensitivity to the output of a bank of modulation filters to detect changes. In addition, the difference in the masking patterns for the two signal types (10 and 45 ms) is also predicted by the differences in the output patterns of the modulation filters. The shorter-duration signal adds energy to higher modulation frequencies, while the longer-duration signal adds modulation energy primarily in the low-frequency region. Consequently, the short-duration signal is masked equally by all three AM rates whereas the long-duration signal is effectively masked by the 4-Hz masker but not by the others.

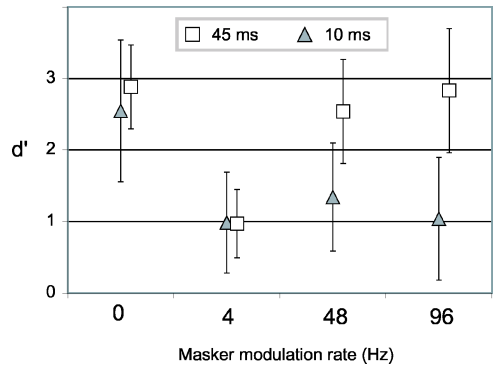


FIG. 5. Average performance values for the listeners in Experiment 2. Error bars indicate ± one standard deviation across listeners.

One alternative explanation for the difference between the masking patterns for the short and long-duration signals is that listeners were detecting spectral splatter for the shorter signal. Since 96-Hz AM generates sidebands that fall outside the critical band centered on 477 Hz (the width is approximately 70 Hz), it is plausible to imagine that the masking observed at the highest masker modulation rate was caused by energetic masking in the carrier-frequency domain rendering the spectral splatter undetectable. Experiment 3, by removing the modulation masking from the carrier-frequency region of the signal, removed this possibility.

IV. EXPERIMENT 3: OFF-FREQUENCY MODULATION MASKING

The results of experiments 1 and 2 are clear indications that modulation sensitivity can be usefully considered to be part of the information listeners use to detect increments added to long-duration ongoing tonal pedestals. Experiment 3 extended the results by testing more modulation rates as

TABLE I. Listener performance for the two conditions in Experiment 2 (477-Hz masker). The listener data for which no unmodulated masker performance could be calculated (250 trials with no errors) are indicated by an asterisk.

		Masker modulation rate			
Listener	$\Delta I/I$ (dB)	0	4	48	96
20 ms signal, 10 ms ramps					
L1	0	1.36	1.08	0.95	0.48
L1	−1.09	1.16	0.39	0.61	0.18
L2	0	2.83	0.46	1.48	1.63
L2	−1.09	1.79	0.37	0.74	0.25
L3	0	3.93	2.16	2.32	2.21
L3	−1.09	2.07	1.65	1.78	0.41
L4	0	3.61	0.71	1.48	1.73
L4	−1.09	2.40	0.00	0.25	0.10
L5	−1.09	3.80	1.59	2.64	2.25
L5	−2.33	2.53	1.44	1.17	1.16
	Mean	2.55	0.99	1.34	1.04
	Std. Dev.	0.99	0.70	0.76	0.86
85 ms signal, 40 ms ramps					
L1	−1.09	2.51	1.79	2.07	1.84
L1	−2.33	1.81	1.02	1.39	1.38
L2	−1.09	2.90	1.21	3.16	3.80
L2	−2.33	3.15	1.52	2.23	2.81
L3	−1.09	⋯	0.89	3.97	4.38
L3	−2.33	3.23	0.61	1.93	2.61
L4	−1.09	3.81	0.38	3.53	3.44
L4	−2.33	2.56	0.40	2.30	3.88
L5	−1.09	3.39	0.94	3.23	3.39
L5	−2.33	2.66	0.96	3.11	2.42
	Mean	2.889	0.98	2.55	2.84
	Std. Dev.	0.59	0.48	0.73	0.87

well as adding a third signal: one with a long duration but with short onset/offset ramps. This signal allows a comparison between a short overall signal and one that simply has short onsets and offsets. In addition, the pedestal remained at 477 Hz, but the modulation was imposed on a masking tone presented more than two octaves away, at 2013 Hz. Any masking observed in experiment 3 provides additional evidence that increment-detection is a phenomenon that seems to occur in the same domain as the detection of amplitude modulation.

A. Methods

The same five listeners participated as in experiment 2, with two additional listeners for some conditions. Three signals were used, two of 85 ms total duration and one of 20 ms, all with a carrier frequency of 477 Hz. The short signal and one of the long signals had raised-cosine onset and offset ramps of 10 ms. The remaining long signal had raised-cosine onset and offset ramps of 40 ms. This resulted in three values of half-amplitude duration: 10, 45, and 75 ms. These signals correspond to the signals with these half-amplitude durations in the previous experiments. All were delayed and added in phase such that they produced increments that were temporally centered in a 1000-ms long pedestal with a carrier-frequency of 477 Hz. The maskers were sinusoidally amplitude-modulated [see Eq. (1)] tones of 1000-ms duration, with a carrier-frequency of 2013 Hz, summed with the pedestals prior to presentation. Modulation frequencies were 0, 4, 8, 12, 24, 32, 48, 64, and 96 Hz. Data were also collected in a condition in which no masker was present. Once the signals had been added to the pedestal, the peak intensities of the signal-plus-pedestal stimulus, relative to the pedestal intensity alone, were $10\log_{10}(\Delta I/I) = -5.86$, -3.85, or -2.33 (producing peak signal-plus-pedestal levels of 61, 61.5, or 62 dB). As in experiment 2, signal levels were set individually and kept constant across conditions. Again some listeners completed full sets of conditions at multiple levels (see Table II).

B. Results

Results averaged across listeners are plotted for all three signal types in Fig. 7 and the data appear in Table II. A repeated-measures analysis of variance was performed on the data, with each listener at each signal level entered independently. For the 10-ms signal, since three different signal levels were used, signal level was added as a covariate. The main effect of AM rate was significant ($F_{10,90}=7.337$, $p = < 0.001$) and there was no interaction with signal level. Planned-comparison t-tests revealed no difference between the performance obtained in the unmasked and 0 Hz (unmodulated) masker conditions ($p > 0.05$). This shows that the addition of an unmodulated pedestal at 2013 Hz did not affect performance, which is what would be predicted by an energy-based model in which energy was estimated indepen-

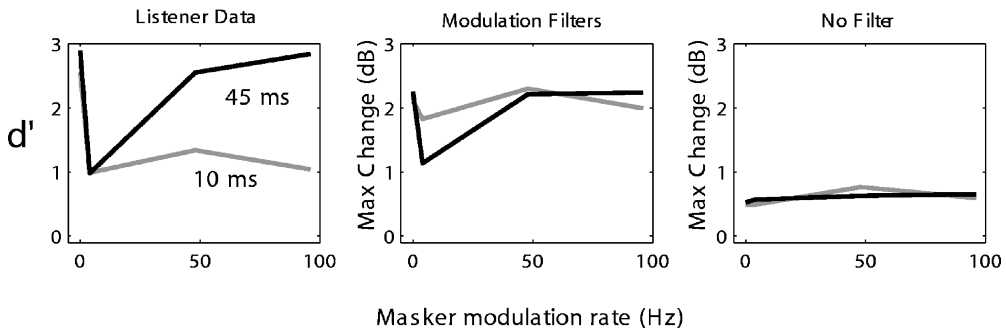


FIG. 6. Model predictions for the performance data (left panel) from experiment 2. For ease of presentation, performance and predictions for the unmodulated masker are plotted at the "0-Hz" point. Modulation filter model predictions are the maximal change in output across a bank of modulation filters. Energy model predictions are the overall change in output across the temporal integration window used in the modulation model. See the text for details.

TABLE II. Listener performance for the three conditions in experiment 3 (2013-Hz masker)

			Masker modulation rate									
Listener	$\Delta I/I$ (dB)	No masker	0	4	8	12	16	24	32	48	64	96
20 ms signal, 10 ms ramps												
L1	−2.33	2.75	1.88	1.54	1.17	1.62	1.49	2.04	1.98	1.35	1.93	1.16
L1	−3.85	1.20	1.04	1.02	0.67	0.58	1.37	1.03	1.34	1.64	1.11	0.99
L2	−2.33	2.56	1.77	1.00	0.82	1.04	1.27	1.74	2.64	1.06	1.28	1.09
L2	−3.85	1.23	1.09	0.10	0.75	0.94	0.54	1.00	1.27	0.65	1.02	0.12
L3	−3.85	2.21	2.77	1.70	2.15	2.51	2.47	2.48	1.83	2.43	2.69	1.70
L3	−5.87	1.78	1.64	1.81	2.11	2.10	2.57	2.30	1.79	1.58	1.61	0.79
L4	−3.85	2.04	1.99	1.18	1.73	1.33	1.51	2.33	2.16	2.56	1.80	1.95
L5	−3.85	2.60	2.39	1.46	0.72	1.68	1.56	2.30	2.19	2.28	2.70	1.47
L5	−5.87	1.57	1.77	0.81	0.37	0.91	0.67	1.18	1.32	1.32	1.69	1.21
L6	−5.87	1.79	2.08	0.22	1.11	1.40	1.03	1.30	1.13	0.97	1.35	0.79
	Mean	1.97	1.84	1.08	1.16	1.41	1.45	1.77	1.76	1.58	1.72	1.13
	Std. Dev.	0.56	0.53	0.58	0.63	0.59	0.66	0.59	0.49	0.65	0.59	0.52
85 ms signal, 10 ms ramps												
L1	−5.87	2.26	1.55	1.54	1.56	2.48	1.90	2.31	2.09	2.29	1.93	2.12
L2	−5.87	3.03	2.36	0.39	2.36	2.92	2.51	2.37	2.38	2.14	2.73	2.01
L3	−5.87	3.06	2.07	1.04	1.63	2.96	2.49	2.60	2.48	2.35	1.93	2.01
L4	−5.87	3.36	3.68	0.83	2.96	3.31	2.44	1.90	2.42	2.93	2.78	2.81
L5	−5.87	2.14	1.65	0.05	0.80	1.15	0.94	1.14	1.08	1.25	1.20	0.79
L7	−3.85	2.56	2.58	0.20	0.88	1.53	2.26	1.77	1.86	2.24	2.33	1.91
	Mean	2.73	2.31	0.67	1.70	2.39	2.09	2.01	2.05	2.20	2.15	1.94
	Std. Dev.	0.49	0.78	0.57	0.84	0.86	0.61	0.53	0.53	0.54	0.59	0.65
85 ms signal, 40 ms ramps												
L1	−5.87	1.56	1.54	0.58	1.02	0.87	1.17	1.31	1.10	1.66	1.46	1.86
L2	−5.87	1.62	1.81	0.39	0.80	1.83	1.95	2.03	1.69	1.88	1.42	1.37
L3	−5.87	1.56	1.89	0.50	1.17	1.69	1.78	1.89	1.54	1.82	1.39	0.92
L4	−5.87	1.19	1.22	0.24	0.55	0.70	0.94	0.83	1.14	0.50	0.93	0.80
L5	−5.87	2.78	2.53	1.04	1.93	1.96	2.13	1.77	1.77	1.95	1.83	2.03
	Mean	1.74	1.80	0.55	1.09	1.41	1.59	1.56	1.45	1.56	1.41	1.40
	Std. Dev.	0.61	0.49	0.30	0.52	0.58	0.51	0.49	0.31	0.60	0.32	0.55

dently for each critical band. Further planned-comparison t-tests showed that the 4, 8, 12, and 96 Hz AM produced performance reliably different from the 0 Hz (unmodulated) masker ($p<0.01$), but performance with the 16, 24, 32, 48, and 64 Hz AM maskers did not ($p>0.05$). As was found in experiment 2, this shows that the 10-ms signal is detected based on a combination of high and low envelope frequencies.

For the 45-ms signal (an 85-ms signal with 40-ms ramps), the main effect of AM rate was significant ($F_{10,40}=7.816$, $p<0.001$). Planned-comparison t-tests showed that the difference between the unmasked and 0 Hz (unmodulated) masker conditions was not statistically reliable ($p>0.05$). In this case, however, the differences between the unmodulated masker and the 4, 8, 12, 32, and 64 Hz modulated masking conditions were all reliable (4 and 8 Hz, $p<0.01$; 12, 32, and 64 Hz, $p<0.05$). Interestingly, this suggests that listeners are capable of using a range of envelope frequencies to perform the task. These results are not the same as those obtained in experiment 2, in which only the lowest modulation frequencies produced substantial masking for the 45-ms signal.

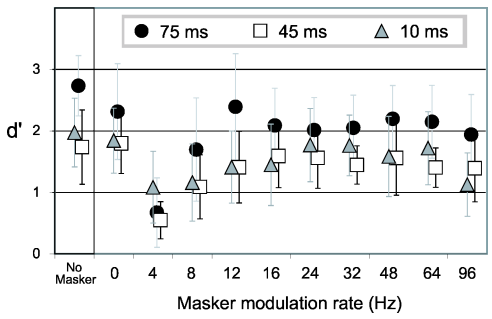


FIG. 7. Average performance values for the listeners in experiment 3. Error bars indicate ± one standard deviation across listeners.

For the 75-ms signal (an 85-ms signal with 10-ms ramps), the main effect of AM rate was significant ($F_{10,50}=10.836$, $p<0.001$). Planned-comparison t-tests showed that the difference between the unmasked and unmodulated masker conditions was not statistically reliable ($p>0.05$). In fact, the only difference that reached significance was that between the unmodulated and the 4 Hz ($p=0.011$). These results are in agreement with the pattern seen in experiment 2, but are at odds with the findings for the 45-ms signal in this experiment. Examination of the patterns displayed by the individual listeners (as can be seen in Table II) suggests that

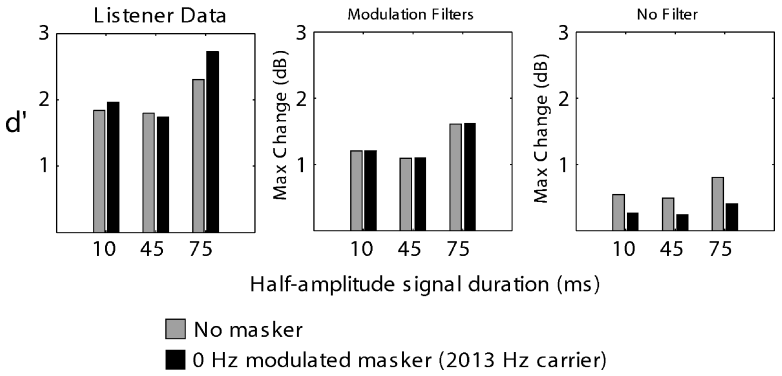


FIG. 8. Model predictions for a subset of the performance data (left panel) from experiment 3. The grey bars indicate detection performance (or predictions) for the three signal durations when no masker was present. The black bars indicate performance (or predictions) when the masker was an unmodulated tone presented at a carrier frequency of 2013 Hz. The models are the same as those plotted in Fig. 6.

there may be differences between listeners that contribute to the variability of masking patterns across the different signal durations.

C. Model predictions

The model predictions were generated in the same manner as for experiment 2. Central to the success of the model is the fact that modulation information is combined across carrier frequencies as suggested by the modeling of Dau and Verhey (1999) and Chi *et al.* (1999). This model structure is also supported by the psychophysical results on studies of AM masking across carrier frequency (e.g., Yost and Sheft, 1989). Figure 8 compares the predictions of the modulation model and the energy model to the listener data for the no masker condition and the unmodulated masker condition. As is clear, the no filter (energy) model predicts a substantial effect of the unmodulated masker. The reason for this is that the predictions are based on a summation of the envelopes despite the fact that they are presented at distant carrier frequencies. Despite this fact, the modulation filter-bank predicts no difference whatsoever. The listener data resemble the prediction of the modulation filters much more closely than they do the prediction of the energy model. The standard deviations shown in Fig. 7 and the individual data presented in Table II make it clear that the apparent variations with masker for the listener data are due to variability across listeners. The various individual listeners were as likely to perform better with no masker as with an unmodulated masker at 2013 Hz.

The model predictions for the off-frequency modulation maskers appear in Fig. 9. These data are a strong indication of the value of considering a model based on a bank of modulation filters as an explanation for these data. Although the energy-integration model fails to predict any change in performance with modulation rate, the modulation filter-bank captures the drop in performance in the presence of 4-Hz AM and the subsequent recovery at higher rates.

D. Discussion

The finding that AM maskers can interfere with increment detection at a distant carrier frequency suggests that whatever mechanism underlies amplitude modulation sensitivity is effective for increment detection as well. As predicted by the changes in the output of low-frequency modulation filters to unmasked signals (Fig. 1, column C), the greatest masking occurred for the lowest AM rate tested. As in experiment 2, however, there was substantial masking at the highest AM rate for the short-duration signal. An explanation based on masking of spectral splatter is not plausible for a masker centered at 2013 Hz, however. In general, the change in output of a bank of modulation filters was successful at predicting the patterns of masking observed in this experiment as in the previous two. The largest changes in output again came from filters tuned to low envelope frequencies.

The most striking result of this study is certainly the masking of a tone at a carrier frequency of 477 Hz by a masker two octaves higher in frequency. This result is similar to those obtained in AM detection experiments (Yost and Sheft, 1989, 1990; Bacon and Konrad, 1993; Bacon and Moore, 1993; Oxenham and Dau, 2001; Gockel *et al.* 2002). Nonetheless, it stands quite alone in the increment detection literature and is, by itself, a compelling argument for the need to reexamine the energy model.

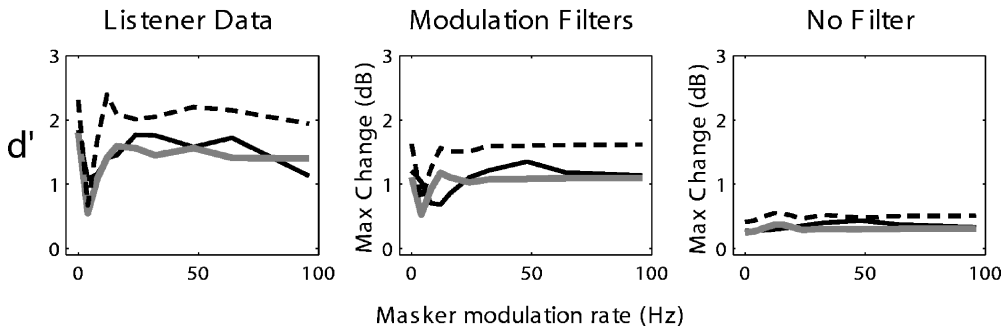


FIG. 9. Model predictions for a subset of the performance data (left panel) from experiment 3. The models are the same as those plotted in Fig. 6.

V. GENERAL DISCUSSION

In detecting increments added to ongoing tones, listener performance is substantially decreased by the presence of amplitude-modulated maskers at low modulation rates. One of the most striking findings was that in experiment 3 amplitude modulation hurt performance even when it was not present at the carrier frequency of the signal. This suggests that the AM filterbank model originally proposed by Bacon and Grantham (1989) and by Houtgast (1989) could be a very effective method for detecting increments. The model used in this article, based on that of Ewert and Dau (2000), supports this idea in that it very effectively predicts performance for most of the experiments described.

Given the success of the simple model of Ewert and Dau (2000), it is possible that a more detailed model might do even better. Two additional models (Dau *et al.*, 1997a, b; Chi *et al.*, 1999) were also explored, but neither was able to substantially improve on the performance of the current, simpler model. Chi *et al.*(1999) has the advantage that it is based upon actual physiological patterns of responding observed in the auditory cortex. Quantitatively (and qualitatively), however, it was less successful than the filters proposed by Ewert and Dau (2000). While it is an important goal of auditory science to bridge the gap between psychophysics and physiology, the current state of knowledge on the physiological basis of modulation sensitivity (recently reviewed by Joris *et al.*, 2004) suggests that there is still much about the physiology that is unknown. In fact, it has even been suggested that a single unit's sensitivity to modulation can vary considerably based on the stimuli presented and time scales over which activity is analyzed (Nelken *et al.*, 2004). Future modeling work will probably need to embrace this non-linear response pattern, but currently there is little work that is capable of capturing the physiological responses let alone extending these patterns to predict psychophysical performance. Consequently, the fact that the model of Ewert and Dau (2000) captured so much of the psychophysical results is certainly a remarkable result.

VI. SUMMARY

Three experiments were conducted on the ability of human listeners to detect changes in the intensity of a 477-Hz tonal pedestal. In the first, increment duration was varied and listener sensitivity was found to increase with duration, with a greater change coinciding with increasing total duration from 20 to 50 ms than with increasing total duration again to 80 ms. This pattern of improvement was predicted by a model based on changes in the output of a bank of modulation-sensitive filters. In the second and third experiments, amplitude modulation was presented throughout the observation interval in order to selectively mask individual modulation filters. On-frequency AM (at the same carrier-frequency as the signal) was found to mask increments with a total duration of 20 ms at rates of 4, 48, and 96 Hz, but 85-ms increments with 40-ms onsets and offsets were only masked by 4-Hz AM. Off-frequency AM (imposed on a carrier frequency of 2013 Hz) was most effective as a masker when the rate was 4 Hz, but there was some effect of 96-Hz AM as well, especially for brief signals. The modulation-based model was more successful in predicting the masking patterns than was a model based on energy, suggesting the an "energy detector" is not the appropriate description of the mechanism by which human listeners demonstrate sensitivity to brief changes in the intensity of a tone.

ACKNOWLEDGMENTS

The authors are extremely grateful to Dr. Torsten Dau and Dr. Shihab Shamma for stimulating discussions on this topic as well as for providing Matlab code. They thank Dr. Anne-Marie Bonnel for extensive discussions of the methodological and theoretical issues related to this research. Dr. H. Steven Colburn provided extremely useful analysis of the model results and parameters as well as encouragement. They are also very grateful to Dr. Andrew Oxenham, the associate editor, as well as to Dr. Christian Lorenzi and an anonymous reviewer, all of whom made comments that were central in developing the model presented here. This work was supported by Grant No. DC 000087 from NIH/NIDCD.

[1]Although thresholds are often obtained with adaptive procedures, estimating threshold from psychometric functions was found to give more stable results in preliminary tests. It also allows the experimenter to ensure that the slopes of the psychometric functions are all the same. If the slopes differ, then the "thresholds" will differ depending on the point at which they are defined. In this case, the slopes were quite similar once the logarithmic transform had been applied.

[2]The performance of listener three for the −1.09 dB signal and the 0-Hz masker did not involve any incorrect responses, so a d' value could not be calculated. For the purpose of the analysis, this point was arbitrarily set to a value of 4.38, which was the highest measurable value obtained by that listener.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Bacon, S. P., and Konrad, D. L. (**1993**). "Modulation detection interference under conditions favoring within- or across-channel processing," J. Acoust. Soc. Am. **93**, 1012–1022.

Bacon, S. P., and Moore, B. C. J. (**1993**). "Modulation detection interference: Some spectral effects," J. Acoust. Soc. Am. **93**, 3442–3453.

Bartlett, N. R. (**1965**). "Thresholds as dependent on some energy relations and characteristics of the subject," in *Vision and Visual Perception*, edited by C. H. Graham (Wiley, New York), pp. 147–175.

Bonnel, A. M., and Hafter, E. R. (**1998**). "Divided attention between simultaneous auditory and visual signals," Percept. Psychophys. **60**, 179–190.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Dallos, P. J., and Olsen, W. O. (**1964**). "Integration of energy at threshold with gradual rise-fall tone pips," J. Acoust. Soc. Am. **36**, 743–751.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., and Verhey, J. L. (**1999**). "Modeling across-frequency processing of amplitude modulation," in *Psychoacoustics, Physiology and Models of Hearing*, edited by T. Dau, V. Hohmann, and B. Kollmeier (World Scientific, Singapore), pp. 229–234.

Dooling, R. J. (**1979**). "Temporal summation of pure tones in birds," J. Acoust. Soc. Am. **65**, 1058–1060.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for

envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.
Garner, W. R., and Miller, G. A. (**1947**). "The masked threshold of pure tones as a function of duration," J. Exp. Psychol. **37**, 293–303.
Gerken, G. M., Bhat, V. K. H., and Hutchison-Clutter, M. (**1990**). "Auditory temporal integration and the power function model," J. Acoust. Soc. Am. **88**, 767–778.
Gockel, H., Carlyon, R. P., and Deeks, J. M. (**2002**). "Effect of modulator asynchrony of sinusoidal and noise modulators on frequency and amplitude modulation detection interference," J. Acoust. Soc. Am. **112**, 2975–2984.
Green, D. M. (**1993**). "Auditory intensity discrimination," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 13–55.
Green, D. M., Birdsall, T. G., and Tanner, W. P. Jr. (**1957**). "Signal detection as a function of signal intensity and duration," J. Acoust. Soc. Am. **29**, 523–531.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Green, D. M., Nachmias, J., Kearney, J. K., and Jeffress, L. A. (**1979**). "Intensity discrimination with gated and continuous sinusoids," J. Acoust. Soc. Am. **66**, 1051–1056.
Hafter, E. R., Bonnel, A.-M., and Gallun, E. (**1998b**). "Detection of change in an auditory stimulus: Psychophysical evidence for a central mechanism," Abstracts of the Twenty-First Midwinter Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL.
Hafter, E. R., Bonnel, A. M., Gallun, E., and Cohen, E. (**1998a**). "A role for memory in divided attention between two independent stimuli," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Q. Summerfield, and R. Meddis, Proceedings of the 11th International Symposium on Hearing (Whurr, London), pp. 228–237.
Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.
Hughes, J. W. (**1946**). "The threshold of audition for short periods of stimulation," Proc. R. Soc. London, Ser. B **133**, 486–490.
Jeffress, L. A. (**1975**). "Masking of tone by tone as a function of duration," J. Acoust. Soc. Am. **58**, 399–403.
Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.
Leshowitz, B., and Raab, D. H. (**1967**). "Effects of stimulus duration on the detection of sinusoids added to continuous pedestals," J. Acoust. Soc. Am. **41**, 489–496.
Leshowitz, B., and Wightman, F. L. (**1971**). "On-frequency masking with continuous sinusoids," J. Acoust. Soc. Am. **49**, 1180–1190.
Macmillan, N. A. (**1971**). "Detection and recognition of increments and decrements in auditory intensity," Percept. Psychophys. **10**, 233–238.
Macmillan, N. A. (**1973**). "Detection and recognition of intensity changes in tone and noise: The detection-recognition disparity," Percept. Psychophys. **13**, 65–75.
Moore, B. C. J., 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**, 3539–3552.
Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (**2004**). "Primary auditory cortex of cats: Feature detection or something else?," Biol. Cybern. **89**, 397–406.
Oxenham, A. J. (**1997**). "Increment and decrement detection in sinusoids as a measure of temporal resolution," J. Acoust. Soc. Am. **102**, 1779–1790.
Oxenham, A. J. (**1998**). "Temporal integration at 6 kHz as a function of masker bandwidth," J. Acoust. Soc. Am. **103**, 1033–1042.
Oxenham, A. J., and Dau, T. (**2001**). "Modulation detection interference: Effects of concurrent and sequential streaming," J. Acoust. Soc. Am. **110**, 402–408.
Tougaard, J. (**1999**). "Receiver operating characteristics and temporal integration in an insect auditory receptor cell," J. Acoust. Soc. Am. **106**, 3711–3718.
Viemeister, N. F. (**1988**). "Psychophysical aspects of auditory intensity coding," in *Auditory Function: Neurobiological Bases of Hearing*, edited by F. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. , 213–241.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Wojtczak, M., and Viemeister, N. F. (**1999**). "Intensity discrimination and detection of amplitude modulation," J. Acoust. Soc. Am. **106**, 1917–1924.
Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**, 848–857.
Yost, W. A., and Sheft, S. (**1990**). "A comparison among three measures of cross-spectral processing of amplitude modulation with tonal signals," J. Acoust. Soc. Am. **87**, 897–900.

Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing[a)]

Shani Whilby, Mary Florentine,[b)] Eva Wagner, and Jeremy Marozeau
Institute for Hearing, Speech and Language and Department of Speech-Language Pathology and Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

(Received 27 December 2004; revised 1 March 2006; accepted 13 March 2006)

The difference in level required to match monaural and binaural loudness of 5- and 200-ms tones was measured for listeners with normal and impaired hearing. Stimuli were 1-kHz tones presented at levels ranging from 10 to 90 dB sensation level. Sixteen listeners (eight normal and eight with losses of primarily cochlear origin) made loudness matches between equal-duration monaural and binaural tones using an adaptive 2AFC procedure. The present results corroborate existing data for 200-ms tones in normal listeners and provide new data for 5-ms tones. On average, the binaural level difference required for equal loudness of monaural and binaural tones is about the same for 5- and 200-ms tones of equal level and changes as a function of level. The group data for normal and impaired listeners are in reasonable agreement with data in the literature. However, the data from some of the impaired listeners deviate markedly from the average, indicating that group data do not accurately represent the behavior of all impaired listeners. Derived loudness functions from the loudness-matching data are reasonably consistent with individual data in the literature. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2193813]

PACS number(s): 43.66.Cb, 43.66.Pn [AO] Pages: 3931–3939

I. INTRODUCTION

Most people with hearing losses listen with both ears. Therefore, understanding the influence of hearing losses on binaural loudness is critical. This is especially important for short-duration sounds, which are pervasive in the environment. The present study examines the relationship between monaural and binaural (diotic) loudnesses of short- and long-duration tones as a function of level for listeners with normal hearing and hearing losses of primarily cochlear origin. Specifically, the difference in the level required to match monaural and binaural loudness [*i.e.*, the Binaural Level Difference for Equal Loudness (BLDEL)] was measured across a wide range of levels.

Listeners with normal hearing show the least BLDEL near threshold, and progressively more BLDEL as the level of the monaural tone increases to 30 or 40 dB above threshold (Fletcher and Munson, 1993; Haggard and Hall, 1982). Maximum amounts of BLDEL are reported to range between 5 dB (Scharf and Fishken, 1970) and 12 dB (Marks, 1978).

Almost all of the existing studies have used long test tones. Only Algom *et al.* (1989) used tones of different durations to study binaural loudness, though no direct matches between monaural and binaural tones were made. To assess loudness, they used free magnitude estimation for tones of six different durations presented monaurally and binaurally. The results show that monaural and binaural loudness functions are nearly parallel for tones ranging in duration from 16 to 256 ms; only the data for 512-ms tones resulted in a shallower monaural than binaural loudness function. Therefore, little is known about BLDEL of short tones and data from equal-loudness matches are not available in the literature.

In addition to the shortage of data on BLDEL of short tones for normal listeners, there are few studies investigating the amount and pattern of BLDEL in impaired listeners. Hawkins *et al.* (1987) studied BLDEL in listeners with high-frequency hearing losses using an adaptive loudness-matching paradigm. The three test levels encompassed only a 20-dB range from most comfortable loudness to loudness-discomfort level, as determined for individual listeners. They reported that BLDEL in their impaired listeners was not significantly different from that for normal listeners. Hall and Harvey (1985) also tested only three levels encompassing a 20-dB range (70, 80, and 90 dB SPL) using an adaptive loudness-matching paradigm. They obtained loudness matches between monaural and binaural tones and found that impaired listeners showed less BLDEL than normal listeners at 70 and 80 dB SPL. The amounts of BLDEL at 90 dB SPL were comparable between the two groups. However, the use of only three levels does not allow an in-depth analysis of the overall pattern of BLDEL.

In summary, there are inconsistent findings across the different studies and a clear need for more data from impaired listeners. The present study provides comprehensive measurements of BLDEL of short and long tones across a wide range of levels in normal and impaired listeners. In addition, a loudness model is applied to the BLDEL data to estimate loudness-growth functions.

[a)]A portion of this work was presented at the twenty-eighth midwinter research meeting of the Association for Research in Otolaryngology, Feb. 19–24, 2005, New Orleans, LA.

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

TABLE I. Audiometric data for the eight impaired listeners.

				Audiometric thresholds in dB HL						
	Sex	Age	Ear	0.25*k*	0.5*k*	1*k*	2*k*	4*k*	8*k*	Possible etiology
HI-1	F	52	L	15	15	40	30	45	60	Unknown
			R	20	15	45	35	50	55	
HI-2	F	57	L	20	40	45	30	20	25	Unknown
			R	15	35	50	35	15	20	
HI-3	M	75	L	25	30	40	75	65	85	Family hx
			R	20	20	45	70	85	NR	
HI-4	M	71	L	15	40	40	45	75	85	Unknown
			R	20	30	40	45	75	NR	
HI-5	M	63	L	25	30	40	45	50	50	Unknown
			R	25	30	45	50	65	50	
HI-6	M	56	L	50	50	55	55	65	105	Noise exposure
			R	50	45	50	55	60	105	
HI-7	M	66	L	20	35	50	55	65	75	Head trauma
			R	15	20	35	35	65	65	
HI-8	M	43	L	15	25	40	40	55	55	Ototoxic rx
			R	20	25	40	40	55	60	

II. METHOD

A. Stimuli

The stimuli were 1 kHz tones with equivalent rectangular durations of 5 and 200 ms. The tones had a 6.67-ms raised-cosine rise and fall. Accordingly, the 5-ms tones consisted only of the rise and fall, whereas the 200-ms tone had a 195-ms steady state portion. The SPL of the tones varied based on the listeners' thresholds at 1 kHz as determined by the procedures outlined below. Fixed levels ranged from 10 to 90 dB sensation level (SL) and did not exceed 106.6 dB SPL for 5-ms tones, or 100 dB SPL for 200-ms tones.

B. Listeners

Sixteen listeners participated in this experiment (seven males and nine females). Eight had normal hearing and eight had hearing losses of primarily cochlear origin. The normal listeners ranged in age from 20 to 73 years (mean age=35). They had no history of hearing difficulties and normal immittance measures. Their pure-tone audiometric thresholds were less than or equal to 10 dB HL at 1 kHz and less than or equal to 15 dB HL at frequencies between 0.25 and 8 kHz (ANSI, 1996). Four of the normal listeners had previous experience with loudness matches. The impaired listeners ranged in age from 43 to 75 years (mean age=60). Their audiological and medical histories were consistent with sensorineural impairment of primarily cochlear origin. All had symmetrical 40- to 60-dB hearing losses at 1000 Hz. Two of the impaired listeners had previous experience with loudness matches. The sex, age, and audiometric data for each impaired listener are given in Table I. Listeners were paid for their participation.

C. Procedure

The procedure consisted of three parts. First, absolute thresholds in the left and right ears were measured for 5- and 200-ms tones to obtain a reference for setting the sensation levels in the subsequent experiments. Next, loudness matches between the two ears were made for each duration so that the loudnesses of the monaural components of the subsequent binaural presentations could be set to be equal. Finally, loudness matches between monaural and binaural tones were made separately for each duration.

1. Absolute thresholds

Absolute thresholds for 5- and 200-ms tones in each ear were measured separately using an adaptive procedure in a two-interval, two-alternative forced-choice (2I, 2AFC) paradigm. Each trial contained two observation intervals, which were separated by 500 ms and marked by lights. The signal was presented with equal *a priori* probability in either the first or second interval. The listener was instructed to indicate which interval contained the tone by pressing a corresponding key on a small computer terminal. A 200-ms light indicated the correct answer 100 ms after the listener's response. The next trial began after a 500-ms delay.

A single threshold measurement was based on three interleaved adaptive tracks. On each trial, one track was selected at random. For each track, the signal level was set approximately 10 dB above the expected threshold of each listener. It decreased after three consecutive correct responses, and increased following one incorrect response. This procedure converged on the signal level yielding 79.4% correct responses (Levitt, 1971). The step size was initially 5 dB and decreased to 2 dB after the second reversal. Reversals occurred when successive signal levels changed direction from decreasing to increasing or *vice versa*. Each track ended after five reversals. The threshold for one track was calculated as the average of the signal levels at the fourth and fifth reversals. One threshold measurement was taken as the average threshold across the three tracks.

2. Equal-loudness matches

The same basic procedures were employed for the loudness matches between ears and between monaural and binaural tones. Loudness matches of equal-duration tones were obtained using a roving-level adaptive procedure with a 2I, 2AFC paradigm. On each trial, the listener heard two 1-kHz tones separated by 600 ms. The listener's task was to indicate which sound was louder by pressing a key on a small computer terminal. The response initiated the next trial after a 1000-ms delay. No feedback was provided.

To reduce biases that may occur when only a single fixed sound is presented in a series of trials, 10 interleaved tracks were used to obtain concurrent loudness matches at five levels with both stimuli varied. Each track consisted of 20 presentations at the given fixed level. On each trial, the track for that trial was selected at random from all tracks not yet completed. Three such matches were obtained for each listener and condition. Because no more than five levels could be tested concurrently, the complete range of fixed levels was divided into a low-level set (i.e., 10, 20, 30, 40, and 50 dB SL) and a high-level set (i.e., 50, 60, 70, 80, and 90 dB SL). The highest level in the low range and the lowest level in the high range were identical. Because of their reduced dynamic range, impaired listeners were run only on the low-level set (i.e., 10, 20, 30, 40, and 50 dB SL). For a more detailed description of the roving-level loudness-matching procedure, see Buus *et al.* (1997).

All variable stimulus levels were determined based on a method of maximum likelihood whereby the computer used a number of psychometric functions after each trial to determine the probabilities of obtaining the listener's responses to all of the previous stimuli. The psychometric function with the highest probability (likelihood) was used to determine the stimulus presentation on the following trial. [For more information, see Florentine *et al.* (2000); for a review, see Formby *et al.* (1996).]

a. Loudness matches of equal-duration tones between ears: Equal loudness matches were obtained between two 1-kHz tones of equal duration (either 5 or 200 ms) presented to opposite ears. For listeners with normal hearing each of the high- and low-level sets was repeated three times for both conditions, resulting in a total of 12 data blocks. Only the low-level sets were presented to impaired listeners and were also repeated three times for each condition, resulting in a total of six blocks. The three equal-loudness matches were averaged for each listener and condition. The final loudness matches between the ears were used to set the right and left ear presentations equally loud for binaural stimuli.

b. Loudness matches between monaural and binaural tones of equal duration: Equal loudness matches were made between monaural and binaural 1-kHz tones of equal duration (either 5 or 200 ms). In five of the 10 interleaved tracks, the level of the monaural tone was fixed; in the other five the level of the binaural tone was fixed. This resulted in four types of blocks for each duration: left monaural *vs.* binaural for low-level and high-level stimulus conditions, and right monaural *vs.* binaural for low-level and high-level stimulus conditions. Each condition was repeated three times on separate days. In total, there were 24 blocks for each listener with normal hearing and 12 for each impaired listener. For each listener and level, four final loudness matches were calculated: the average from the tracks with the monaural tone fixed and the average from the tracks with the binaural tone fixed, each for 5- and 200-ms tones.

D. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, sampled the listeners' responses, and executed the adaptive procedures. The tones were generated digitally with a 50-kHz sample rate and reproduced by a 16-bit digital-to-analog (D/A) converter (TDT DD1). For each ear, the output from the D/A converter was sent through an attenuator (TDT PA4) and low-pass filter (TDT FT5, f_c=20 kHz, 190 dB/octave) before being sent to the right and left earphone of a Sony MDR-V6 headset *via* a headphone amplifier (TDT HB6). The listeners were seated in a sound-attenuating booth.

For routine calibration, the output of the headphone amplifier was led to a 16-bit analog-to-digital (A/D) converter (TDT DD1), so that the computer could sample the wave form, calculate its spectrum and rms voltage, and display the results before each block of matches. [The reported SPLs presume an output of 116 dB SPL for an input of 1 V rms, which is approximately that measured at 1 kHz in a 6-cc coupler (B&K 4152).]

E. Data analysis

To examine the statistical significance of the effects of stimulus variables and differences among listeners, an analysis of variance for repeated measures (ANOVA, sensation level × duration × fixed mode × monaural ear × listener) was performed. The dependent variable was the level difference ($L_{\text{monaural}} - L_{\text{binaural}}$) between two equally loud monaural and binaural stimuli of the same duration. For all tests, the outcome was considered significant when $p \leqslant 0.05$.

III. RESULTS

Individual and group results for the loudness matches between monaural and binaural tones of equal duration and threshold measurements are shown in Fig. 1 for normal listeners and in Fig. 2 for impaired listeners. Results from the loudness matches of equal-duration tones between ears are not shown, since they only provided the basis for the presentation of the binaural stimuli and the levels at the two ears matched within a couple of dB. Mean thresholds for listeners with normal hearing range from 22.8 to 32.8 dB SPL for 5-ms tones (average=26.3 dB), and from 9 to 18.4 dB SPL for 200-ms tones (average=12.5 dB). Mean thresholds for impaired listeners range from 57.3 dB SPL to 69.2 dB SPL for 5-ms tones (average=62.1 dB), and from 48 dB SPL to 59.8 dB SPL for 200-ms tones (average=53.5 dB).

For each listener, the BLDEL is plotted as a function of the SPL of the monaural tone. Circles and squares represent data for 5- and 200-ms tones, respectively. Data obtained when the monaural tone was fixed in level are shown with open symbols; closed symbols indicate data obtained when the level of the monaural tone was varied. For listeners with normal hearing, the BLDEL functions for both tone durations

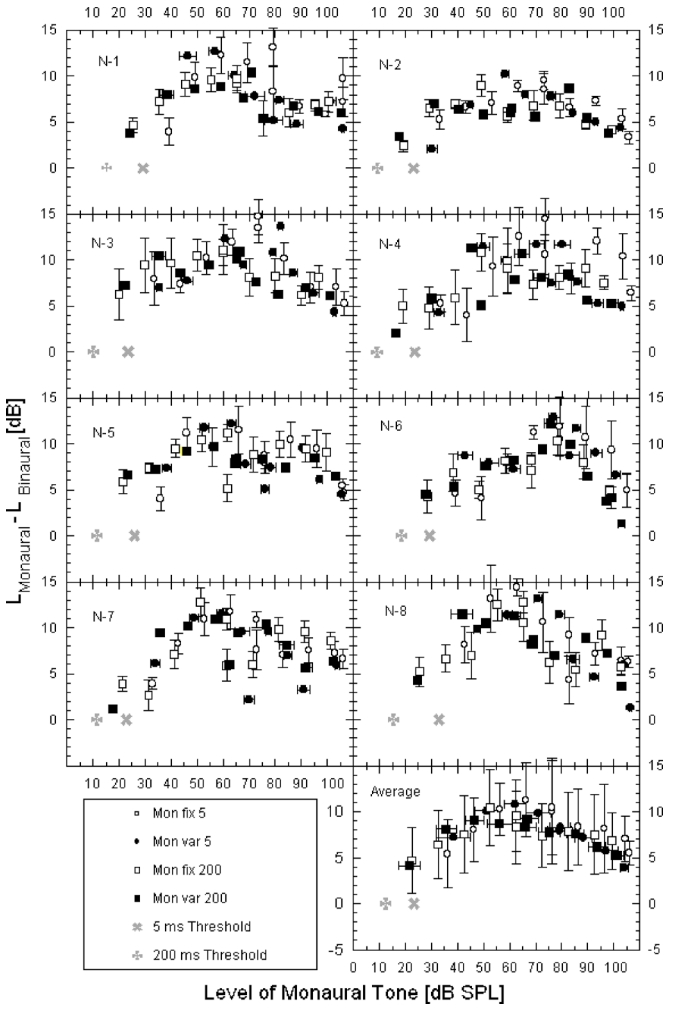


FIG. 1. The level difference between monaural and binaural tones needed to obtain equal loudness (i.e., the binaural level difference for equal loudness, BLDEL) is plotted as a function of the level of the monaural tone. Individual and group data are shown for 1-kHz tones in normal listeners. The level difference between monaural and binaural tones needed to obtain equal loudness is plotted as a function of the level of the monaural tone in dB SPL. Circles and squares represent data for 5- and 200-ms tones, respectively. Data obtained when the monaural tone was fixed in level are shown with open symbols; closed symbols indicate data obtained when the level of the monaural tone was varied. Error bars represent standard errors for individual data and standard deviations for group data. Gray symbols show monaural thresholds for 5-ms tones and for 200-ms tones, as shown in the figure legend.

show the same pattern: the amount of BLDEL changes as a function of level (ranging from about 2 to 15 dB) and is largest at moderate levels. The maximum BLDEL varies among normal listeners and peaks at levels between 55 and 75 dB SPL depending on the individual. The results of the ANOVA show no significant effect of tone duration on BLDEL ($p=0.3204$), but a significant effect of level on the amount of BLDEL ($p \leqslant 0.0001$). Individual listeners make highly consistent loudness matches as indicated by the small standard errors (average is 1.6 dB; range is from 0.3 to 3.7 dB). Differences across listeners are also relatively small; the mean standard deviation across levels is 3.9 dB.

Figure 2 shows the amount of BLDEL for impaired listeners. Again, summation is independent of tone duration ($p=0.5051$) but does depend on level ($p=0.0047$), with the

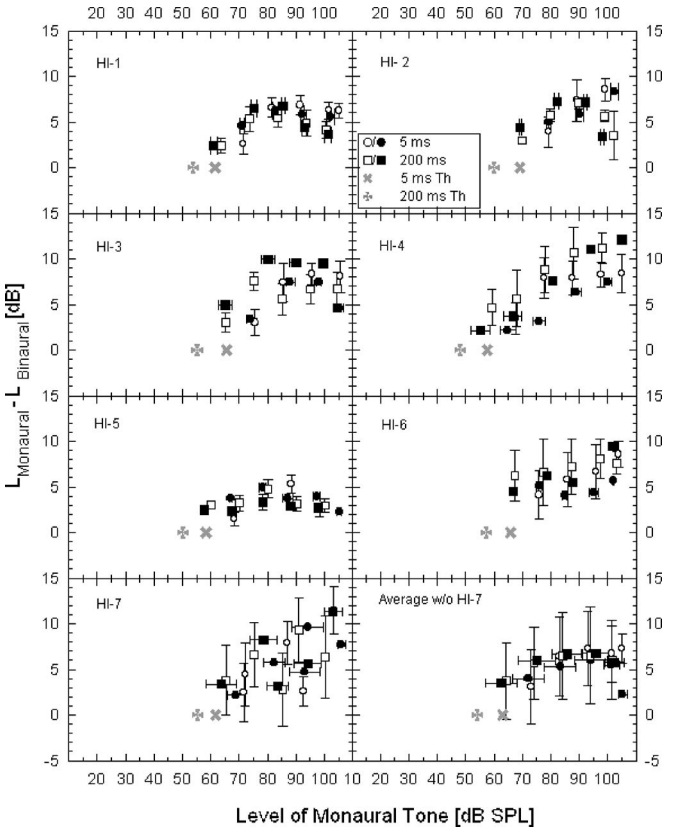


FIG. 2. Individual and group BLDEL for 1-kHz tones in impaired listeners plotted in the same manner as Fig. 1. The data from HI-7 are not included in the average graph.

amount ranging from 1.5 to 12 dB. Although variability across listeners is larger than that found in listeners with normal hearing (average standard deviation=4.5 dB), all listeners except HI-7 were consistent in their loudness judgments (average standard error is 1.4 dB; range is from 0.2 to 3.7). One of the eight listeners (HI-8) had difficulties performing the task and was not able to provide reliable results. His data showed no clear pattern and are not included in the figure. Four listeners (HI-1, HI-2, HI-3, HI-5) show a pattern similar to that found in normal listeners with greatest amounts of BLDEL at moderate SLs. Two of the remaining listeners (HI-4, HI-6) show increasing BLDEL with increasing level. HI-7 started taking medication for a heart condition, after the medical history of HI-7 was taken and before the onset of data collection. This prescription medication has known side effects causing confusion and mental changes, which were noted by the experimenters. His data were therefore excluded from average data shown in Fig. 2 and from all further group data analysis. (*NB*: Because the data from the impaired listeners appear to represent at least two populations, it seems unwise to average the data over all the impaired listeners. The present data were averaged with the sole purpose of comparison with other data in the literature.)

IV. DISCUSSION

A. Monaural and binaural loudness in normal listeners

The present data from normal listeners are compared with data in the literature in Fig. 3.[1] Stimuli and procedures

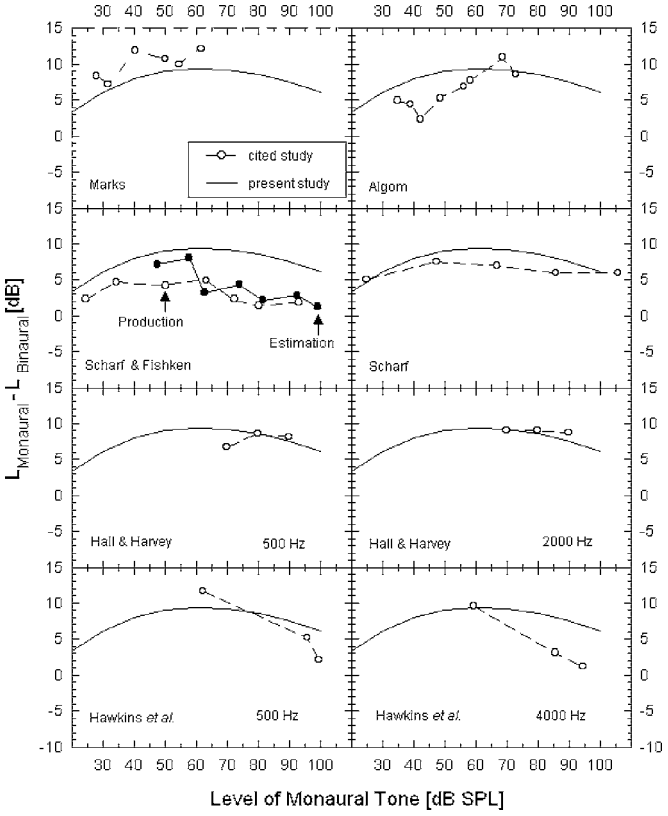


FIG. 3. Comparison among studies of BLDEL for tones in normal listeners. The solid line represents a fit to the average data for both durations from the present study. Additional data are taken from Marks (1978) in the top left panel, Algom *et al.* (1989) in the top right panel, Scharf and Fishken (1970) in the left upper middle panel, Scharf (1968) in the right upper middle panel, Hall and Harvey (1985) in the left and right lower middle panels, and Hawkins *et al.* (1987) in the left and right bottom panels. See the text for a more detailed description.

varied significantly. Details of these studies are summarized in Table II. Data from Marks (1978) and Algom *et al.* (1989) are in reasonable agreement with the present study, showing maximum amounts of BLDEL at moderate levels. The lack of data at higher SLs limits the possibility for comparison of the overall pattern of BLDEL. The Scharf and Fishken (1970) data obtained using magnitude estimation show less BLDEL than the present data, as well as a reduction in the overall amount of BLDEL with increasing levels, but do not include lower SLs for comparison. Their results using magnitude production cover a wide range of levels, and show a flatter function with less BLDEL than obtained in the present study. However, the function derived by Scharf (1968) using loudness matches shows a pattern of BLDEL similar to that of the present study. Data from Hall and Harvey (1985) are in reasonable agreement with our data. Unfortunately, they covered only a small level range. In agreement with the present study, Hawkins *et al.* (1987) found a maximum amount of BLDEL at the moderate level tested (60 dB SPL) and smaller amounts of BLDEL at the two high levels tested (95 and 100 dB SPL). However, the amounts of BLDEL differ notably from those of the present study, especially at the high levels. Differences among the reported results may be due to the various psychophysical procedures, durations, frequencies, and more restricted ranges of levels tested. There is a wide range of maximum amounts of BLDEL (4.9–12.1 dB) and data from the present study fall within that range (9.1 dB).

The present data corroborate the existing data for 200-ms tones and extend the literature by presenting BLDEL data for 5-ms tones. Only Algom *et al.* (1989) used tones of different durations to study binaural loudness and they used free magnitude estimation in which direct loudness matches between monaural and binaural tones were never made. The present data show that the magnitude and pattern of BLDEL is about the same for 5- and 200-ms tones in normal listeners.

B. Monaural and binaural loudness in impaired listeners

In Fig. 4 data from two studies with impaired listeners were replotted in the same manner as in Fig. 3. Despite large individual differences in the present data for impaired listeners, averages were calculated for comparison with those of the following two studies. Hall and Harvey (1985) and Hawkins *et al.* (1987) only presented average data. The results of Hall and Harvey (1985) show reasonable agreement with those of the present study despite the different test frequencies. The amount of BLDEL increases with increasing level up to 90 dB SPL for both frequencies tested. The overall amount of BLDEL is slightly greater than that of the present study for the 500-Hz test tone, but is comparable for the 2000-Hz test tone. Data collected by Hawkins *et al.*

TABLE II. Summary of stimulus and procedural parameters of seven studies.

Study	Procedure	Duration (ms)	Frequency (kHz)
Present study	Adaptive, 2I 2AFC	5, 200	1
Marks (1978)	Magnitude estimation	1000	1
Algom *et al.* (1989)	Magnitude estimation	16, 32, 64, 128, 256, 512 (only 256-ms data plotted)	2
Scharf and Fishken (1970)	Magnitude production Magnitude estimation	1000	1
Scharf (1968)	Loudness matches	750	1
Hall and Harvey (1985)	Adaptive, 2I 2AFC	300	0.5 2
Hawkins *et al.* (1987)	Adaptive, forced choice	200	0.5 4

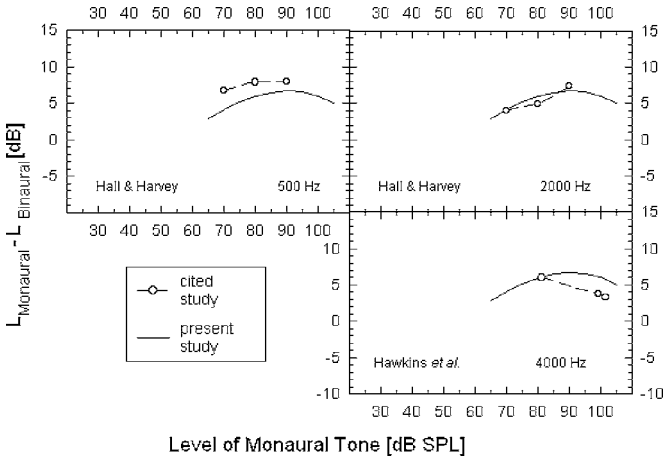


FIG. 4. Comparison among studies of BLDEL for tones in impaired listeners. The level difference between monaural and binaural tones needed to obtain equal loudness is plotted as a function of the level of the monaural tone. The solid line represents a fit to the average data from the present study. Additional data are taken from Hall and Harvey (1985) in the top right and left panels, and from Hawkins *et al.* (1987) in the bottom right panel.

(1987) for a 500-Hz test tone were excluded from this comparison because listeners had high-frequency hearing losses with normal thresholds at 500 Hz (average=9 dB HL). At 4000 Hz their data show slightly less BLDEL than that found in the present study, although there is a decline in the amount of BLDEL at higher levels. Considering the ranges of levels tested in each study, average data from the present study are in good agreement with the average data in the literature. If we only look at the average data we could conclude that impaired listeners show about the same amount of BLDEL, as did Hall and Harvey (1985) and Hawkins *et al.* (1987).

Although the average data indicate that normal and impaired listeners show about the same amount of BLDEL, comparisons of group data from normal and impaired listeners should be made with caution because of the notable individual differences. In other words, the mean data suggest that impaired listeners show similar shaped functions as normal listeners, but less overall BLDEL. The individual data indicate that differences among individuals are much larger for impaired than for normal listeners.

In contrast to the normal listeners, who show a consistent general pattern, the impaired listeners appear to fall into two groups. One group exhibits functions that are similar in shape, but not in magnitude, to those of normal listeners. The BLDEL data from the other group increase and can rise to about 12 dB at 105 dB SPL. In addition, not all impaired listeners may show the same magnitude and shape of the function for the 5- and 200-ms tones. For example, HI-2 shows less BLDEL for the 200-ms tones at about 100 dB SPL than for the 5-ms tones. The opposite is true for HI-4 who shows less BLDEL for the 5-ms tones at about 90 to 105 dB SPL.

These findings may have implications for the rehabilitation of persons with hearing losses, especially those fitted with binaural hearing aids. Although there is no standard procedure for fitting people with binaural aids, one common method prescribes a 3-dB reduction in gain when going from monaural to binaural hearing aids. This method appears to work quite well for some impaired listeners, but it may not work well for all. The results of the present study suggest that the use of average data may be misleading due to individual differences among impaired listeners. In addition, the algorithms used for hearing aids assume the same BLDEL for short and long tones. This appears to be an incorrect assumption at high levels for two of the present impaired listeners. Therefore, the consideration of individual differences may be important for understanding binaural loudness in impaired listeners.

C. Modeling

Loudness functions were derived from the BLDEL data from each listener by assuming that the Binaural Equal-Loudness-Ratio Hypothesis (BELRH) was true for the present data. The BELRH, an extension of the equal-loudness-ratio hypothesis (Epstein and Florentine, 2005; Florentine *et al.*, 1996), states that the ratio between the loudnesses of binaural and monaural tones at the same level is independent of level. In other words, when loudness is plotted on a logarithmic scale as a function of SPL, the binaural loudness function is vertically shifted by a constant amount from the monaural loudness function.

The monaural and binaural loudness functions were modeled as third-order polynomials on a logarithmic scale. They were fit using a least-squares method so that the horizontal distances between the functions on a logarithmic scale matched the BLDEL data.

The monaural loudness function, F_m, was defined as

$$F_m(L) = 10^{a_m L^3 + b_m L^2 + c_m L + d_m}, \tag{1}$$

where L is the level in dB SPL of the monaural stimulus, a_m, b_m, c_m, and d_m are the free parameters of the model. The binaural loudness function, F_b, can be defined by the same type of function,

$$F_b(L) = 10^{a_b L^3 + b_b L^2 + c_b L + d_b}, \tag{2}$$

where L is the level in dB SPL of the binaural stimulus and a_b, b_b, c_b, and d_b are the free parameters specific to the binaural loudness function. The BELRH implies that, for every level L, the ratio between these two functions is constant: K_{ELRH},

$$F_b(L)/F_m(L) = K_{\mathrm{ELRH}}. \tag{3}$$

On a logarithmic scale, the BELRH implies that the two functions are parallel and are separated by a constant quantity equal to the logarithm of K_{ELRH}, therefore

$$a_m = a_b,$$

$$b_m = b_b,$$

$$c_m = c_b,$$

$$d_m = d_b - \log(K_{\mathrm{ELRH}}), \tag{4}$$

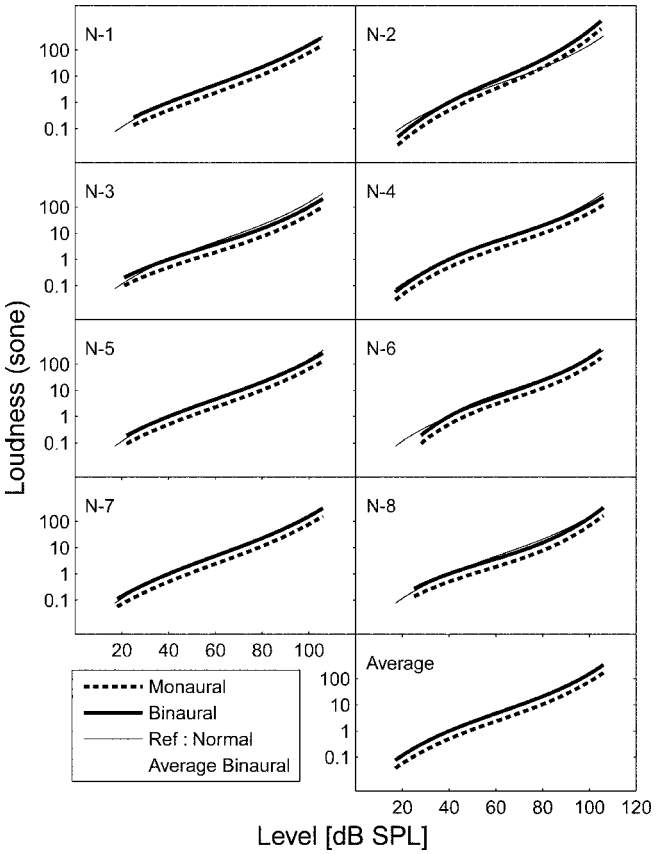


FIG. 5. Individual derived loudness functions for eight normal listeners. The solid lines correspond to the binaural loudness functions, the dashed lines to the monaural loudness functions. The bottom right panel shows the average data. The thin line shows average binaural loudness function.

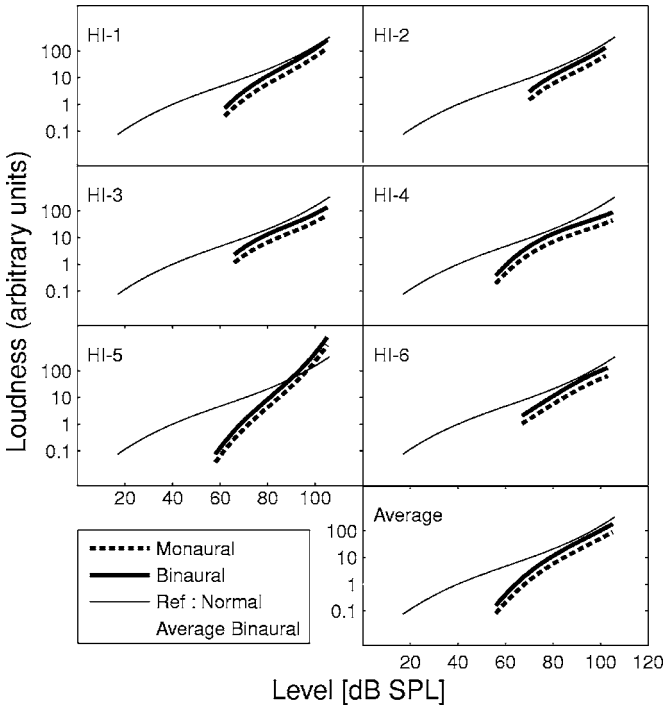


FIG. 6. Individual derived loudness functions for six impaired listeners plotted in the same manner as Fig. 5. Each loudness function for the impaired listeners was set in a manner consistent with average data in the literature (see text). The thin line shows average normal data from Fig. 5.

The constant, K_{ELRH}, describes the amount of binaural loudness summation (BLS). It is well known that a sound presented to both ears is louder than the same sound presented to only one ear (*e.g.*, Fletcher and Munson, 1993; Scharf and Fishken, 1970; Marks, 1978; Hall and Harvey, 1985; Algom *et al.*, 1989). Generally, it is assumed that the total loudness of a binaural sound is equal to the sum of the loudnesses in individual normal ears [for review, see Marks (1978) and Hellman (1991)]. Although Scharf and Fishken (1970) maintain that BLS depends on level and is not always complete, the majority of the available reviews indicate that setting the constant, K_{ELRH}, to 2 is reasonable, at least at levels exceeding 20 dB SPL (Marks, 1978). The same constant was also used by Fletcher (1953) to extract loudness functions from comparisons between binaural and monaural tones. Therefore, perfect summation was assumed to plot Figs. 5 and 6.

For normal listeners, the value of d_b was set such that the function F_b gives a value of one sone at a level of 40 dB SPL (ISO R532). For impaired listeners, the value of d_b was set such that each individual average monaural loudness at 85 dB SPL match the overall average monaural loudness of the normal listeners at 80 dB SPL.

Now, the model is comprised of only three free parameters: a, b, and c. For a fixed, selected level of the monaural stimulus, L_m, a third-order polynomial fit of the BLDEL data was used to estimate the level of the binaural stimulus, L_b, at which the loudnesses of the monaural and binaural stimuli were equal. (An analysis of the fit error as function of order shows that using a higher order fit does not substantially decrease the error.) This can be expressed by the following equations:

$$
\begin{aligned}
&F_b(L_b) = F_m(L_m),\\
&\log(F_b(L_b)) = \log(F_m(L_m)),\\
&a(L_m^3 - L_b^3) + b(L_m^2 - L_b^2) + c(L_m - L_b) = d_b - d_m\\
&\quad = \log(K_{\mathrm{ELRH}}).
\end{aligned}
\tag{5}
$$

Finally, the values of the last three free parameters were determined using a least-squares fit to minimize the error of the fit between the model and the BLDEL data.

D. Application of the model to derive loudness functions for normal and impaired listeners

The amount of BLDEL should be inversely proportional to the slope of the loudness function when plotted as loudness versus level. Accordingly, loudness functions in Figs. 5 and 6 were derived from the BLDEL data for each listener as described earlier. The derived normal loudness functions for individual listeners, shown in Fig. 5, are consistent with the shape of directly measured individual loudness functions. The slopes of the loudness functions for normal listeners change as a function of level and are lowest at moderate levels, consistent with temporal-integration data (Florentine *et al.*, 1996; Buus and Florentine, 2001) and other individual data in the literature (e.g., Epstein and Florentine, 2005). Slopes fit to the individual data above 40 dB in the classical

manner (i.e., fitting a power function to all the data) are slightly steeper than the typical 0.3 reported in the literature (Hellman, 1999). If the assumptions made to derive the loudness functions were correct, the existence of less than perfect binaural loudness summation would be indicated. For example, if perfect binaural summation is not assumed and a ratio $K_{ELRH}=1.8$ is used, the slopes of the derived loudness functions approximate the typical 0.3. The existence of perfect binaural summation is still an open issue that awaits further study.

In any case, the individual loudness functions look quite reasonable and indicate that deriving loudness functions from BLDEL data is a new, albeit indirect, method to measure loudness functions. Before this new method is employed, more direct testing is needed between the derived and directly measured loudness functions. Unfortunately, these data could not be obtained due to time limitations.

The derived loudness functions for individual impaired listeners are shown in Fig. 6. An additional assumption was made to plot these data, i.e., that the mean monaural normal loudness at 80 dB SPL is equal to the mean monaural loudness of the impaired listeners at 85 dB SPL. This assumption was made on the bases of average data from impaired listeners with absolute thresholds of 45 dB HL (from Hellman, 1999, Fig. 14) and is consistent with average data from Miskolczy-Fodor (Hellman, 1964, Fig. 18). It is likely to be inaccurate in some cases because not all impaired listeners have similar loudness growth functions. The derived loudness functions in Fig. 6 show the very wide range of loudness functions that can be observed in impaired listeners. Impaired listeners can have different rates of loudness growth (Hellman and Meiselman, 1993; Hellman, 1999; Buus and Florentine, 2002; Florentine, 2004). The data from listener HI-5 are consistent with the shapes of many directly measured individual loudness functions in the literature, showing rapid loudness growth that may exceed normal (Hellman, 1999, Fig. 13, listener LP). The low-SL slopes of the matching functions—once clearly above threshold (Florentine and Buus, 2002)—generally increase as hearing loss increases, similar to the general tendency observed by Miskolczy-Fodor (1960) and Steinberg and Gardner (1937). The more linear response observed in impaired listeners is thought to result from a loss of compression (for review see Moore, 1995) and it is likely that the different degrees of linearization lead to the different shapes of the loudness functions. The derived loudness functions for the other impaired listeners may be consistent with the concept of loudness being greater than normal at and near threshold in some impaired listeners than normal listeners (Buus and Florentine, 2002; Florentine and Buus, 2002; for an alternative view, see Moore, 2004) and appear consistent with indirect measures of loudness using a time-reaction procedure (Florentine *et al.*, 2004). Given the wide range of loudness data from impaired listeners, the derived loudness functions for the impaired listeners look fairly reasonable and suggest that individual loudness functions may be derived from BLDEL data, although more assumptions need to be made for the impaired listeners than normal listeners.

V. SUMMARY

The present group data for normal and impaired listeners are in reasonable agreement with data in the literature. For both normal and impaired listeners, the data indicate that the loudness of a tone—whether 5 ms or 200 ms—is greater if it is heard in two ears than if it is heard in only one ear. The BLDEL changes with level and is greatest at moderate levels. For normal listeners, BLDEL is about 4 dB at 20 dB SPL, increases to about 10 dB at 60 dB SPL, and decreases to about 5–6 dB at 100 dB SPL. For impaired listeners with moderate losses, the group data show about 3 dB of BLDEL at 65 dB SPL, which increases to about 6 dB at 90 dB SPL and decreases to about 5 dB at 100 dB SPL. However, there are clear individual differences in the data from the impaired listeners, indicating that group data do not represent the behavior of all impaired listeners. Loudness functions derived from the BLDEL data from normal listeners are consistent with the shapes of directly measured individual loudness functions. Slopes fit to individual data are slightly steeper than the typical 0.3. However, the individual loudness functions look quite reasonable and indicate that deriving loudness functions from BLDEL data may be a new, indirect method to measure individual loudness functions. Derived loudness functions for individual impaired listeners show the very wide range of loudness functions that can be observed in impaired listeners.

ACKNOWLEDGMENTS

Søren Buus contributed substantially to this work; he passed away prior to its completion. The authors thank Michael Epstein, Rhona Hellman, Bertram Scharf, and Sandra Cleveland for helpful comments. Brian Moore and Andrew Oxenham provided comments that substantially improved the paper. NIH/NIDCD Grants Nos. R01DC02241 and RO1DC02241-07S1 supported this research.

[1]Data from each study were collected using different stimuli and procedures, and have been replotted to show the difference in level between equally loud monaural and binaural tones as a function of the SPL of the monaural tone. Data from Fig. 5 of Marks (1978) showed the estimated loudness of monaural and binaural tones based on measures of magnitude estimation over a level range of 10 to 50 dB SPL. Data from Fig. 1 and Fig. 2 of Algom *et al.* (1989) showed loudness estimates made through magnitude estimation for monaural and binaural tones of different durations over a level range from 30 to 70 dB SPL. Only data for the 256-ms test tone, which is the duration that is closest to that in our study, were used for comparison. Data from Fig. 1 of Scharf and Fishken (1970) showed the numbers assigned through magnitude estimation and magnitude production to monaural and binaural tones over a level range from 20 to 110 dB SPL. Data from Fig. 1 of Scharf (1968) showed the difference in level between equally loud monaural and binaural tones as a function of the SPL of the binaural tone. Data from Fig. 1 of Hall and Harvey (1985) showed the diotic SPL of 500 and 2000 Hz tones as a function of the level in SPL of monaural tones over a range of 60 to 95 dB SPL. Data from each study were extracted and used to determine the levels of monaural and binaural tones necessary to evoke equal loudness. The level of each binaural tone was then subtracted from the level of an equally loud monaural tone in order to determine the difference in level of equally loud monaural and binaural stimuli.

Algom, D., Rubin, A., and Cohen-Raz, L. (**1989**). "Binaural and temporal integration of the loudness of tones and noises," Plasma Chem. Plasma Process. **46**, 155–166.

ANSI (**1996**). "Specification for audiometers," *ANSI S3.6-1996* (American National Standards Institute, New York).
Buus, S., and Florentine, M. (**2001**). "Modifications to the power function for loudness," in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, T. Lachmann (Pabst, Berlin), pp. 236–241.
Buus, S., and Florentine, M. (**2002**). "Growth of loudness in cochlearly impaired listeners: Recruitment reconsidered," J. Assoc. Res. Otolaryngol. **3**, 120–139.
Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**, 669–680.
Epstein, M., and Florentine, M. (**2005**). "A test of equal-loudness-ratio hypothesis using cross-modality matching functions," J. Acoust. Soc. Am. **118**, 907–913.
Fletcher, H. (**1953**). *Speech and Hearing in Communication*, 2nd ed. (Van Nostrand, New York).
Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.
Florentine, M. (**2004**). "Softness imperception: Defining a puzzling problem," Hearing Health. **20**, 31–34.
Florentine, M., and Buus, S. (**2002**). "Evidence for normal loudness growth near threshold in cochlear hearing loss," in *Genetics and the Function of the Auditory System, GN ReSound*, edited by L. Tranebjaerg *et al.* (Tåstrup, Denmark), pp. 411–426.
Florentine, M., Buus, S., and Geng, W. (**2000**). "Toward a clinical procedure for narrowband gap detection I: a psychophysical procedure," Audiology **39**, 161–167.
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 Rosenberg, M. (**2004**). "Reaction-time data support the existence of softness imperception in cochlear hearing loss," in *Auditory Signal Processing: Physiology, Psychoacoustics and Models*, edited by S. McAdams *et al.* (Springer, New York).
Formby, C., Sherlock, L. P., and Green, D. M. (**1996**). "Evaluation of a maximum likelihood procedure for measuring pure-tone thresholds under computer control," J. Am. Acad. Audiol **7**, 125–129.
Haggard, M. P., and Hall, J. W. (**1982**). "Forms of binaural summation and the implications of individual variability for binaural hearing aids," Scand. Audiol. Suppl. **15**, 47–63.
Hall, J. W., and Harvey, A. D. (**1985**). "Diotic loudness summation in normal and impaired hearing," J. Speech Hear. Res. **28**, 445–448.
Hawkins, D. B., Prosek, R. A., Walden, B. E., and Montgomery, A. A. (**1987**). "Binaural loudness summation in the hearing impaired," J. Speech Hear. Res. **30**, 37–43.
Hellman, R. (**1964**). "Loudness functions of a 1000-cps tone in the presence of a masking noise," J. Acoust. Soc. Am. **36**, 1618–1627.
Hellman, R. (**1991**). "Loudness measurements by magnitude scaling: Implications for intensity coding," in *Ratio Scaling of Psychological Magnitude*, edited by S. J. Bolanowski and G. A. Gescheider, pp. 215–228.
Hellman, R. P. (**1999**). "Cross-modality matching: A tool for measuring loudness in sensorineural impairment," Ear Hear. **20**, 193–213.
Hellman, R. P., and Meiselman, C. H. (**1993**). "Rate of loudness growth for pure tones in normal and impaired hearing," J. Acoust. Soc. Am. **93**, 966–975.
ISO R 532 (1966). International Organization for Standardization, Acoustics method for calculating loudness level.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am. **64**, 107–113.
Miskolczy-Fodor, F. (**1960**). "Relation between loudness and duration of tonal pulses. III. Response in cases of abnormal loudness function," J. Acoust. Soc. Am. **32**, 486–492.
Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford Medical Publications, Oxford University Press).
Moore, B. C. (**2004**). "Testing the concept of softness imperception: loudness near threshold for hearing-impaired ears," J. Acoust. Soc. Am. **115**, 3103–3111.
Scharf, B. (**1968**). "Binaural loudness summation as a function of bandwidth," *The 6th International Congress on Acoustics*, Tokyo, Japan.
Scharf, B., and Fishken, D. (**1970**). "Binaural summation of loudness: reconsidered," J. Exp. Psychol. **86**, 374–379.
Steinberg, J. C., and Gardner, M. B. (**1937**). "The dependence of hearing impairment on sound intensity," J. Acoust. Soc. Am. **9**, 11–23.

Informational masking of speech in children: Auditory-visual integration

Frederic Wightman[a] and Doris Kistler
Heuser Hearing Institute and Department of Psychological and Brain Sciences, University of Louisville, Louisville, Kentucky 40292

Douglas Brungart
Human Effectiveness Directorate, Air Force Research Laboratory, 2610 Seventh Street, Wright Patterson AFB, OH 45433

(Received 19 July 2005; revised 5 March 2006; accepted 15 March 2006)

The focus of this study was the release from informational masking that could be obtained in a speech task by viewing a video of the target talker. A closed-set speech recognition paradigm was used to measure informational masking in 23 children (ages 6–16 years) and 10 adults. An audio-only condition required attention to a monaural target speech message that was presented to the same ear with a time-synchronized distracter message. In an audiovisual condition, a synchronized video of the target talker was also presented to assess the release from informational masking that could be achieved by speechreading. Children required higher target/distracter ratios than adults to reach comparable performance levels in the audio-only condition, reflecting a greater extent of informational masking in these listeners. There was a monotonic age effect, such that even the children in the oldest age group (12–16.9 years) demonstrated performance somewhat poorer than adults. Older children and adults improved significantly in the audiovisual condition, producing a release from informational masking of 15 dB or more in some adult listeners. Audiovisual presentation produced no informational masking release for the youngest children. Across all ages, the benefit of a synchronized video was strongly associated with speechreading ability. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195121]

PACS number(s): 43.66.Dc, 43.71.Ft [JHG] Pages: 3940–3949

I. INTRODUCTION

The ability to segregate and selectively attend to the components of an auditory scene is central to our appreciation of and interaction with the world around us. Often the components of an auditory scene interfere with each other and degrade our auditory source segregation ability. One type of interference is energetic masking, thought to be a consequence of temporal and spectral overlap of the target and distracting sounds. The extent of energetic masking is accurately estimated by filter-bank models of the auditory periphery (Moore and Glasberg, 1987; Slaney, 1993, 1994). Informational masking is another type of interference that inhibits source segregation and is thought to occur when distracting sounds are highly variable or are perceptually similar to target sounds (Durlach *et al.*, 2003; Lutfi, 1990). The distinction between energetic and informational masking is readily appreciated in the case of a pure-tone detection task. When a pure tone is presented in a background of wideband Gaussian noise, the detection threshold for the tone is elevated almost exclusively by energetic masking, since the noise is relatively static and has a very different quality than the tonal signal. However, when the pure tone signal is presented in a background of other tones that have random frequencies and levels, informational masking occurs because the distracting sound is variable and has a tonal quality. Informational masking of this sort produces a substantial decrement in detection performance, represented by as much as a 40-dB threshold elevation in some conditions and in some listeners (Neff and Callaghan, 1988; Oh and Lutfi, 1998).

Recent results from our laboratory suggest that informational masking in pure-tone detection tasks is much greater in young children than in adults (Lutfi *et al.*, 2003; Oh *et al.*, 2001; Wightman *et al.*, 2003). Hall *et al.* (2005) report similar findings with complex tonal stimuli. However, the degree to which these results generalize to speech stimuli has only recently been demonstrated.

With a speech signal and speech distracter that overlap each other temporally and spectrally, it is sometimes difficult to determine how much of the observed masking effects are informational and how much are energetic. The clearest identification of the informational masking effects of speech distracters is found in the work of Brungart and colleagues (Brungart, 2001b; Brungart and Simpson, 2002a, b, 2004; Brungart *et al.*, 2005, 2001). This work shows that in certain tasks, in particular the Coordinate Response Measure (CRM) task developed at the Air Force Research Laboratory (Bolia *et al.*, 2000), nearly all of the interference in speech target recognition produced by speech distracters is informational and this interference can amount to an overall threshold shift of as much as 10 dB.

Children tested in the CRM paradigm appear to demonstrate much larger amounts of informational masking than adults (Wightman and Kistler, 2005). For the youngest chil-

[a]Electronic mail: fred.wightman@louisville.edu

 0001-4966/2006/119(6)/3940/10/$22.50

dren (ages 4–5 years) masking is more than 15 dB greater than in adults. A clear and monotonic age effect is also observed, with children as old as 16 years still not performing at adult levels. These results are generally consistent with other studies of children's speech recognition with speech distracters (Fallon *et al.*, 2000; Hall *et al.*, 2002). Interpretation of the age effect shown by Wightman and Kistler (2005) in terms of informational masking is complicated by the many reports of higher energetic masking thresholds in children (e.g., Allen and Wightman, 1994; Allen *et al.*, 1989; Schneider *et al.*, 1989). However, it appears to be the case that only the youngest children (i.e., preschoolers) produce markedly higher masked thresholds (Allen *et al.*, 1989; Schneider *et al.*, 1989). For example, recent data from our laboratory show that children older than about 6.5 years demonstrate adultlike detection thresholds for a tone masked by wideband noise (Oh *et al.*, 2001; Wightman *et al.*, 2003). Moreover, with a male target talker in the CRM paradigm, the switch from a male to a female distracter produces an 8 -dB release from masking in all age groups tested, including the adults (Wightman and Kistler, 2005). Differences in energetic masking between male and female distracters cannot fully explain the 8-dB masking release (Brungart, 2001b). Finally, an analysis of the errors made by the children suggested that the children confused the distracter and the target; errors came primarily from the distracter message. If energetic masking were involved, one might assume that target audibility would be decreased (see also Brungart, 2001b) and that this would lead to random distributions of errors. Thus, we argue that in the CRM speech task, the interference produced by the distracter is dominated by informational masking in all age groups tested.

There are many aspects of everyday listening that may mitigate informational masking effects. For example, spatial separation of the target signal from the distracters might be expected to produce a significant release from informational masking, and a number of experiments with adults confirm this expectation (Freyman *et al.*, 2001; Freyman *et al.*, 1999; Helfer and Freyman, 2005; Kidd *et al.*, 1998, 2005b). Unfortunately, initial research with children suggests a much smaller benefit of spatial separation (Hall *et al.*, 2005; Litovsky, 2005; Wightman *et al.*, 2003).

In many real world listening situations, the listener has the advantage of being able to see the target talker. It is well known that for adults integration of the information from the auditory signal and the visual cues obtained from speechreading (lipreading) can provide a substantial improvement in speech understanding in a background of noise (Sumby and Pollack, 1954; Summerfield, 1979) or multitalker babble (Sommers *et al.*, 2005). Recent experiments on speech recognition in speech backgrounds suggest that integration of auditory and visual sources of information (A/V integration) produces as much as a 10-dB release from informational masking (Helfer and Freyman, 2005). For most listeners this would mean the difference between understanding nothing and understanding everything. Because speech recognition is such an important aspect of development in children, and since children are often forced to listen to speech in noisy classroom environments where informational masking may be especially troublesome, we view A/V integration as an important topic to address in research on children.

It is important to note here that we use the term "A/V integration" to refer to the process whereby a listener achieves some improvement in speech understanding by viewing a simultaneous visual representation of the talker. However, as Grant and colleagues argue (Grant, 2002; Grant and Seitz, 1998, 2000; Grant *et al.*, 1998; van Wassenhove *et al.*, 2005), the benefit of combining auditory and visual cues is the result of at least two independent processes, information encoding and information integration. Thus, individual differences in A/V benefit, including age effects, might be the result of differences in either encoding or integration or both. The experiments reported here do not permit us to disentangle the separate contributions of the two processes. Nevertheless, consistent with many other reports in the literature, we will continue to use the term "A/V integration" to refer to the end result of both processes, and, where appropriate, we will discuss the encoding and integration processes separately.

There have been several previous studies of A/V integration in children. Research on infants clearly suggests that multisensory speech information is perceived. For example, Kuhl and Meltzoff (1982) demonstrated that infants looked longer at a face, the movements of which matched a speech sound, than at a face that did not match. Also, Rosenblum *et al.* (1997) reported results suggesting that infants are influenced by the McGurk effect (McGurk and MacDonald, 1976) whereby a given speech syllable is heard differently when presented synchronously with a video of a talker speaking a different speech syllable. In another study of a McGurk-like effect in infants, Desjardins and Werker (2004) showed that the effect is much less robust than in adults. Finally, a recent report by Hollich *et al.* (2005) suggested that infants might be able to segregate a target speech stream from a speech distracter if a synchronized video display of the target talker were present.

Results from studies of somewhat older children suggest that, although A/V synchrony may be perceived, A/V integration may not always be used to assist in auditory source segregation. Research with preschool and school-aged children (Desjardins *et al.*, 1997; Kishon-Rabin and Henkin, 2000; Massaro, 1984; Massaro *et al.*, 1986) reveals less A/V integration in children than in adults. The lack of A/V integration in children is not unexpected. For example, children are known to perform poorer than adults in tasks requiring face-processing (Aylward *et al.*, 2005; de Gelder *et al.*, 1998; Doherty-Sneddon *et al.*, 2001; Mondloch *et al.*, 2004; 2003; Schwarzer, 2000; Taylor *et al.*, 2004). Additionally, there are results which suggest that most children are relatively poor at speechreading (Massaro, 1984; Massaro *et al.*, 1986), a skill that is clearly necessary for A/V integration in speech tasks. Finally, the "auditory dominance effect" is much larger in children (Monsen and Engerbretson, 1983; Napolitano and Sloutsky, 2004; Robinson and Sloutsky, 2004; Sloutsky and Napolitano, 2003). Auditory dominance refers to the fact that when individuals are presented with simultaneous auditory and visual stimuli, attention is captured by the auditory stimulus.

Previous research on A/V integration, including that with infants and young children, has focused on listening in quiet, so the extent to which the results might generalize to more realistic noisy conditions is not clear. The purpose of the experiment described here is to study A/V integration (and the resultant release from informational masking) in children of various ages using a paradigm involving speech recognition in the presence of a speech distracter. The Coordinate Response Measure task, used in previous research from our laboratory (Wightman and Kistler, 2005), will be used here. This task has several advantages. First, a large number of studies with adults (Brungart, 2001a, b; Brungart and Simpson, 2002b, 2004, 2005; Brungart *et al.*, 2001, 2005; Kidd *et al.*, 2003, 2005a, b, c) indicates that performance in the CRM task is dominated by informational masking. Second, the amount of informational masking produced by the time-synchronized speech distracter in the CRM task is large, so that release from informational masking is easily measured (Arbogast *et al.*, 2002, 2005; Wightman and Kistler, 2005). Finally, as shown in our own study (Wightman and Kistler, 2005), reliable data revealing large amounts of informational masking can be obtained from children as young as 4 years performing in the task.

II. METHODS

A. Listeners

Ten adults and 23 school-aged children served as participants in this experiment. Four of the adults and 16 of the children had also served (9 months earlier) as listeners in our previous study using the CRM task (Wightman and Kistler, 2005). Children and adults were recruited from the University of Wisconsin and University of Louisville communities. The adults ranged in age from 18 to 31.9 years. For convenience in data interpretation, the children were divided into three age groups: six in the 6–8.9-year group, seven in the 9–11.9-year group, and ten in the 12–16.9-year group. All adults and children passed a 20 dB HL screening for hearing loss at octave frequencies from 0.25 to 8 kHz. All children passed the annual vision screening performed in their schools and no adult reported an uncorrected visual deficit. The children were tested for middle-ear problems (routine tympanometry) before the first session and again if necessary. None of the adults or children recruited was excluded due to inability to perform the task. One 11-year-old child did not complete the experiment due to scheduling difficulties.

B. Stimuli

Speech stimuli were taken from the corpus of high-quality, digitally recorded CRM stimuli made available by Bolia *et al.* (2000). The corpus includes 2048 phrases of the form, "Ready, *call sign*, go to *color number* now." Eight talkers (four male, four female) are recorded, each speaking 256 different phrases (eight different call signs, "baron," "ringo," "tango," etc.; eight numbers, 1–8; and four colors, red, white, green, blue). The target phrase was always spoken by talker 0 (male) from the corpus, using the call sign "baron." The distracter phrases always used a different male talker (1–3), call sign, color, and number. The distracter phrase used on each trial was chosen randomly with replacement from the available phrases.

As a check on the extent of energetic masking, some of the younger children were tested with a modulated noise distracter. The modulated noise was a speech-spectrum noise (long-term spectrum derived from the CRM phrases) with a temporal envelope determined by one of the distracter phrases, randomly selected on each trial as with the speech distracters. The envelope was derived by full-wave rectification and convolution with a 7.2-ms rectangular window, exactly as described by Brungart (2001b).

The audio stimulus materials were produced digitally (CRM stimuli taken from the distribution CD), converted to analog form (22 050-Hz sample rate) by a control PC, mixed, amplified, and presented to listeners via calibrated Beyer DT990-Pro headphones. The target and distracter phrases were time aligned on the distribution CD such that the word "ready" for target and distracter phrases started synchronously. Because of small differences in speaking rate and word length the durations of the phrases were slightly different.

The videos of the target talker (same talker as used for the target in the current experiment) were recorded in the corner of a large anechoic chamber at the Air Force Research Laboratory (Brungart and Simpson, 2005). The video was captured with a digital camera (Sony Digital Handycam) located roughly 1.5 m in front of the talker, who stood in front of a black, acoustically transparent background. The audio was recorded directly onto the videotape. The talker was instructed to repeat each of the 32 possible target messages in the CRM corpus (i.e., those with the target call sign "baron") at a monotone level while keeping his head as still as possible. Breaks were inserted between the CRM phrases to avoid any effects of coarticulation between consecutive recordings. The resulting videotapes were downloaded onto a PC where they were partitioned into individual AVI files for each of the 32 recorded phrases. Then a commercially available video editor (VirtualDub, www.virtualdub.org) was used to crop the frames of the AVI files around the locations of the talker's head, convert them from color to grayscale, and compress them into the Indeo 5.1 codec.

To produce the test stimuli for the current study, the original video tracks were individually time aligned with the corresponding audio target message from the original high-quality CRM audio corpus using custom MATLAB software. The audio track on the original videotape was not used because of its rather poor quality. Initially, the high-quality audio was inserted in place of the videotape audio such that the starting points of the two audio tracks were the same. Since the same highly practiced talker was used for both recordings this initial alignment was generally satisfactory. This was expected because this specific talker had previously been described as "extremely consistent" (Brungart and Simpson, 2005). In some cases, to compensate for slight differences in the duration of pauses between the original and the high-quality audio sentences, video frames were added (duplicates) or deleted using the VirtualDub software until alignment seemed perfect. No modifications were made to the

audio files. No more than 2 frames were added to or deleted from the video tracks at any single point. At the playback rate of 29 fps, this would amount to a time shift of less than 70 ms, which is about equal to the adult threshold for detection of A/V asynchrony (Grant *et al.*, 2004; Lewkowicz, 1996). Thus, for the adults, potentially detectable misalignments were probably rendered undetectable by the manipulation. Infant asynchrony detection thresholds are much larger (Lewkowicz, 1996), so there is no reason to expect that our child or adult listeners would perceive any misalignment of the audio and video. To check for proper alignment, several lab personnel viewed the videos as they were simultaneously presented with the audio. None reported a misalignment, although no formal assessment of detectability was made. The final time-aligned video file and the corresponding audio files (target and distracter) were combined into a single file (AVI format) on a trial-by-trial basis using VirtualDub software.

All conditions involved trials in which a single target and single distracter were presented to the listener's right ear. The overall level of the distracter in the target ear was held constant for all conditions at approximately 65 dB SPL. The level of the target was varied randomly from trial to trial in order to obtain complete psychometric functions, percent correct versus target/distracter ratio (T/D), from each listener. Depending on listener and condition the target/distracter ratio (T/D) ranged from −35 to +15 dB (in 5-dB steps). Thus the highest level of the target was 80 dB SPL.

C. Conditions

All listeners were tested in three conditions. In the video-only condition, no audio was presented and listeners responded to the video alone. This condition allowed assessment of a listener's untrained speechreading ability. Two other conditions were evaluated: the audio-only condition in which only the audio target and distracter messages were presented, and the audiovisual condition in which the audio was presented along with a simultaneous video of the target talker saying the target message. Data from the video-only condition were obtained in the first and last sessions. Data from the audio-only and audiovisual conditions were obtained in all sessions with the two conditions alternated.

Approximately 6–8 months after the completion of the experiment, seven of the younger listeners (four in the 6–8.9-year-old group and three in the 9–11.9-years-old group) were tested in a condition identical to the audio-only condition but using a modulated noise distracter. This condition, presumably involving mostly energetic masking, was included as a test of our assumption that masking in the audio-only condition and the audiovisual conditions was dominated by informational masking.

D. Procedure

Listeners sat in a sound-isolated room in front of a computer display. The display showed a start button and 32 response buttons arranged in four colored matrices of eight buttons each, numbered 1–8. For the conditions in which the video of the target talker was presented, a box approximately 6 in. wide and 8 in. tall appeared in the center of the screen in which the video could be seen. Individual trials were initiated by the listener by a mouse-click on the start button. After hearing the phrases, the listener moved the mouse cursor to the matrix of the heard color and clicked on the number corresponding to the heard number. No feedback was given regarding the correctness of the response. The response method was identical in the video-only condition. However, the listeners were told that they would see, but not hear, the "baron" talker. They were instructed to "watch the man's face" and respond according to what they thought the man said.

All participants completed a practice run of 30 trials listening to the target talker with no distracter to assure perfect performance at five levels ranging from 45 to 65 dB SPL. Next each listener completed a practice run of 60 trials in each of the audio-only and audiovisual conditions followed by a practice run of 30 trials in the video-only condition. The practice runs were used to determine the target-distracter (T/D) levels to be used for the test runs. In the audio-only condition, the level of the target was varied randomly (five or six levels) from trial to trial so that an entire psychometric function, from near perfect performance to chance, could be estimated during each session. The levels were 5 dB apart so that a 20-dB (five levels) or a 25-dB (six levels) range was covered. In the audiovisual condition all listeners completed runs in which the levels were 5 dB apart. Listeners who performed well above 50% correct at the lowest level were tested in one or more runs in which the levels were 10 dB apart (covering a 50-dB range) to determine if performance had reached plateau.

Both children and adults were tested in 60 or 120 trial blocks in the audio-only and audiovisual conditions and in 30 or 60 trial blocks in the video-only condition. Children completed 240–300 trials in each of the audio-only and audiovisual conditions and 30–90 trials in the video-only condition. Adults completed 120 trials in the video-only condition and 480 trials in the other two conditions.

Seven children were tested in the modulated noise condition. These children completed 210–240 trials in this condition, and, in the same session, an additional 180 trials in the audio-only condition.

Testing was conducted over the course of several sessions. Sessions were approximately 1–2 h long for both children and adults. Frequent breaks during each session were encouraged. Sessions were scheduled at the participant's convenience, usually once or twice per week. Most listeners completed the experiment in two to four sessions. Listeners were paid $8/h for their participation.

III. RESULTS AND DISCUSSION

A. Data analysis

Although complete psychometric functions were obtained from the listeners in all conditions, the irregular form of many of the functions led us not to fit them with a smooth curve (e.g., logistic) and extract parameters of the fitted functions. The irregularity was characterized in most cases by a plateau in performance at T/D ratios of between 0 and

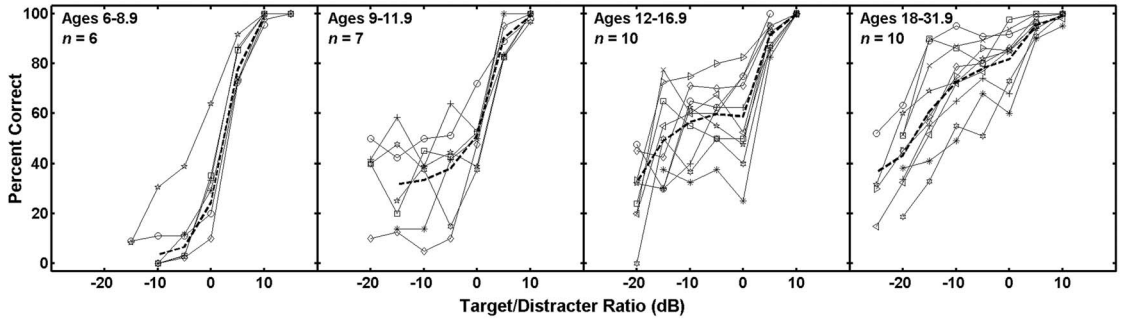


FIG. 1. Psychometric functions for individual listeners in the audio-only condition of the CRM speech recognition task. The four panels show data from listeners in the four different age groups. The different symbols in each panel represent different listeners. The mean psychometric function in each group is shown by the dashed line. The mean excludes data from the one "outlier" in the 6–9.9 years age group whose data are represented by open stars.

−10 dB. This plateau has been observed in several previous studies using both the CRM and other paradigms (Brungart, 2001b; Brungart and Simpson, 2002b; Dirks and Bower, 1969; Egan *et al.*, 1954; Wightman and Kistler, 2005). Here we will present complete psychometric functions.

Individual differences are large in informational masking studies, especially those involving children (Lutfi *et al.*, 2003; Oh *et al.*, 2001; Wightman *et al.*, 2003; Wightman and Kistler, 2005), so averaging must be done with caution. In a previous study involving similar conditions (Wightman and Kistler, 2005), we argued that averaging produced psychometric functions that were reasonable representations of the individual psychometric functions of the members of the age group. The results described here show larger intersubject differences so some individual psychometric functions will be discussed in addition to the averages.

To quantify the release from informational masking provided by the simultaneous video display, an "$A/V_{benefit}$" score (Grant and Seitz, 1998) was computed for each individual. The $A/V_{benefit}$ score is defined as

$$A/V_{benefit} = \frac{(A/V - A)}{(1 - A)}.$$

A/V and A are proportion correct recognition scores from the audiovisual and audio-only conditions, respectively, averaged over T/D ratios of −15 to 0 dB where informational masking is expected to be maximal. Thus the $A/V_{benefit}$ score reflects the difference between the recognition scores in the audiovisual and audio-only conditions relative to the amount of improvement possible given the audio-only score.

B. Audio-only condition

Figure 1 shows the data from individual listeners in each age group in the audio-only condition. In the data from each age group, the dashed line represents mean performance. As was the case in the previous study (Wightman and Kistler, 2005) individual differences are large. However, as before, we conclude that, in general, the mean psychometric functions appear to represent fairly the general shape of the psychometric functions in each group. The non-monotonicities (dips) in many of the individual functions (especially in the data from the 12–16-years olds) are not well captured by the mean. However, those non-monotonicities are not statistically significant. Because of the limited amount of data at each T/D ratio (less than 50 trials for the children) the confidence limits around each percent correct value near the middle of the function are quite large, in some cases more than ±15%.

The audio-only condition in this experiment is nearly identical to the "monaural" condition of our previous experiment (Wightman and Kistler, 2005). The only differences are procedural: the stimulus levels in this experiment were presented randomly rather than in an up-down staircase, and no feedback was given. Not surprisingly, since many of the same listeners participated in both experiments, the mean data from this condition match well with the mean data from the comparable condition of the previous experiment (Wightman and Kistler, 2005). Figure 2 shows the two sets of mean psychometric functions (means weighted according to the number of trials included for each listener at each T/D) with the previous data age-grouped according to the current scheme. The only obvious difference is the apparent lack of

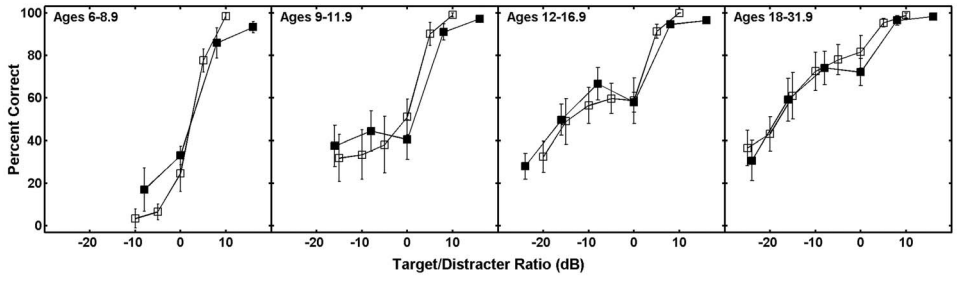


FIG. 2. Mean psychometric functions in the audio-only condition of the CRM task for listeners in each of the four age groups. Data from the current study (excluding that from the "outlier" shown in Fig. 1) are represented by open symbols and data from the previous study (Wightman and Kistler, 2005) by filled symbols. The error bars represent 95% confidence intervals for the mean.

the performance plateau in the data from the adults and older children at T/Ds from 0 to −10 dB. Although some individual listeners showed the plateau (Fig. 1), it was not as common an observation in this as compared to the earlier study. It seems reasonable to suggest that the use of the up-down staircase in the previous study may have facilitated the use of the level-difference segregation strategy whereby even though the target may be less intense than the distracter, it is still intelligible and can be recognized as the "softer talker." Random level presentation may not draw a listener's attention to this strategy, since the target levels do not systematically increase and decrease. It is also possible that the lack of feedback contributed to the apparent ineffectiveness of the level-difference segregation strategy. However, it should be noted that the overall level of performance was apparently not influenced by either the lack of feedback or the random level presentation. Other than the lack of a performance plateau in the current data, the two data sets are nearly identical.

In both data sets, and in our previous studies of informational masking with tonal stimuli (Oh *et al.*, 2001; Wightman *et al.*, 2003), individual variability was greater in the intermediate age groups than in either the adult group or the youngest group of children. This is inconsistent with data from other detection and discrimination studies from our laboratory in which variability is highest in preschoolers (Allen and Wightman, 1994, 1995; Allen *et al.*, 1989). However, it seems plausible that the non-monotonic change in variability could result from the developmental course of selective attention strategies. Attentional strategies for dealing with informational masking may not develop until the early school years, and then develop at different rates in different children. Supporting this view are the results of several studies (e.g., Geffen and Sexton, 1978; Geffen and Wale, 1979; Sexton and Geffen, 1979) which suggest that strategies for focusing attention do not begin to develop until about age 7 and continue to develop until the teenage years. Gibson (1969) argued that selective attention develops as a consequence of perceptual learning. If so it is reasonable to expect different rates of development in different children since each child's individual environment would play an important role in perceptual learning.

Figure 3 shows the data from the seven children who were tested with the modulated noise distracter and retested in the audio-only condition. Note that in each case, performance with the noise distracter was substantially better than with the speech distracter, amounting to a shift in the psychometric function of about 10 dB for all but one listener. In other words, the noise (presumed to be an energetic masker) was 10 dB less effective as a masker than the speech, supporting the view that masking in this experiment was dominated by informational masking. Figure 3 also shows that, for all but one listener, the retest performance in the audio-only condition was the same as in the original test. One listener (LBV) improved considerably at T/D ratios less than 0 dB; the performance plateau at T/D ratios between 0 and −10 dB improved by about 35%, reflecting much more adult-like performance (see Fig. 2). Also, note that the mean adult data from Brungart's modulated noise condition (Brungart, 2001b) are nearly identical to the modulated noise data from the seven children tested here (lower right panel of Fig. 3). This suggests that in the CRM task, children do not show substantially more energetic masking than adults.

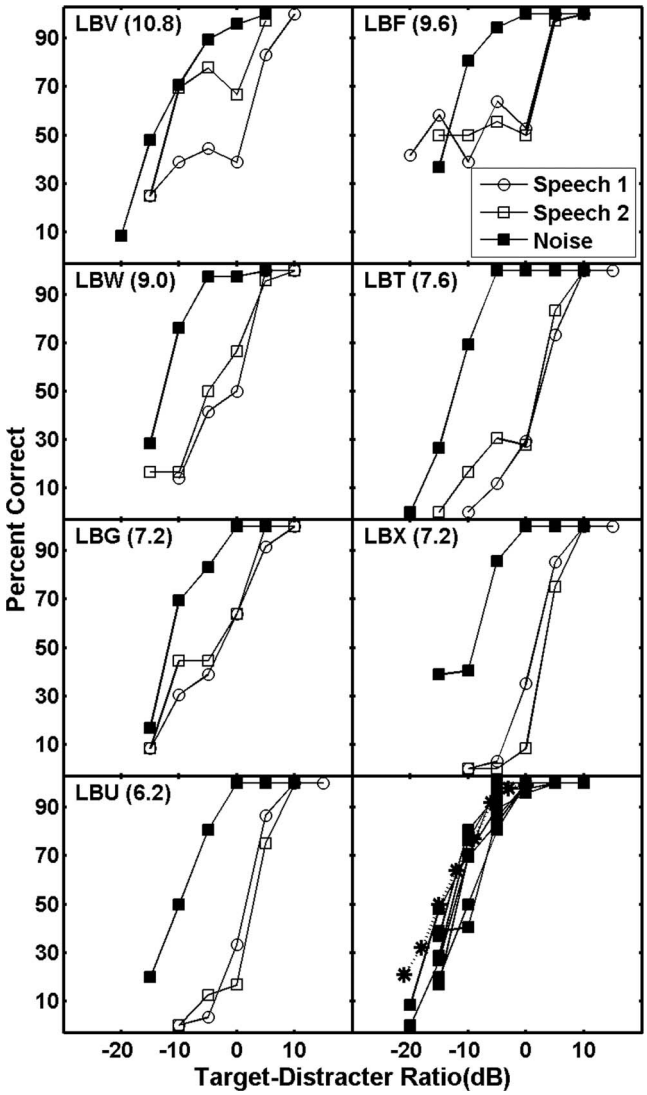


FIG. 3. Individual psychometric functions from seven children tested with modulated noise as a distracter (audio-only) and retested with the speech distracter. The original data obtained with a speech distracter are labeled "Speech 1" and the retest data are labeled "Speech 2." The bottom right panel shows all the individual modulated noise functions along with the data obtained in a similar condition by (Brungart, 2001b). Brungart's data are plotted as stars.

C. Video-only condition

Figure 4 shows a scatterplot of the scores in the video-only condition as a function of age. The individual differences in listeners' abilities to speechread in this experiment are striking. Among the adults, some scored as high as 85% but most were a bit lower. The lowest score in the adult group was about 48%. This result is consistent with that reported recently by Brungart and Simpson (2005) who also used the CRM task. Although there is an obvious age effect, with the younger children showing relatively poorer speechreading abilities, even among the youngest children (group median score of 10.8%) there is one who scored 80%.

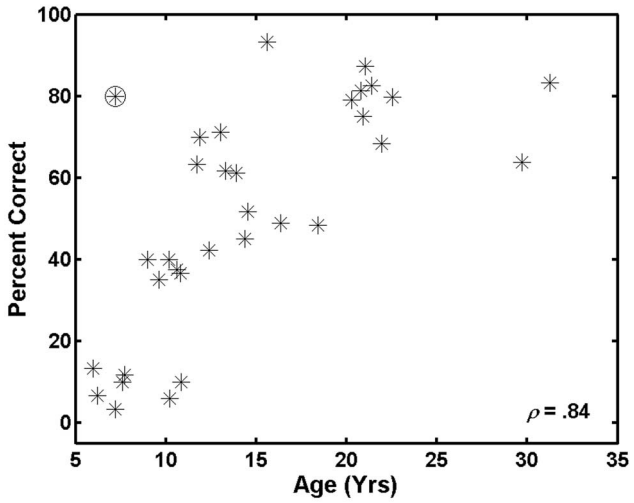


FIG. 4. Speechreading scores (video-only condition) for all listeners in the current study plotted as a function of age. The circled symbol represents the score for the young listener whose data were excluded from the mean computations shown in Figs. 1, 2, 5, and 6. The correlation listed in the inset is a Spearman *r*, computed with the data indicated by the circled symbol removed.

Large individual differences in speechreading ability are common. Watson *et al.* (1996) reported a range of scores from 50 subjects on a CID sentence test of speechreading from 4% to 87% total words correct (mean 37%; sd 17%). In a study of the speechreading ability of 60 undergraduates on a sentence test, Yakel *et al.* (2000) reported a mean score of 51% correct keywords reported with a standard deviation of 12 keywords. This suggests that more than 30% of the subjects had scores below about 39% or above 63%. These are but two examples of the results of the many experiments on speechreading in normal-hearing adults that show large individual differences [see Campbell *et al.* (1998) for a review of the classical work].

Individual differences in speechreading ability are also large in children. In one relatively recent study, Lyxell and Holmberg (2000) obtained a range of scores from 0% correct to 41% correct on a sentence test administered to 23 normal-hearing children. These results are consistent with those from the classic studies of children's speechreading reported by Massaro (1984; Massaro *et al.*, 1986) and demonstrate not only large individual differences but a level of speechreading proficiency that is generally lower in children than in adults. Both of these features of previous data can be seen in the data shown in Fig. 4.

D. Audiovisual condition

Figure 5 shows the individual performance of all listeners in the audiovisual condition. In the data from each age group, the dashed line represents mean performance. The data from the listener in the youngest group who performed much better than all the others were not included in the computation of the mean. This is the same listener who produced a video-only score of 80%. As was the case in the audio-only condition, large individual differences are evident in the data shown in Fig. 5, especially in the age 12–16.9-year group.

Mean psychometric functions from the audiovisual condition are plotted in Fig. 6. This figure also shows the mean functions from the audio-only condition and the mean speechreading scores. This figure clarifies the general developmental course of audiovisual integration and release from masking. The results suggest that on average there is no A/V release in children up to age 9, very little in children from 9 to 11.9, and much more in older children and adults. However, it seems clear that except for a few children in the age 12–16.9-year group, adultlike A/V integration and release from masking are not seen until the later teenage years. One should not expect much A/V integration when video encoding is lacking, as in those listeners who scored poorly in the video-only condition. Unfortunately, the current experiment does not allow us to identify whether a low $A/V_{benefit}$ score is a result of poor encoding, poor integration, or both.

Figure 6 suggests a strong relationship between A/V release from masking and speechreading ability, both of which increase with age. Recall that Fig. 4 shows the speechreading score for each participant as a function of age. Figure 7 shows the $A/V_{benefit}$ score for each participant as a function of age. The correlation between $A/V_{benefit}$ and age is 0.74. However, when speechreading score is factored out, the correlation of $A/V_{benefit}$ and age is only 0.03. The lack of correlation suggests that the age dependence we see in $A/V_{benefit}$ is almost entirely a result of the age-related changes in speechreading ability, which probably reflects age-related changes in the encoding of the visual information in speech. However, this does not imply that $A/V_{benefit}$ is determined only by speechreading ability. When the variance

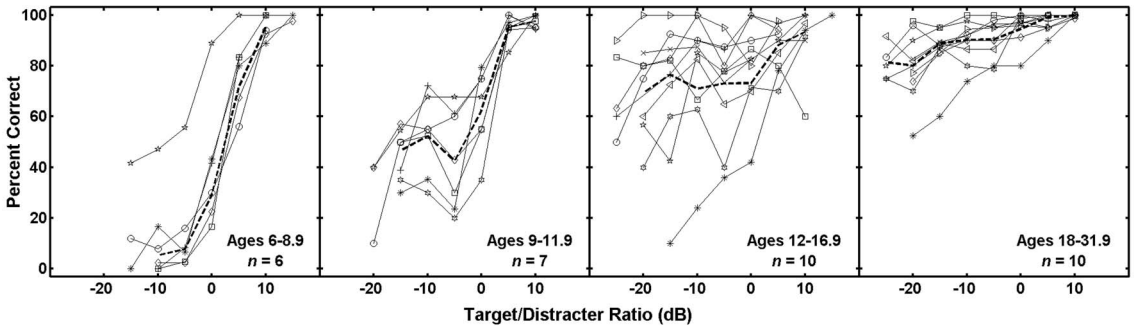


FIG. 5. Same as Fig. 1, except here the individual psychometric functions are from the audiovisual condition.

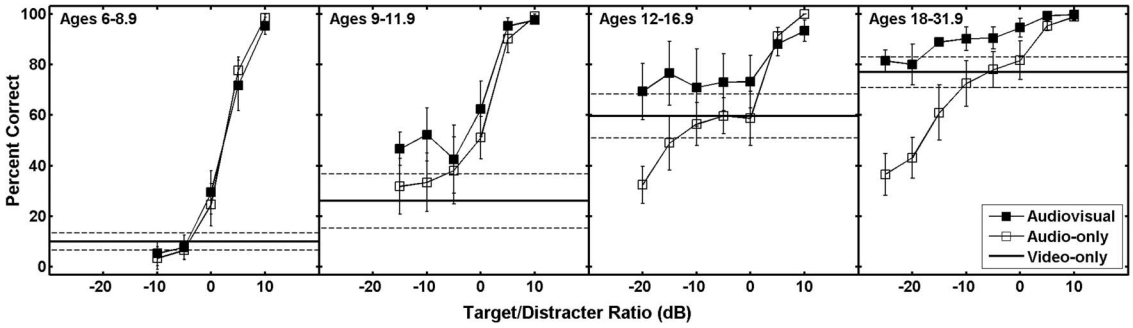


FIG. 6. Mean psychometric functions in the audio-only condition (open symbols) and the audiovisual condition (filled symbols) of the CRM task for listeners in each of the four age groups. Solid horizontal lines show mean performance in the video-only condition. The error bars and dashed lines represent 95% confidence intervals for the mean.

due to age is controlled, the correlation between $A/V_{benefit}$ and speechreading is 0.65, implying that other factors contribute to $A/V_{benefit}$. Thus, these data alone do not provide strong support for an age-related change in the ability of a listener to integrate auditory and visual information. This conclusion is identical to that reached by Sommers *et al.* (2005) in their study of younger and older adults.

Although interpreting data from individual participants is almost always problematic, there is one individual in this study who produced data that have very suggestive features. The data from this individual, a 7-year-old child, are shown in Fig. 8. The speechreading (video-only condition) score for this child was 80%, thus considerably better than any of the other children and on a par with scores from the adults. If speechreading were the sole determinant of A/V release from informational masking, we would expect a large A/V release from this individual. However, as the data show, the impact of the added video in the audiovisual condition was modest, producing about a 4–5-dB shift in the psychometric function. The $A/V_{benefit}$ score for this individual was only 0.35. Note that in the case of adults, for whom the mean speechreading score was almost 80%, the mean $A/V_{benefit}$ score was 0.65, and the lower end of the psychometric function in the audiovisual condition asymptotes at the speechreading score (Fig. 6). However, for the 7-year-old with an 80% speechreading score, the lower end of the psychometric function did not asymptote, suggesting a different strategy of combining the audio and video information. For this individual one might hypothesize that encoding of the video information was quite good, but integration of auditory and video information was less than optimal.

Given the importance of speechreading revealed by our data, we might speculate that speechreading training of young children may increase the release from informational masking they can achieve in everyday listening situations. The results from the study reported by Massaro *et al.* (1993) support this suggestion. In that study, speechreading training of adults on syllables, words, and sentences significantly improved A/V speech perception.

The results reported here do not permit a determination of the extent to which A/V integration reduced only informational masking. Many previous studies have reported an A/V benefit for adults recognizing speech targets in noise backgrounds (see, for example, Grant, 2002; Grant and Seitz,

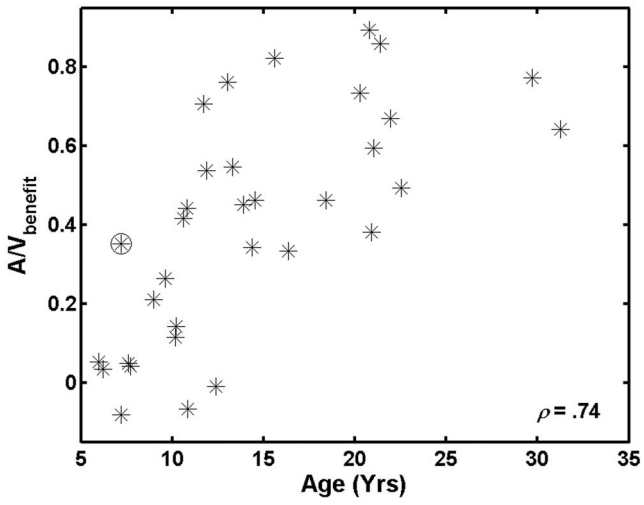


FIG. 7. $A/V_{Benefit}$ scores for all listeners in the current study plotted as a function of age. The circled symbol represents the score for the young listener whose data were excluded from the mean computation shown in Figs. 1, 2, 5, and 6. The correlation listed in the inset is a Spearman *r*, computed with the data indicated by the circled symbol removed.

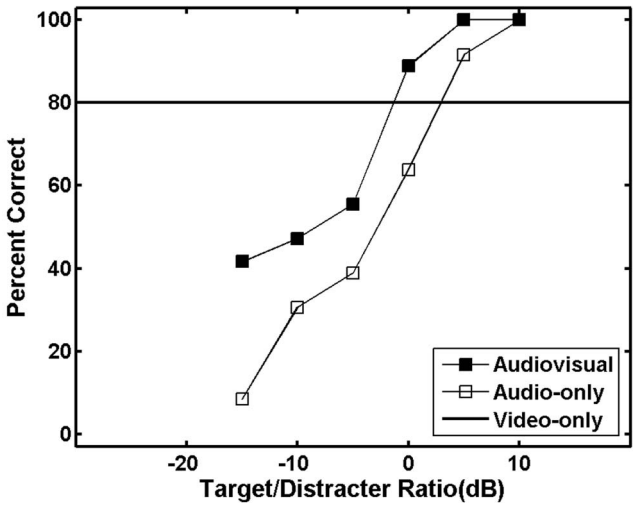


FIG. 8. Same as Fig. 5 except this figure includes only the data from the 7-years-old who performed better than the others in the same age group. This is the listener whose data were excluded from the mean computation shown in Figs. 1, 2, 5, and 6.

1998; Grant *et al.*, 1998; Sommers *et al.*, 2005; Sumby and Pollack, 1954). Our results are generally consistent with the previous findings. However, since noise was used as the masker in the previous studies, the contribution of informational masking in these studies was probably minimal. In our experiment, informational masking was dominant. Given the conditions studied here, we cannot provide independent measures of A/V release from energetic and informational masking.

A recent study of informational masking reported by Helfer and Freyman (2005) shows an A/V release of almost 10 dB with a two-talker distracter. This is quite similar to the extent of A/V release we report here (Fig. 6) with adult listeners. There are no previous studies of which we are aware on A/V release from informational masking in children.

IV. CONCLUSIONS

A/V release from informational masking was measured in 23 children and 10 adults using a task requiring recognition of a target speech message in the presence of a single speech competitor. As observed in previous studies, the adults tested here achieved a large release from informational masking in the task as a result of watching a video of the target talker during the task. The effect was largest at the lowest T/D ratios; at T/D=−20 dB, the video produced a mean increase of about 40% in the recognition score.

The youngest children obtained very little benefit from the video. This was at least in part a result of the fact that their speechreading scores were low. Obviously, to achieve a benefit, visual information must first be encoded, and with such low speechreading scores, there is little evidence that the youngest children were encoding information from the video. As mentioned earlier, several previous studies have also shown that children produce low speechreading scores (Lyxell and Holmberg, 2000; Massaro, 1984; Massaro *et al.*, 1986). The results of one particularly intriguing study (Doherty-Sneddon *et al.*, 2001) suggested that in some tasks looking at a face would actually interfere with a child's ability to attend to an auditory message.

Children in intermediate age ranges obtained a smaller but potentially useful release from informational masking with the video. In the 12–16-years-old group, the improvement at a T/D of −20 dB was about 30%, although individual differences were large. The dependence of the extent of release from informational masking on age was a result of the strong association of age and speechreading ability.

ACKNOWLEDGMENTS

The authors would like to thank Jen Junion Dienger, Rebekah Drska, the teachers in the University of Wisconsin Waisman Early Childhood Program, and the staff of the Heuser Hearing Institute for their contributions to the research. The research was supported financially by a grant to the first author from the National Institutes of Health (Grant No. R01-HD023333-14).

Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.

Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

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

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Aylward, E. H., Park, J. E., Field, K. M., Parsons, A. C., Richards, T. L., Cramer, S. C., and Meltzoff, A. N. (**2005**). "Brain activation during face perception: Evidence of a developmental change," J. Cogn Neurosci. **17**, 308–319.

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

Brungart, D. S. (**2001a**). "Evaluation of speech intelligibility with the coordinate response measure," J. Acoust. Soc. Am. **109**, 2276–2279.

Brungart, D. S. (**2001b**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2002a**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., and Simpson, B. D. (**2002b**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D. S., and Simpson, B. D. (**2004**). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. **115**, 301–310.

Brungart, D. S., and Simpson, B. D. (**2005**). "Interference from audio distracters during speechreading," J. Acoust. Soc. Am. **118**, 3889–3902.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., 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.

Campbell, R., Dodd, B., and Burnham, D. K. (Eds.). (**1998**). *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-Visual Speech* (Psychology Press, Hove, East Sussex, UK).

de Gelder, B., Vroomen, J., and Laeng, B. (**1998**). "Impaired speechreading related to arrested development of face processing," in Auditory-Visual Speech Processing (AVSP'98), edited by D. Burnham, J. Robert-Ribes, and E. Vatikiotis-Bateson (http://www.isca-speech.org/archive/avsp98, Terrigal-Sydney, Australia), pp. 157–161.

Desjardins, R. N., and Werker, J. F. (**2004**). "Is the integration of heard and seen speech mandatory for infants?" Dev. Psychobiol. **45**, 187–203.

Desjardins, R. N., Rogers, J., and Werker, J. F. (**1997**). "An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks," J. Exp. Child Psychol. **66**, 85–110.

Dirks, D. D., and Bower, D. R. (**1969**). "Masking effects of speech competing messages," J. Speech Hear. Res. **12**, 229–245.

Doherty-Sneddon, G., Bonner, L., and Bruce, V. (**2001**). "Cognitive demands of face monitoring: Evidence for visuospatial overload," Mem. Cognit. **29**, 909–919.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). "Some factors affecting multi-channel listening," J. Acoust. Soc. Am. **26**, 774–782.

Fallon, M., Trehub, S. E., and Schneider, B. A. (**2000**). "Children's perception of speech in multitalker babble," J. Acoust. Soc. Am. **108**, 3023–3029.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Geffen, G., and Sexton, M. A. (**1978**). "The development of auditory strategies of attention," Dev. Psychol. **14**, 11–17.
Geffen, G., and Wale, J. (**1979**). "Development of selective listening and hemispheric asymmetry," Dev. Psychol. **15**, 138–146.
Gibson, E. J. (**1969**). *Principles of Perceptual Learning and Development* (Appleton-Century-Crofts, New York).
Grant, K. W. (**2002**). "Measures of auditory-visual integration for speech understanding: A theoretical perspective," J. Acoust. Soc. Am. **112**, 30–33.
Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2450.
Grant, K. W., and Seitz, P. F. (**2000**). "The use of visible speech cues for improving auditory detection of spoken sentences," J. Acoust. Soc. Am. **108**, 1197–1208.
Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.
Grant, K. W., Wassenhove, V. v., and Poeppel, D. (**2004**). "Detection of auditory (cross-spectral) and auditory-visual (cross-modal) synchrony," Speech Commun. **44**, 43–53.
Hall, J. W., III, Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.
Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (**2002**). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," Ear Hear. **23**, 159–165.
Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.
Hollich, G., Newman, R. S., and Jusczyk, P. W. (**2005**). "Infants' use of synchronized visual information to separate streams of speech," Child Dev. **76**, 598–613.
Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005a**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.
Kidd, G. Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005b**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acust. Acta Acust. 91, 526–536.
Kidd, G. Jr., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005c**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.
Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.
Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.
Kishon-Rabin, L., and Henkin, Y. (**2000**). "Age-related changes in the visual perception of phonologically significant contrasts," Br. J. Audiol. **34**, 363–374.
Kuhl, P. K., and Meltzoff, A. N. (**1982**). "The bimodal perception of speech in infancy," Science **218**, 1138–1141.
Lewkowicz, D. J. (**1996**). "Perception of auditory-visual temporal synchrony in human infants," J. Exp. Psychol. Hum. Percept. Perform. **22**, 1094–1106.
Litovsky, R. Y. (**2005**). "Speech intelligibility and spatial release from masking in young children," J. Acoust. Soc. Am. **117**, 3091–3099.
Lutfi, R. A. (**1990**). "How much masking is informational masking?" J. Acoust. Soc. Am. **88**, 2607–2610.
Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003**). "One factor underlies individual differences in auditory informational masking within and across age groups," Percept. Psychophys. **65**, 396–406.
Lyxell, B., and Holmberg, I. (**2000**). "Visual speech reading and cognitive performance in hearing-impaired and normal hearing children (11–14 years)," Br. J. Educ. Psychol. **70**, 505–518.
Massaro, D. W. (**1984**). "Children's perception of visual and auditory speech," Child Dev. **55**, 1777–1788.
Massaro, D. W., Cohen, M. M., and Gesi, A. T. (**1993**). "Long-term training, transfer, and retention in learning to lipread," Percept. Psychophys. **53**, 549–562.
Massaro, D. W., Thompson, L. A., Barron, B., and Laren, E. (**1986**). "Developmental changes in visual and auditory contributions to speech perception," J. Exp. Child Psychol. **41**, 93–113.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Mondloch, C. J., Dobson, K. S., Parsons, J., and Maurer, D. (**2004**). "Why 8-years-olds cannot tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features," J. Exp. Child Psychol. **89**, 159–181.
Mondloch, C. J., Geldart, S., Maurer, D., and Le Grand, R. (**2003**). "Developmental changes in face processing skills," J. Exp. Child Psychol. **86**, 67–84.
Monsen, R. B., and Engerbretson, A. M. (**1983**). "The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction," J. Speech Hear. Res. **26**, 91–97.
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.
Napolitano, A. C., and Sloutsky, V. M. (**2004**). "Is a picture worth a thousand words? The flexible nature of modality dominance in young children," Child Dev. **75**, 1850–1870.
Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.
Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.
Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.
Robinson, C. W., and Sloutsky, V. M. (**2004**). "Auditory dominance and its change in the course of development," Child Dev. **75**, 1387–1401.
Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (**1997**). "The McGurk effect in infants," Percept. Psychophys. **59**, 347–357.
Schneider, B. A., Trehub, S. E., Morrongiello, B. A., and Thorpe, L. A. (**1989**). "Developmental changes in masked thresholds," J. Acoust. Soc. Am. **86**, 1733–1742.
Schwarzer, G. (**2000**). "Development of face processing: The effect of face inversion," Child Dev. **71**, 391–401.
Sexton, M. A., and Geffen, G. (**1979**). "Development of three strategies of attention in dichotic monitoring," Dev. Psychol. **15**, 299–310.
Slaney, M. (**1993**). *An efficient implementation of the Patterson-Holdsworth auditory filter bank* (35), Apple Computer, Inc., Cupertino, CA.
Slaney, M. (**1994**). Auditory Toolbox: Apple Computer, Inc., Cupertino, CA.
Sloutsky, V. M., and Napolitano, A. C. (**2003**). "Is a picture worth a thousand words? Preference for auditory modality in young children," Child Dev. **74**, 822–833.
Sommers, M. S., Tye-Murray, N., and Spehar, B. (**2005**). "Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults," Ear Hear. **26**, 263–275.
Sumby, W., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, Q. (**1979**). "Use of visual information for phonetic perception," Phonetica **36**, 314–331.
Taylor, M. J., Batty, M., and Itier, R. J. (**2004**). "The faces of development: A review of early face processing over childhood," J. Cogn Neurosci. **16**, 1426–1442.
van Wassenhove, V., Grant, K. W., and Poeppel, D. (**2005**). "Visual speech speeds up the neural processing of auditory speech," Proc. Natl. Acad. Sci. U.S.A. **102**, 1181–1186.
Watson, C. S., Qiu, W. W., Chamberlain, M. M., and Li, X. (**1996**). "Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**, 1153–1162.
Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distracters," J. Acoust. Soc. Am. **118**, 3164–3176.
Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.
Yakel, D. A., Rosenblum, L. D., and Fortier, M. A. (**2000**). "Effects of talker variability on speechreading," Percept. Psychophys. **62**, 1405–1412.

The detection of increments and decrements is not facilitated by abrupt onsets or offsets

Christopher J. Plack[a)]
Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom

Frederick J. Gallun
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

Ervin R. Hafter
Department of Psychology, University of California, Berkeley, California 94720-1650

Andrew Raimond
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom

(Received 27 September 2005; revised 27 March 2006; accepted 29 March 2006)

Two experiments measured thresholds for the detection of increments and decrements in the intensity of a quasi-continuous broadband-noise (experiment 1) or increments in a 477-Hz pure-tone pedestal (experiment 2). A variety of onset and offset ramps for the intensity change were tested, from instantaneous onsets or offsets to ramps lasting several tens of milliseconds. For increments and decrements with equal duration, the characteristics of the ramps had little effect on performance. Abrupt rise times, which are associated with strong transient responses in auditory neurons, did not facilitate detection in comparison to much slower rise times. The temporal window model of temporal resolution provided a good account of the data when the decision statistic was the maximum magnitude of the change in the output of the window produced by the increment or decrement, but provided a poor account of the data when the decision statistic was the maximum rate of change in the output of the window over time. Overall the results suggest that, in the absence of cues in the audio-frequency domain, rapid changes in envelope contribute little to near-threshold increment or decrement detection. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2198184]

PACS number(s): 43.66.Fe, 43.66.Ba, 43.66.Mk [JHG] Pages: 3950–3959

I. INTRODUCTION

The threshold for detecting a brief change in the intensity of a continuous, or quasi-continuous, pedestal has been widely used over recent years as a measure of temporal resolution and intensity coding in the auditory system (Bacon and Viemeister, 1994; Glasberg *et al.*, 2001; Heinz and Formby, 1999; Moore and Peters, 1997; Moore *et al.*, 1993, 1996, 1999; Oxenham, 1997; Peters *et al.*, 1995; Plack and Moore, 1991; Wojtczak *et al.*, 2003). However, there is still some debate regarding which cues are used by the listener to detect these intensity changes. Although in some cases increments and decrements may be detected by analysis of the spectral profile (Heinz and Formby, 1999), or by detection of the spectral splatter associated with abrupt intensity changes in narrow-band signals (Leshowitz and Wightman, 1972), this article will focus on "within-channel" cues that depend on the *temporal* profile or envelope of the sound. The two hypotheses to be investigated here are the following:

(i) Detection is based on the *magnitude* of the change.
(ii) Detection is based on the *rate* of the change.

These hypotheses can be considered in terms of the temporal window model of temporal resolution (Moore *et al.*, 1988; Oxenham and Moore, 1994; Plack and Oxenham, 1998; Plack *et al.*, 2002), which assumes that the internal representation of a stimulus is smoothed by a sliding temporal integrator to produce a "temporal excitation pattern" (TEP). The decision statistic is usually the maximum difference in magnitude at the output of the window between the stimulus with the intensity change and a reference stimulus without the intensity change. This statistic is consistent with hypothesis (*i*). In the case of the detection of brief increments or decrements, the statistic is effectively the size of the bump or dip in the TEP.

Oxenham (1997) measured the detection of increments and decrements in a 500-ms, 8-kHz sinusoidal pedestal. Increment and decrement durations from 2 to 16 ms (half-amplitude durations, 2-ms ramps) were tested. As duration increased, the smallest detectable change in intensity decreased. Oxenham simulated the data using a version of the temporal window model that incorporated an initial compressive nonlinearity representing cochlear processing. He found that the model did not produce a good description of the data when the decision statistic was the maximum difference in the output of the temporal window. Specifically, a version of the model using parameters fit to the decrement data overes-

[a)]Electronic mail: c.plack@lancs.ac.uk

 0001-4966/2006/119(6)/3950/10/$22.50

timated thresholds for the increment data. A similar discrepancy was shown when the results were compared to predictions of the model of Forrest and Green (1987). This model uses a half-wave rectifier followed by a low-pass filter to simulate limited temporal resolution. The decision statistic is the ratio of the maximum and minimum output of the filter, a criterion similar to the maximum difference statistic used by the temporal window model.

Oxenham found that the predictions of the temporal window model could be greatly improved if the decision statistic was not the maximum difference in the output of the window, but the maximum *rate of change* in the output of the window (i.e., the maximum slope of the TEP). This statistic is consistent with hypothesis (*ii*). In a subsequent paper, Oxenham (1998) found evidence for the use of abrupt changes (onsets and offsets) for the detection of signals in narrowband noise. Oxenham modeled these results by including information on the envelope spectrum in the region between 80 and 150 Hz, in other words, *high-frequency* envelope energy associated with rapid intensity changes. Other authors have also suggested that onset and/or offset information can be used to detect intensity changes. Macmillan (1971, 1973) proposed that the auditory system contains a change detector, utilizing neural transients, that is insensitive both to the *direction* of the change (i.e., whether it is an increment or a decrement) and to the *duration* of the change. The change detector is assumed to operate in conjunction with another detector, based on temporal integration, which is sensitive to change direction and duration. Carlyon and Moore (1986) measured Weber fractions for the detection of an increment of approximately 30-ms duration (including 5-ms onset and offset ramps) on a continuous 6.5-kHz pedestal. Masking the onset and offset ramps with bursts of bandpass noise produced an increase in the Weber fraction when the pedestal level was 55 or 80 dB SPL (although no increase was observed at 35 dB SPL). Carlyon and Moore suggested that listeners used the transient neural response to the onset of the increment as a cue. Finally, Moore and Peters (1997) showed that the detection of increments was much less dependent on increment duration than would be predicted by a detector using the overall energy in the increment. They also argued that their data were consistent with the use of onset information.

One way to test these hypotheses is to manipulate the rate of change in the envelope of the stimulus by varying the duration of the onset and offset ramps for the increment or decrement. If detection depends on an abrupt change in the stimulus or in its internal representation, then increasing the ramp duration should impair performance. On the other hand, if detection is based on the maximum magnitude of the change, then the characteristics of the ramps should be unimportant, as long as the maximum change stays constant.

In an investigation of the influence of spectral splatter on change detection, Macmillan (1973) measured increment and decrement detection with 1-kHz pure-tone pedestals, using stimuli with equal onset and offset ramp durations varying from 0.5 to 10 ms. For 25- and 50-ms nominal increment/decrement durations (steady-state durations were adjusted slightly to give equal overall energy for the different ramp conditions), thresholds increased slightly with increasing ramp duration. For example, for one listener for a 25-ms nominal duration, the Weber fraction $[10 \log_{10}(\Delta I/I)]$ for decrement detection was −9.5 dB for 0.5-ms ramps and −7.9 dB for 5-ms ramps. However, most differences were less than this. There is also the possibility, despite the use of a broadband noise masker in some conditions, that thresholds for the shortest ramp durations may have been affected by the detection of spectral splatter.

The effects of ramp duration have also been investigated in a gap-detection task. Ison *et al.* (2002) found that a temporal gap in broadband noise with a fast offset ramp (i.e., a fast rise time after the silent interval) produced greater inhibition of the acoustic startle reflex in the mouse than a gap with a slow offset ramp. This suggests that fast offsets are more salient. In a study on human listeners, however, Allen *et al.* (2002) reported no difference in detectability (using a "yes/no" task) between a gap in a broadband noise having a shallow onset ramp and a time-reversed stimulus with a shallow offset ramp, for ramp durations from 0 to 8 ms. In a second experiment, they found that gaps with "equally effective" durations (defined in one instance as the time interval between half-amplitude points on the ramps) were roughly equally detectable, irrespective of the onset and offset ramp durations. These results suggest that human listeners do not use the rate of change of intensity as a detection cue.

The present experiments extend the findings of Macmillan (1973), with a much larger range of ramp durations and the inclusion of asymmetrical ramp conditions. In addition, the first experiment used a broadband-noise pedestal to eliminate cues from spectral splatter. Experiments from two different laboratories are reported, each conducted in ignorance of the other. Although the experiments address similar issues, the procedures and stimuli used were different. We feel that the results complement each other and provide a more stringent test of the hypotheses than would be available from either experiment in isolation. Following Oxenham (1997), the data from both experiments were simulated using two versions of the temporal window model: one with a decision statistic based on the maximum difference in the output of the window and one with a decision statistic based on the maximum rate of change in the output of the window. This was done to validate the threshold measurements in terms of the expected internal representations of the stimuli and to provide a further test of the two hypotheses. For example, if detection is based on rate of change, then the model using rate of change as the decision statistic should provide the best fit to the data.

II. EXPERIMENT 1: INCREMENT AND DECREMENT DETECTION FOR BROADBAND NOISE

A. Stimuli

The pedestal was a broadband (0–20 kHz) Gaussian noise presented at a spectrum level of 20 dB ($re.\ 2 \times 10^{-5}\ \mathrm{N/m^2}$). The noise in the target interval consisted of a 5-ms raised-cosine onset ramp, a 200-ms steady-state portion, a portion containing an increment or decrement in intensity, a further 200-ms steady-state portion, and a 5-ms

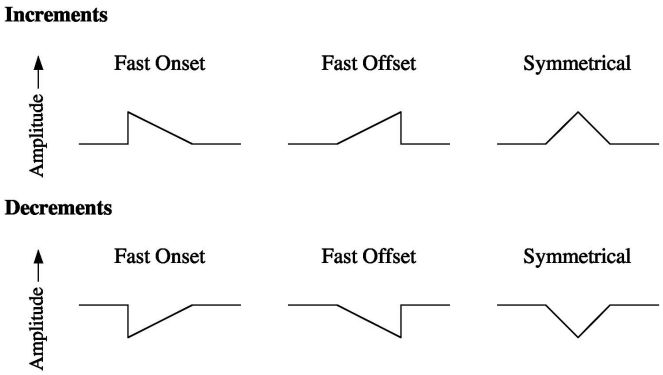


FIG. 1. A schematic illustration of the temporal envelopes of the increments and decrements presented in experiment 1.

raised-cosine offset ramp. The comparison interval was the same as the target interval, except that the increment/decrement was set to 0 dB. The increment or decrement was gated on and/or off with linear-amplitude ramps and had no steady state. The total duration of the increment/decrement was either 5, 10, 20, or 40 ms. In the "fast-onset" conditions, the onset ramp was instantaneous (except for any slight smoothing produced by the antialiasing filter), and the offset ramp had durations of 5, 10, 20, and 40 ms. In the "fast-offset" conditions, the offset ramp was instantaneous, and the onset ramp had durations of 5, 10, 20, and 40 ms. In the "symmetrical" conditions, the onset and offset ramps had the same duration. The ramp durations were 2.5, 5, 10, and 20 ms. Figure 1 illustrates the temporal characteristics of the stimuli.

All stimuli were generated digitally with 32-bit resolution and were output by an RME Digi96/8 PAD 24-bit soundcard at a clocking rate of 48 kHz. The headphone output of the soundcard was fed, via a patch-panel in a double-walled sound-attenuating booth, to Sennheiser 580 headphones. Stimuli were presented to the right ear.

B. Procedure

Increment and decrement thresholds were estimated using a two-interval, two-alternative, forced-choice procedure. The interstimulus interval was 500 ms. The level of the increment/decrement was defined as the difference between the pedestal level and the level at the time of maximum excursion of the increment or decrement (peak or trough level): in other words, the *maximum change* in dB. This value was varied using a geometric two-down one-up adaptive procedure that tracked the 70.7% correct point on the psychometric function (Levitt, 1971). The starting level was 20 dB. The step size was a factor of 2 for the first four reversals and a factor of 1.414 for the following 12 reversals. For each replication, the threshold estimate was the geometric mean of the last 12 reversals (expressed as ΔL in dB). At least six replications were conducted for each condition, and the final threshold was the geometric mean of the estimates from the individual replications. In each experimental session, one replication for each of the 24 conditions (two change directions×three ramp configurations×four change durations) was conducted, presented in a random order.

The experiment was controlled by custom-made software from a PC workstation located outside the sound booth. Listeners viewed a computer monitor through a window in the booth and decisions were recorded via a computer keyboard. Lights on the monitor display flashed on and off concurrently with each stimulus presentation and provided feedback at the end of each trial.

C. Listeners

Four young (21–23 years) normal-hearing listeners were tested. Audiometric thresholds were less than 15 dB HL for frequencies from 250 Hz to 8 kHz. Listeners were given two hours of training on the task before data collection began.

D. Results

The individual data are presented in Fig. 2. Thresholds are plotted with respect to the total duration of the increment or decrement (i.e., 5, 10, 20, or 40 ms for each of the conditions). Decrement thresholds are plotted as negative values. The mean standard errors of the data presented in the figure (calculated by taking the antilog of the standard errors of the logarithms of the dB values) were 1.12, 1.10, 1.17, and 1.14 for L1, L2, L3, and L4, respectively (these values can be used to multiply or divide the thresholds in dB to give the error above or below the threshold). The pattern of results is similar for the four listeners. Thresholds, expressed as absolute deviations in dB, decrease as duration is increased. Crucially, there appear to be no consistent differences between the different ramp configurations. For a given change direction and change duration, an abrupt onset did not produce lower thresholds compared to the time-reversed condition with a slow onset ramp or compared to the symmetrical condition with a ramped onset and offset. Expressed as absolute deviations in dB, thresholds are similar for the increment and decrement conditions, except for the 5-ms duration, for which the decrement threshold is higher. The finding that increments are more detectable than decrements at short durations is consistent with previous studies (de Boer, 1986; Oxenham, 1997).

A three-way analysis of variance was performed on these values, with factors change direction (increment or decrement), ramp configuration (fast onset, fast offset, or symmetrical), and change duration (5, 10, 20, or 40 ms). Absolute values were used, so that the magnitude of the effect could be compared between increments and decrements. The analysis revealed a highly significant main effect of change duration $[F(3,9)=111.870, p<0.00005]$ and a highly significant interaction between change direction and change duration $[F(3,9)=37.777, p<0.00005]$. None of the other effects, including those of the ramp configuration factor $[F(2,6)=0.604, p=0.577]$ and of the interaction of change direction and ramp configuration $[F(2,6)=0.908, p=0.453]$, were significant. *Posthoc* comparisons (Tukey) revealed significant differences ($p<0.05$) between all pairs of durations except 20 vs. 40 ms. A significant difference was also found between increments and decrements at the 5-ms duration, but not at the other durations.

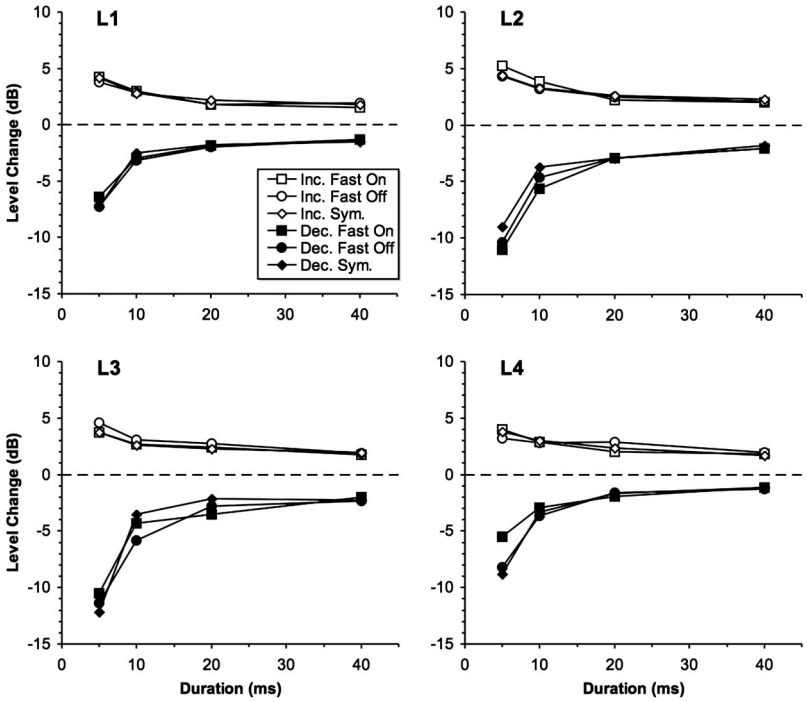


FIG. 2. The results of experiment 1 showing the threshold levels of the increments (open symbols) and decrements (filled symbols). Duration refers to the total duration of the increment or decrement.

III. EXPERIMENT 2: INCREMENT DETECTION FOR PURE TONES

A. Stimuli

The pedestal was a 477-Hz, 60-dB SPL pure tone, with a 980-ms steady state and 10-ms, raised-cosine, onset and offset ramps. The 477-Hz frequency was chosen for compatibility with an additional set of conditions being run on the same listeners (described in Gallun and Hafter, this issue). Increments were produced by adding a 477-Hz pure tone *in phase* to the pedestal. The total increment duration was either 20, 40, or 150 ms. The increments were gated with raised-cosine ramps, with no steady state. For the three "fast-onset" conditions, the ramp durations were 2-ms onset, 18-ms offset; 10-ms onset, 30-ms offset; and 10-ms onset, 140-ms offset. The three "fast-offset" conditions were time-reversed versions of the fast-onset conditions. The increments were temporally centered on the pedestal, i.e., the temporal center of the increment was 500 ms after the onset of the pedestal. Individual increments were set to have ΔL values of 0.5, 0.8, 1, 1.3, and 1.5 dB (defined in the same way as in experiment 1).

All stimuli were generated digitally with 16-bit resolution at a clocking rate of 50 kHz. They were output via a locally constructed 16-bit DAC with a low-pass reconstruction filter set to 20 kHz. Stimuli were presented monaurally over the left channel of Stax SR-5 electrostatic headphones to subjects seated in a double-walled sound-attenuating booth.

B. Procedure

On each trial the listener was presented with a single tone, either the pedestal alone or the pedestal plus increment (chosen at random). The listener was required to indicate whether or not the tone contained an increment by using a mouse to select the word "change" or "no change" on a computer screen. Only one increment level was presented in a block and a minimum of four blocks were obtained with that level. Individual conditions were run in blocks of 50 trials, with 25 trials containing the pedestal alone and 25 containing the pedestal plus increment.

The percentage correct responses for each condition and listener were used to calculate the detection index, d' (Green and Swets, 1966). Linear fits were made to psychometric functions plotting $\log_{10}(d')$ versus increment intensity in units of $10\log_{10}(\Delta I/I)$. If the average d' value was below 0.2 or above 3.5, the result was not included in the estimate of the psychometric function. Various attempts at fitting linear functions to the data suggested that converting from ΔL and excluding extreme d' values is necessary to ensure that the psychometric functions are well described by straight lines (which they were). A threshold for each listener was defined as the level associated with a d' of 1 on that listener's psychometric function. Thresholds were then converted back to ΔL values $[10\log_{10}((\Delta I+I)/I)]$ for ease of comparison with the data from experiment 1.

C. Listeners

Four young (18–24 years) normal-hearing listeners were tested. Audiometric thresholds in the test ear were less than 10 dB HL for frequencies from 125 Hz to 4 kHz. Extensive training preceded the experiment and a listener's performance was judged to be stable before data collection began.

D. Results

The individual data are presented in Fig. 3. Thresholds expressed as the size of the increment in dB [transformed from the $10\log_{10}(\Delta I/I)$ values] are plotted versus the total duration of the increments. For all three durations, there is no

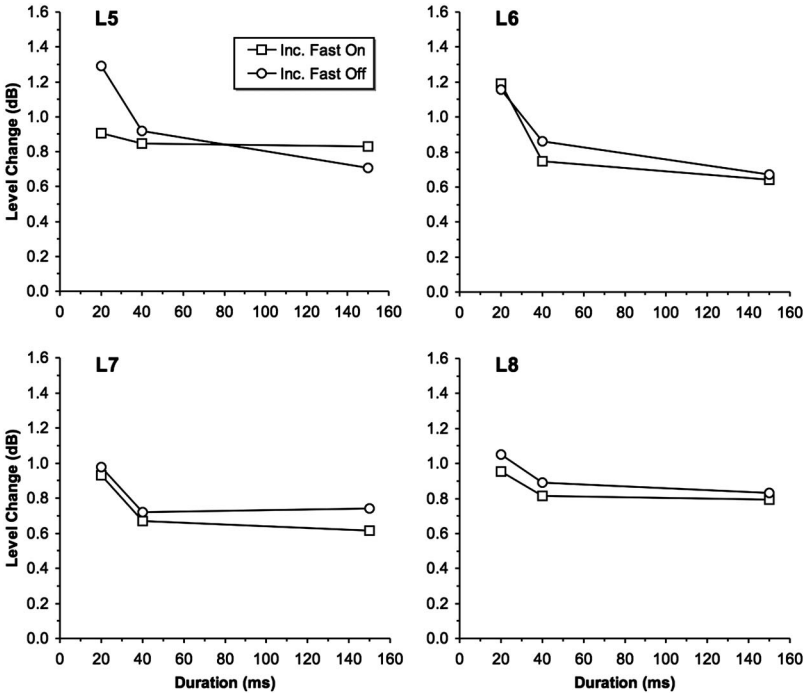


FIG. 3. The results of experiment 2, showing increment thresholds as a function of duration.

consistent difference between fast and slow onsets. One listener (L5) shows better performance for the fast onset at the shortest duration, but this does not occur at the other durations or for the other listeners.

A two-way analysis of variance was conducted on these values, with factors ramp configuration (fast onset or fast offset) and change duration (20, 40, or 150 ms). The main effect of duration was significant $[F(2,18)=16,599, p<0.001]$ but the effect of ramp configuration did not quite reach significance $[F(1,18)=4.194, p=0.055]$. This near-significant effect was due to the difference between L5's thresholds for the two ramp configurations at the shortest duration, which may be regarded as an outlier. There was no significant interaction between duration and ramp configuration $[F(2,18)=0.039, p=0.962]$. *Posthoc* analyses showed that the difference between the 20-ms signals and the other two durations was significant at the $p<0.01$ level, but there was no difference between the 40-ms signals and the 150-ms signals ($p=0.430$).

IV. ANALYSIS

Simulations with the temporal window model were used to determine whether a maximum difference, or maximum rate of change, decision statistic gave the best fit to the experimental results. In addition, envelope spectra were calculated to determine if the results were consistent with the use of high-frequency envelope information, as would be predicted by a detection mechanism utilizing transients. The analysis was intended to provide a further test of the two hypotheses.

A. Temporal window model

The model was based on that described by Plack and Oxenham (1998), and the reader is referred to that article for further details of the model and the fitting procedure. The only difference between this implementation and the earlier implementation was in the use of rate of change as an additional decision statistic. The parameters of the model [Eqs. (1)–(3)] were all taken from the earlier study.

The input to the model was the intensity envelope of the stimulus (ignoring fluctuations associated with the stochastic nature of the noise). The envelope was first subjected to a compressive nonlinearity to simulate cochlear processing:

$$I_{\text{out}} = I_{\text{in}}^{0.16}, \tag{1}$$

where I_{in} is the input intensity and I_{out} is the output of the compression stage. Following Plack and Oxenham (1998), the compression exponent is that derived by Oxenham and Plack (1997) from their behavioral measurements of cochlear nonlinearity.[1] The compressed waveform was then smoothed by a sliding temporal integrator or temporal window. The output of this stage (O) was the integrated product of the compressed waveform with the temporal window, as a function of the center time of the window. (This is equivalent to a convolution of the compressed waveform with the time-reversed temporal window.) The shape of the temporal window was described by a combination of exponential functions:

$$W(t) = 0.975 \exp(t/4) + 0.025 \exp(t/29) \tag{2}$$

for times before the center of the window ($t=0$) and

$$W(t) = \exp(-t/3.5) \tag{3}$$

for times after the center of the window, where $W(t)$ is an intensity-weighting function and t is time measured relative to the center (or maximum) of the weighting function (in ms).

In the first version of the model, the decision statistic was the *maximum difference* in the output of the model $[10 \log_{10}(O_T/O_C)$, where O_T is the output of the model in response to the target interval and O_C is the output of the

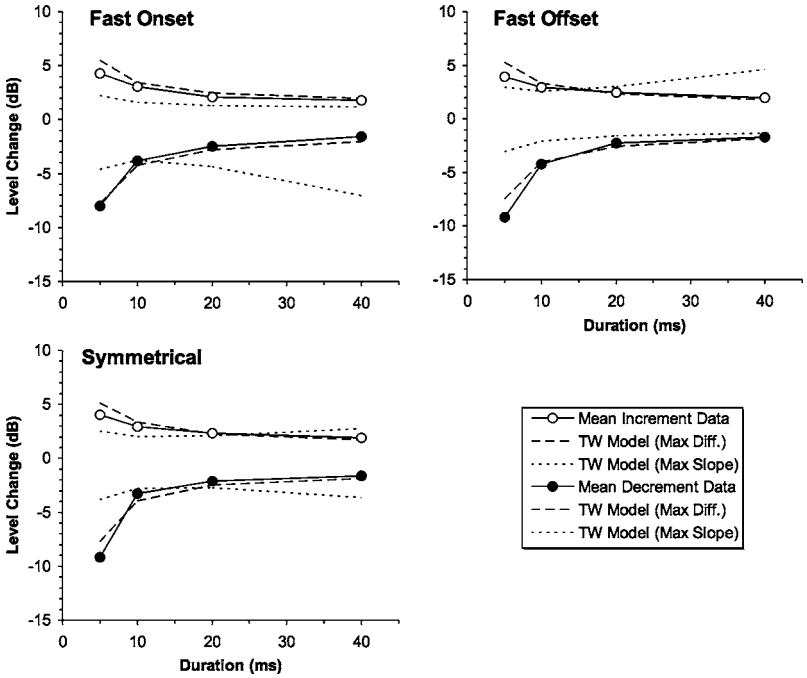


FIG. 4. The predictions of the temporal window model for the mean data from experiment 1. Continuous lines and symbols show the experimental data. Dashed lines show the thresholds predicted using a maximum difference decision statistic. Dotted lines show the thresholds predicted using a maximum slope decision statistic.

model in response to the comparison interval[2]] produced by the increment or decrement. For each condition, the target and the comparison intervals were passed through the model, and the center time of the window found for which the difference between the target and comparison intervals at the output of the model was greatest. For each condition, the level of the increment or decrement was varied adaptively until this difference was equal to the free parameter k. This produced a set of predicted thresholds for all the conditions. The process was repeated with different values of k, until the value of k was found that minimized the squared deviations of the predicted values from the mean values (across listeners) from the experiment. k was allowed to vary independently for the two experiments.

In the second version of the model, the decision statistic was the *maximum slope* in the output of the model $[d(10\log_{10}(O_T))/dt$, where t is time in ms] produced by the increment or decrement. Logarithmic units were used so that the model would predict a constant Weber fraction as a function of level.[3] For each condition, only the target interval was passed through the model, and the center time of the window found for which the slope at the output (the change in level between that center time and a center time one sample earlier) was greatest. For each condition, the level of the increment or decrement was varied adaptively until the slope value was equal to the free parameter s. The procedure of adaptively varying s to minimize the squared deviations was the same as that described above.

The predictions of the two versions of the model for experiment 1 are shown in Fig. 4. The first version of the model, using maximum difference as the decision statistic (the best-fitting value of k was 0.169 dB), accounts well for the results. There is a slight tendency for the model to overestimate thresholds for increments, and underestimate thresholds for decrements, at the shortest duration, but the discrepancy is smaller than that reported by Oxenham (1997). On the other hand, the second version of the model, using maximum slope as the decision statistic (the best-fitting value of s was 0.0151 dB/ms), provides a poor fit to the data. In particular, for the fast-offset increment conditions and fast-onset decrement conditions, the model predicts that threshold should *increase* for the longest durations (in direct contrast to the data). This is because shallow rising ramps tend to produce shallow slopes in the TEP, so the level change needs to be increased to match the criterion.

The predictions of the model for experiment 2 are shown in Fig. 5. These data were combined with additional thresholds for symmetrical conditions reported by Gallun and Hafter (this issue). The total increment durations for the additional data were 20, 50, and 85 ms, including 10-ms raised-cosine onset and offset ramps. In all other respects, the stimuli were the same as for experiment 2, as were the experimental procedure and listeners. The first version of the model (left-hand panel), using maximum difference as the decision statistic (k=0.087 dB), provides a good fit to the data. The second version of the model (left-hand panel), using maximum slope as the decision statistic (s =0.0026 dB/ms), provides a very poor fit to the data. Again, as ramp duration is increased in the fast-offset conditions, the model predicts that threshold should increase rather than decrease. Overall, analysis with the temporal window model provides partial support for the hypothesis that detection is based on the *magnitude* of the intensity change and evidence against the hypothesis that detection is based on the *rate* of the intensity change.

B. Envelope spectra

Another popular account of temporal resolution is the modulation filterbank model of Dau and colleagues (Dau *et al.*, 1997a, b). This model assumes that the auditory system exhibits frequency selectivity in the envelope domain, in that

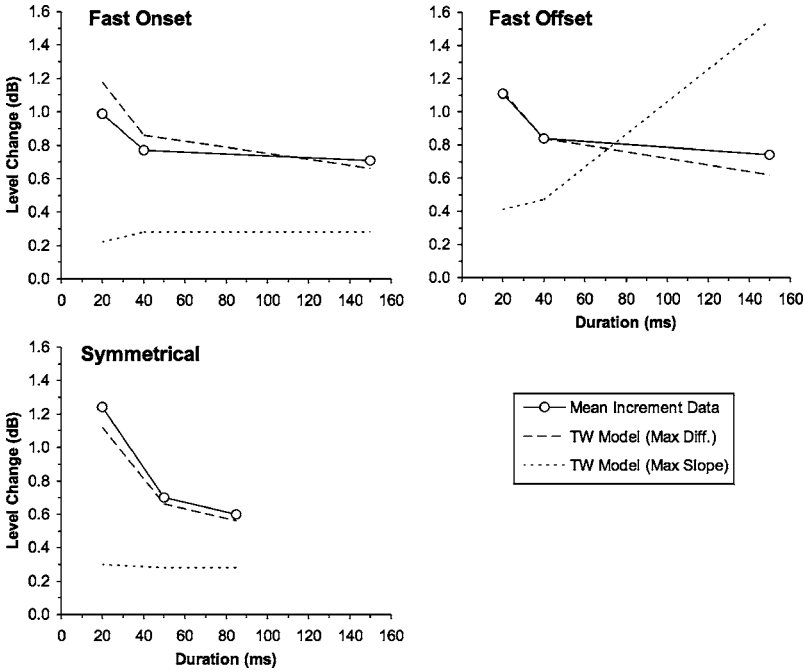


FIG. 5. The predictions of the temporal window model for the mean data from experiment 2, combined with symmetrical increment thresholds from Gallun and Hafter (this issue). Continuous lines and symbols show the experimental data. Dashed lines show the thresholds predicted using a maximum difference decision statistic. Dotted lines show the thresholds predicted using a maximum slope decision statistic.

the system has the ability to analyze selectively the output of a single modulation filter. To determine which modulation frequencies may have been used by the listeners in performing the tasks described here, a spectral analysis was performed on the envelopes of a selection of the increments from the two experiments and the additional increment stimuli reported by Gallun and Hafter (this issue) (fast-offset stimuli were not included because the magnitude spectra are identical to the equivalent fast-onset conditions). In each case, the analysis window contained only the portion of the stimulus that was incremented and not the preceding and following pedestal fringes. The size of the increments was fixed.

The resulting envelope spectra are shown in Fig. 6. Consider first the spectra for the stimuli from experiment 1. The stimuli with a 5-ms increment duration have similar spectra. This is expected, since the frequency axis only extends up to 250 Hz: The difference between the fast-onset (0-ms onset ramp) and symmetrical (2.5-ms onset ramp) stimuli is most evident at modulation frequencies higher than 250 Hz. On the other hand, there are differences in the spectra for the two stimuli with a 40-ms increment duration: As expected, the fast-onset stimulus has much more high-frequency energy.

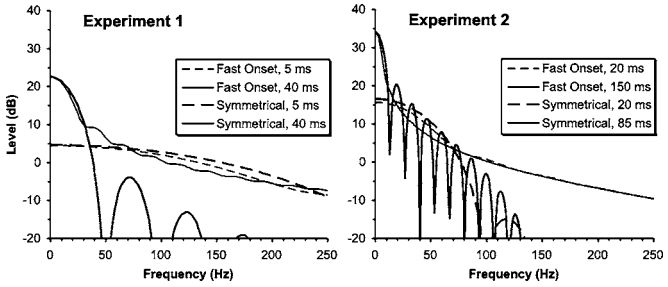


FIG. 6. Envelope spectra for a selection of the increments from the two experiments, and for the symmetrical stimuli reported by Gallun and Hafter (this issue) (presented with the experiment 2 spectra). The legend indicates the ramp condition and the total increment duration.

However, the thresholds for these two stimuli are almost identical (Fig. 2). Similar thresholds seem to correspond to similar amounts of low-frequency energy ($<$40 Hz) in the modulation spectra, rather than to similar amounts of high-frequency energy. The same pattern is observed in the spectra for the stimuli from experiment 2 and the stimuli reported by Gallun and Hafter (this issue). The 20-ms stimuli have similar thresholds (Fig. 5), and similar energy below 80 Hz. The 150-ms fast-onset stimulus and 85-ms symmetrical stimulus have similar thresholds, and similar energy below 90 Hz.

In summary, the high-frequency energy that is associated with abrupt intensity changes does not correlate well with threshold in the detection task. Instead, low thresholds seem to be associated with large amounts of low-frequency modulation energy. If the increments and decrements were detected via a modulation filterbank, it seems likely that filters with low center frequencies were used.

V. DISCUSSION

The results of the two experiments are consistent with the gap-detection experiments of Allen *et al.* (2002) that suggest that the main cue for the detection of intensity changes is the magnitude of the change in the stimulus rather than the rate of change of magnitude (or the slope of the TEP). The present results provide support for the use of magnitude of change as the decision statistic in models of temporal processing (Florentine *et al.*, 1999; Forrest and Green, 1987; Moore *et al.*, 1988; Oxenham and Moore, 1994; Plack and Oxenham, 1998). However, it is not easy to reconcile the modeling results with those of Oxenham (1997). His analysis, using a similar version of the temporal model to that employed here, was consistent with the use of the maximum *slope* of the TEP as the decision statistic. A version of the model using maximum *difference* as the decision statistic could not predict the differences in performance for incre-

ments and decrements. Part of the discrepancy may be related to the generally longer increment and decrement durations used in the present study compared to those that Oxenham used to test the temporal window model (6–18 ms). For the shortest duration in experiment 1 (5 ms), the version of the temporal window model using maximum difference as the decision statistic (Fig. 4) overestimated increment thresholds and underestimated decrement thresholds, a pattern at least qualitatively consistent with that reported by Oxenham. However, it is clear that the duration difference is not a satisfactory explanation for the size of the discrepancy, which may possibly be attributed to some other difference between the stimuli in the two studies (for example, the high-frequency sinusoidal pedestal used by Oxenham, or the lack of a steady-state portion in the present stimuli). It should be noted that Oxenham did not test explicitly the effects of rapid onsets and offsets by varying ramp duration. In other words, his data, while exposing a limitation in the model, do not address directly the issue of whether change detection benefits from rapid transients.

The present findings are easier to reconcile with the results of Moore and Peters (1997). These authors argued that onset information is more important than ongoing information, on the basis that the detection of increments is much less dependent on increment duration than would be predicted by a detector using the overall energy in the increment. However, the temporal window, being a "leaky" integrator of limited temporal extent, does not integrate the entire energy of the increment, and so predicts a smaller effect of duration than a nonleaky integrator. The temporal window model, using the maximum difference decision statistic ($k=0.479$ dB^4), was used to simulate the results of Moore and Peters for a 4-kHz pedestal, with a 200-ms pedestal fringe before increment/decrement onset. The model accounted well for the effects of duration (rms error of 0.62 dB), although again, at the shortest duration, the increment threshold was overestimated slightly and the decrement threshold was underestimated slightly. The main point, however, is that these data should not be used as evidence against the use of ongoing differences in intensity. For example, for the 210-ms increment, the optimum center time of the temporal window in the simulation was 180 ms after the start of the increment. In more recent work, Moore *et al.* (1999) and Glasberg *et al.* (2001) have shown that the effects of duration can be simulated using the temporal window model with a temporal template as the decision device. This model uses information regarding the intensity change over the whole duration of the increment or decrement.

There is some physiological evidence that the representation of decrements in auditory-nerve fibers is more sensitive to prior adaptation than the representation of increments. The change in firing rate produced by an increment in intensity is not affected by the duration of the preceding pedestal. However, measured over a 1-ms time window, the change in firing rate produced by a decrement in intensity decreases as the time interval between pedestal onset and decrement onset is increased from 0 to 45 ms. (Smith *et al.*, 1985). It is possible that the nonlinear effects of adaptation, a process not simulated in the temporal window model, may account for the failure of the model to provide accurate predictions at short durations (including the poor predictions reported by Oxenham, 1997). Moore and Peters (1997) found that increasing the pedestal fringe duration from 200 ms to continuous produced a marked increase in detection threshold for short decrements, an effect which cannot be explained by the temporal window model. They suggested that neural adaptation is the cause of the poor performance for short decrements.

Another aspect of intensity discrimination that is not explained by the temporal window model in its present form is the finding that, for changes of 50 ms and less, performance for detecting an increment or decrement is similar or superior to performance for discriminating the same increment from the same decrement (Macmillan, 1971, 1973). In contrast, the maximum difference statistic of the temporal window model predicts that, for a given size of increment and decrement, discrimination should be superior to detection. As suggested by Macmillan, the auditory system may combine information from a direction-sensitive detector with that from a direction-insensitive (or "change") detector. If this is the case, the present results suggest that the change detector is not dependent on rapid transients. In particular, it might be suggested that the spectral splatter associated with the 2-ms ramps that were used for the 20-ms pure-tone signals in experiment 2 would have provided a direction-insensitive cue to detection. However, the finding that thresholds for the 20-ms increments were significantly higher than the thresholds for the longer-duration increments (with shallow ramps) suggests that spectral splatter did not have a large effect on performance. In this respect, the conclusions are similar to those of Macmillan (1973).

A final complication to this story regards the effects of added noise. As described in the Introduction, Oxenham (1998) found evidence for the use of abrupt changes (onsets and offsets) for the detection of signals in narrow-band noise. Narrow-band noise contains low-frequency envelope fluctuations, which may effectively mask the low-frequency modulation energy in the signal. Hence, the high-frequency energy associated with the onsets and offsets of the signal may become a more salient cue. Similarly, in increment and decrement detection experiments with sinusoidal pedestals in which noise is added to mask spectral splatter, the low-frequency fluctuations in the noise at the output of the auditory filters may impair performance for low pedestal frequencies (see Glasberg *et al.*, 2001) and increase the importance of the high-frequency envelope information in the onsets and offsets. Although the pedestal used in experiment 1 in the present study was a broadband noise, we regard it as unlikely that fluctuations in the noise had a major effect on performance. Listeners were free to use a high-frequency, wide-bandwidth, auditory filter to detect the intensity changes. For a Gaussian noise, the proportion of low-frequency envelope energy decreases with increasing bandwidth.

Although the present results suggest that, at least in the absence of additional noise, abrupt onsets and offsets do not facilitate increment and decrement detection near threshold, physiological measurements have shown that rapid increases in intensity are particularly well represented by auditory neu-

rons. For example, Barsz *et al.* (1998) found that the representation in mouse inferior colliculus neurons was stronger for gaps with rapid offsets (fast rise times) than for stimuli with slow offsets: For gaps with slow offsets, a longer gap duration was required to produce a significant change in firing rate compared to a control stimulus without the gap. At all levels in the auditory system, neurons show a general tendency to produce a greater firing rate at stimulus onset. Indeed, there are some neurons, for example the onset units in the cochlear nucleus (Pfeiffer, 1966), which respond selectively to stimulus transients. However, it may sometimes be misleading to relate single-unit responses to the behavior of the whole organism. We do not yet have a clear understanding of how cortical centers make use of the multiple representations of acoustic information in the brainstem.

Furthermore, the physiological measurements were often made using stimuli with a high signal-to-noise ratio. It is possible that the physiology and the psychophysics might converge for supra-threshold stimulus encoding and discrimination. Fishbach *et al.* (2001) provide a neural model of auditory edge detection based on the detection of transients. Their model can account for the single-unit response to ramps of different durations, and also for psychophysical results such as the subjective categorization of a change between two intensities as continuous or discontinuous based on ramp duration, and the segregation of partials in a complex tone based on ramp duration (Bregman *et al.*, 1994). Hence the present findings should not be interpreted as evidence against the use of transients in general, but as evidence against the use of transients in the detection of *near-threshold* increments and decrements in intensity.

VI. CONCLUSIONS

(i) For increments and decrements with equal durations, detection threshold is affected very little by the characteristics of the onset and offset ramps. In particular, abrupt onsets or offsets do not facilitate detection.

(ii) A version of the temporal window model, using the maximum difference in the model output as the decision statistic, produced a good fit to the data. Another version, using the maximum slope in the model output over time as the decision statistic, produced a poor fit to the data.

(iii) A spectral analysis of the envelopes of the stimuli confirmed that the high-frequency components associated with rapid intensity changes are relatively unimportant in increment and decrement detection.

ACKNOWLEDGMENTS

The authors are grateful to Andrew Oxenham, John Grose, and two anonymous reviewers for helpful comments on an earlier version of this manuscript.

[1]In a subsequent analysis, versions of the model using compression exponents of 0.5 (amplitude envelope) and 1 (intensity envelope) were also fit to the data from experiment 1, using the maximum difference decision statistic. All three versions produced a good fit to the data, with rms errors of 0.70, 0.39, and 0.49 dB for compression exponents of 0.16, 0.5, and 1, respectively. Hence it appears that the predictions are not influenced greatly by the choice of compression exponent.

[2]An across-interval level comparison was used to simplify the analysis. A fit to the experiment 1 data was also conducted using a comparison of the change across *time* within each interval (from 100 ms after onset to the time of maximum change) as the decision statistic [specifically, $10\log_{10}(O_T/O_{T100}) - 10\log_{10}(O_C/O_{C100})$, where O_{T100} and O_{C100} are the model outputs for a center time 100 ms after onset, for the target and comparison intervals, respectively]. The predictions were almost identical to those from the model described in the main text, with an rms error of 0.70 dB in each case. Using change across time as the statistic would allow the model to make more accurate predictions for experiments in which level is roved between the two intervals.

[3]Use of dO_T/dt as the maximum slope statistic produced very similar results, with an almost identical goodness of fit [rms errors of 2.5 dB with the use of dO_T/dt, and 2.4 dB with the use of $d(10\log_{10}(O_T))/dt$].

[4]The best fitting k values vary between experiment 1 (0.169 dB), experiment 2 (0.087 dB), and the Moore and Peters results (0.479 dB). Larger k values reflect worse discrimination. The variability might reflect performance difference between individual listeners, although it seems more likely that differences between the stimuli had a greater effect. The broadband pedestals used in experiment 1 may provide more information across frequency channels than the pure-tone pedestals used by Moore and Peters. Although pure-tone pedestals were also used in experiment 2, the frequency was low and the envelope changes abrupt in some conditions leading to spectral splatter that may have aided detection. If stimulus variability is responsible for the variation in the k values, then this represents a limitation of the model in its present form.

Allen, P. D., Virag, T. M., and Ison, J. R. (**2002**). "Humans detect gaps in broadband noise according to effective gap duration without additional cues from abrupt envelope changes," J. Acoust. Soc. Am. **112**, 2967–2974.

Bacon, S. P., and Viemeister, N. F. (**1994**). "Intensity discrimination and increment detection at 16 kHz," J. Acoust. Soc. Am. **95**, 2616–2621.

Barsz, K., Benson, P. K., and Walton, J. P. (**1998**). "Gap encoding by inferior collicular neurons is altered by minimal changes in signal envelope," Hear. Res. **115**, 13–26.

Bregman, A. S., Ahad, P., Kim, J., and Melnerich, L. (**1994**). "Resetting the pitch-analysis system: 1. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone," Percept. Psychophys. **56**, 155–162.

Carlyon, R. P., and Moore, B. C. J. (**1986**). "Continuous versus gated pedestals and the 'severe departure' from Weber's Law," J. Acoust. Soc. Am. **79**, 453–460.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

de Boer, E. (**1986**). "On thresholds of short-duration intensity increments and decrements," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York), pp. 429–436.

Fishbach, A., Nelken, I., and Yeshurun, Y. (**2001**). "Auditory edge detection: A neural model for physiological and psychoacoustical responses to amplitude transients," J. Neurophysiol. **85**, 2303–2323.

Florentine, M., Buus, S., and Geng, W. (**1999**). "Psychometric functions for gap detection in a yes-no procedure," J. Acoust. Soc. Am. **106**, 3512–3520.

Forrest, T. G., and Green, D. M. (**1987**). "Detection of partially filled gaps in noise and the temporal modulation transfer function," J. Acoust. Soc. Am. **82**, 1933–1943.

Gallun, F. J., and Hafter, E. R. (**this issue**). "Amplitude modulation as a mechanism for increment detection," J. Acoust. Soc. Am..

Glasberg, B. R., Moore, B. C. J., and Peters, R. W. (**2001**). "The influence of external and internal noise on the detection of increments and decrements in the level of sinusoids," Hear. Res. **155**, 41–53.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Heinz, M. G., and Formby, C. (**1999**). "Detection of time- and bandlimited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**, 313–326.

Ison, J. R., Castro, J., Allen, P. D., Virag, T. M., and Walton, J. P. (**2002**). "The relative detectability for mice of gaps having different ramp durations at their onset and offset boundaries," J. Acoust. Soc. Am. **112**, 740–747.

Leshowitz, B., and Wightman, F. L. (**1972**). "On the importance of considering the signal's frequency spectrum: Some comments on Macmillan's 'Detection and recognition of increments and decrements in auditory intensity' experiment," Percept. Psychophys. **12**, 209–210.

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

Macmillan, N. A. (**1971**). "Detection and recognition of increments and decrements in auditory intensity," Percept. Psychophys. **10**, 233–238.

Macmillan, N. A. (**1973**). "Detection and recognition of intensity changes in tone and noise: The detection-recognition disparity," Percept. Psychophys. **13**, 65–75.

Moore, B. C. J., and Peters, R. W. (**1997**). "Detection of increments and decrements in sinusoids as a function of frequency, increment, and decrement duration and pedestal duration," J. Acoust. Soc. Am. **102**, 2954–2965.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). "Effects of frequency on the detection of decrements and increments in sinusoids," J. Acoust. Soc. Am. **94**, 3190–3198.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1996**). "Detection of decrements and increments in sinusoids at high overall levels," J. Acoust. Soc. Am. **99**, 3669–3677.

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**, 3539–3552.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.

Oxenham, A. J. (**1997**). "Increment and decrement detection in sinusoids as a measure of temporal resolution," J. Acoust. Soc. Am. **102**, 1779–1790.

Oxenham, A. J. (**1998**). "Temporal integration at 6 kHz as a function of masker bandwidth," J. Acoust. Soc. Am. **103**, 1033–1042.

Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.

Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (**1995**). "Effects of level and frequency on the detection of decrements and increments in sinusoids," J. Acoust. Soc. Am. **97**, 3791–3799.

Pfeiffer, R. R. (**1966**). "Classification of response patterns of spike discharges for units in the cochlear nucleus: tone-burst stimulation," Exp. Brain Res. **1**, 220–235.

Plack, C. J., and Moore, B. C. J. (**1991**). "Decrement detection in normal and impaired ears," J. Acoust. Soc. Am. **90**, 3069–3076.

Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.

Plack, C. J., Oxenham, A. J., and Drga, V. (**2002**). "Linear and nonlinear processes in temporal masking," Acustica **88**, 348–358.

Smith, R. L., Brachman, M. L., and Frisina, R. D. (**1985**). "Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments," J. Acoust. Soc. Am. **78**, 1310–1316.

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.

Effect of masker-frequency variability on the detection performance of infants and adults[a)]

Lori J. Leibold[b)] and Lynne A. Werner
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington, 98195

(Received 4 April 2005; revised 5 April 2006; accepted 6 April 2006)

The effect of masker-frequency variability on the detection performance of 7–9 month-old infants and adults was examined. Listeners detected a 300-ms 1000-Hz pure tone masked by: (1) A random-frequency two-tone complex; (2) a fixed-frequency two-tone complex; or (3) a broadband noise. Maskers repeated at 300-ms intervals throughout testing at 60 dB SPL. The signal was presented simultaneously with one presentation of the masker. Thresholds were determined adaptively using an observer-based method. Infants' thresholds were higher than adults' in all conditions, but infants' and adults' thresholds changed with masker condition in qualitatively similar ways. The fixed two-tone complex produced masking for both age groups, but more masking for infants than for adults. For infants and adults, the random two-tone complex produced more masking than broadband noise, but the difference was greater for infants than for adults. For infants and adults, the random two-tone complex produced more masking than the fixed two-tone complex, and the difference between these conditions was similar for both age groups. These results suggest that infants are more susceptible to informational masking than adults in the absence of spectral variability. Whether infants are more susceptible to the effects of masker-frequency variability than adults remains to be clarified. [DOI: 10.1121/1.2200150]

PACS number(s): 43.66.Lj [RAL] Pages: 3960–3970

I. INTRODUCTION

Acoustic variability is inherent in the natural environment of a developing infant. Relevant signals can occur at unpredictable times, particularly when the listener has limited experience with sounds. Moreover, the spectral, amplitude, and temporal characteristics of these signals are dynamic and variable. Understanding how infants learn to identify and distinguish important acoustic signals from a variable background is critical if we hope to understand how hearing develops.

Many investigators have examined the effects of stimulus variability on the detection and discrimination performance of trained adults (e.g., Watson *et al.*, 1975; Neff and Green, 1987; Neff and Callaghan, 1988; Neff and Dethlefs, 1995; Kidd *et al.*, 1994; Kidd *et al.*, 2003; Oh and Lutfi, 1998; Richards and Neff, 2004; Alexander and Lutfi, 2004). In this context, stimulus variability is created by randomly changing the acoustic properties of either the relevant target (the signal) or the irrelevant background (the masker). Considerable masking can be produced by varying the frequency content of the masker with each presentation; threshold elevations as high as 50 dB have been observed for fixed-frequency signals in the presence of random-frequency multitone maskers (e.g., Neff and Green, 1987). Unlike peripheral or "energetic" masking, this masking can be observed when the frequency components that comprise the masker are located well outside the auditory filter centered on the signal. That is, masking is observed even though the peripheral auditory system provides sufficient resolution to prevent the masker from affecting the audibility of the signal. In the literature, masking produced in excess of energetic masking (Fletcher, 1940) has been referred to as "informational" masking (reviewed by Durlach *et al.*, 2003b).

Informational masking is widely believed to involve central auditory processes and may reflect difficulties in such complex auditory skills as sound source segregation (e.g., Kidd *et al.*, 1994; Neff, 1995)—the process by which acoustic components are identified as coming from one or more sound sources—and analytic listening strategy (e.g., Lutfi *et al.*, 2003b)—the extent to which a listener attends to the relevant signal of a complex sound and ignores irrelevant maskers. However, the mechanisms responsible for informational masking are not fully understood. In part, this lack of understanding reflects the difficulty involved in accounting for the large differences in performance observed across individuals.

Recently, investigators have had considerable success accounting for these large individual differences using Lutfi's component-relative-entropy (CoRE) model (Lutfi, 1993). The CoRE model is similar to models of energy detection in that it assumes detection is based on the energy output of the auditory filter centered on the signal frequency. However, the CoRE model differs from traditional energy-detector models by considering potential contributions from auditory filters not centered on the signal frequency. Investigations by Lutfi

[a)]Portions of this article are based upon a thesis by the first author submitted to the Graduate School of the University of Washington in partial fulfillment of the requirements for the Doctor of Philosophy degree. Portions of these results were presented at the Acoustical Society of America meeting in Nashville, TN in May 2003 [Leibold and Werner, J. Acoust. Soc. Am.**113**, 2208 (2003)] and at the Scientific and Technology Meeting of the American Auditory Society in Scottsdale, AZ in March 2005.

[b)]Current affiliation: Boys Town National Research Hospital, Omaha, NE 68131. Electronic mail: leiboldl@boystown.org

and his colleagues have demonstrated that individual differences in informational masking are largely related to differences in the number and the frequency range of monitored auditory filters. For example, Alexander and Lutfi (2004) have shown that weighting efficiency, the extent to which a listener exclusively monitors the auditory filter centered on the signal frequency, successfully predicted the amount of informational masking. Listeners with high weighting efficiencies were resistant to informational masking whereas listeners with low weighting efficiencies were susceptible to informational masking. Presumably, adult listeners generally have high weighting efficiencies when masker variability is reduced.

Although decision weights have not been obtained from infants under similar conditions, perceptual development might be expected to influence performance in the presence of masker-frequency variability. It is interesting to note similarities between adults' performance in informational masking tasks and infants' performance in conventional psychoacoustic tasks, including elevated detection thresholds and large individual differences. Several findings suggest that infants' weighting efficiency under conditions of low stimulus variability is similar to that of adults under conditions of high stimulus variability. For example, Bargones and Werner (1994) measured "listening bands" for 7- to 9-month-old infants and for adults using the probe-signal paradigm (Greenberg and Larkin, 1968). During a typical probe-signal task in the frequency domain, listeners detect tones in noise, and a single "expected" frequency target signal is presented on the majority of trials. Probes at other "unexpected" frequencies are presented on a minority of trials. In the Bargones and Werner study, adults detected the expected signal at a higher rate than the unexpected probes. This result is consistent with previous results from other laboratories and indicates that adults listen selectively for an expected frequency in a detection task, effectively enhancing detection of the expected frequency at the expense of unexpected frequencies (e.g., Dai *et al.*, 1991). In terms of the CoRE model, adults do not detect unexpected frequencies because they give little weight to the output of filters centered on the unexpected frequencies. In sharp contrast, infants did not appear to listen selectively in the frequency domain. Infants detected expected and unexpected frequencies equally well, suggesting that they give weight to auditory filters centered at unexpected frequencies.

A consequence of infants' inefficient weighting strategy is that they exhibit masking in conditions where neither energetic nor informational masking has been observed in adults. For example, Werner and Bargones (1991) have described "distraction effects" in 6-month-old infants. Thresholds for a 1000-Hz pure tone were measured in quiet and in the presence of a bandpass noise. The noise (4000–10,000 Hz) was higher in frequency than the 1000-Hz tone and was presented at an overall level of 40 dB SPL. Adult thresholds were similar in noise and in quiet. In contrast, infant thresholds were elevated by approximately 10 dB in the presence of the noise relative to thresholds obtained in quiet. Furthermore, infant thresholds did not change when the level of the noise was changed from 40 to 50 dB SPL. Given that the frequency of the noise band was remote from the auditory filter of the 1000-Hz signal and that increasing the level of the noise did not further increase infants' thresholds, it is difficult to attribute the threshold elevations to energetic masking. Investigations of infants' psychometric function for detection also support the idea that infants use an inefficient weighting strategy (Bargones *et al.*, 1995; Werner and Boike, 2001). However, note that infants apparently have a low weighting efficiency for pure-tone detection when the masker frequency is fixed. Put another way, infants act like uncertain adults even when the stimuli are not variable. We use the term "variability" to refer to the characteristics of the sound, and the term "uncertainty" to refer to the presumed variability in the sensory effects of sound, because effects of internal and external variability might vary with development.

Although informational masking has not been systematically examined during infancy, it has been examined during the preschool period (e.g., Allen and Wightman, 1995; Hall *et al.*, 2005; Oh *et al.*, 2001; Wightman *et al.*, 2003; Wightman and Kistler, 2005). Studies have consistently reported that young children are more susceptible to the effects of masker-frequency variability than adults. For example, Allen and Wightman (1995) obtained detection thresholds for a 1000-Hz pure tone in the presence of broadband noise alone and in the presence of a broadband noise plus a one-component random-frequency tonal masker in a group of adults and a group of 3–5-year-old children. The average adult threshold obtained in the combined noise and random tone condition was 66.2 dB SPL compared with an average child threshold of 81 dB SPL. Note that the data of only 7 of 17 preschoolers tested were used to calculate this average threshold. The remaining children had detection thresholds above 90 dB SPL. Thus, children generally performed much worse than adults. However, large individual differences in performance were evident for both age groups. Among the seven children who provided measurable thresholds, the difference in threshold between the noise-plus distracter and the noise-alone conditions was often similar to that observed for many of the adults. In another study, Oh *et al.*, 2001 obtained thresholds from preschoolers and adults for a 1000-Hz pure tone presented simultaneously with either a broadband or a multitone masker comprised of 2 to 906 components. The amplitude and frequency of each component of the multitone masker varied randomly on each presentation. Both groups demonstrated large individual differences in threshold in the presence of the multitone masker. However, the average child's threshold was worse than the average adult's threshold. For some children, as much as 83 dB of masking was observed. In contrast, only small age differences in masking were observed for the broadband noise masker. Recently, Lutfi *et al.* (2003b) have asked whether individual differences in informational masking, including differences between children and adults, reflected fundamental differences in detection strategy rather than age differences in weighting efficiency. Principal component (PC) analysis was used to determine the number of PCs that were required to account for the variance in individual informational masking. The hypothesis was that if children and adults used qualitatively different detection strategies, at least two PCs would be re-

quired. The results indicated that a single PC accounted for 83% of the variance within and across age groups, and that PC was highly correlated with the number of auditory filters the listener monitored in detection. Preschoolers generally monitored a greater number of filters than did adults, suggesting that the apparent increase in susceptibility to informational masking observed during the preschool period occurs because children listen less selectively and monitor more filters than adults do during these tasks. This finding is consistent with those of Werner and Bargones (1991) and Bargones and Werner (1994), except that preschoolers appear to monitor a larger number of filters under conditions of masker-frequency variability while infants appear to do so even when masker frequency is fixed. To our knowledge, no published studies have examined preschoolers' listening strategies for the detection of a fixed-frequency pure tone signal in the presence of fixed-frequency remote maskers.

The purpose of the current investigation was to examine the detection performance of infants under conditions of masker-frequency variability. It was difficult to predict how infants would perform under such conditions. Given the apparent susceptibility to informational masking observed during the preschool period, informational masking effects might be very large during infancy, with large differences in performance across infant subjects. Alternatively, variability is often high during times of rapid maturation. Perhaps listening abilities develop more slowly during infancy than during the preschool period. As a result, infants' performance might be consistently poor. Finally, infants appear to use an inefficient weighting strategy in the absence of masker-frequency variability, monitoring irrelevant auditory filters. One explanation for this phenomenon could be that infants' internal representation of the stimulus varies, even when the stimulus does not. The introduction of physical stimulus variability might have little additional effect on infants' performance.

II. METHOD

A. Listeners

The listeners were 33 7–9-month-old infants and 31 18–30-year-old adults. The average age at the initial testing session was 32.9 wk [standard deviation (SD)=4.1 wk] for infants and 21.4 yr (SD=1.7 yr) for adults. Listener selection criteria included: (1) No risk factors for hearing loss as assessed by parent or self-report, (2) no more than two episodes of otitis media, (3) not under treatment for otitis media within the prior week, and (4) healthy on the test date. In addition, screening tympanometry was performed on every subject at each session. Peak admittance of at least 0.2 mmhos at a pressure between −200 and 50 daPa was required to pass the screening. No adult listener had more than two years of musical training nor had any listener participated in previous psychoacoustic studies. An additional seven infants and one adult were tested but were excluded from data analysis, because they did not meet the training or testing criteria.

B. Stimuli and equipment

For all conditions, the signal was a 1000-Hz pure tone. The duration of the signal was 300-ms, including a 10-ms rise/fall time. There were three types of maskers: Random-frequency two-tone, fixed-frequency two-tone, and broadband noise (300–3000 Hz). Both the random and fixed two-tone maskers consisted of two simultaneous components. The rationale for using a two-component masker is that this condition tends to produce lower thresholds relative to 10- and 20-component maskers, but the largest range in performance across listeners (e.g., Neff and Dethlefs, 1995). Thus, maturational effects might well be expected. In addition, any masking produced by two remote-frequency masker components is unlikely to be energy-based. For the random two-tone masker, one component was drawn at random from a rectangular distribution with a range of 300–920 Hz. The second component was drawn at random from a rectangular distribution with a range of 1080–3000 Hz. The "protected region" of 920 to 1080 Hz around the signal frequency was designed to reduce the possible contribution of peripheral (energy-based) masking within a presumed auditory filter centered at the signal. For the fixed two-tone masker, the two components were 581-Hz and 2920-Hz pure tones.[1] The general properties (e.g., two flanking tones, stimulus level) of the fixed-frequency masker were designed to match those of the random-frequency masker, but the frequency content of the masker remained unchanged throughout testing. The rationale for including a fixed two-tone masker condition is that it provides a comparison condition of minimal acoustic variability. In contrast to the minimal-variability broadband noise condition, the fixed two-tone condition is expected to produce little or no energy-based masking. Maskers were repeated at 300-ms intervals throughout the entire testing session. The signal, when present, was gated synchronously with one presentation of the masker. All maskers were presented at a total power of 60 dB SPL. A schematic representation of the stimuli is presented in Fig. 1.

Stimuli were generated digitally and then low-pass filtered at 4000 Hz using Tucker-Davis Technologies (TDT) system III programmable hardware (RP2). The signal and masker were attenuated separately using two programmable attenuators (TDT PA5), mixed, and fed to a headphone buffer (TDT HB7). The experiment was controlled by a computer using custom software. Testing was conducted in a double-walled sound-attenuating room (IAC). All stimuli were presented to the listener's right ear through an ER1 insert earphone. When necessary, the foam tip of the insert earphone was trimmed to fit the ear canal.

C. Procedure

Listeners were randomly assigned to one of the three masker conditions. Adults were tested in a single visit to the laboratory. Infants were tested in three separate visits occurring within a 2-wk period. Visits for both age groups were approximately 45 min in length. For listeners who completed testing quickly enough, data were collected in a second masker condition. The assignment of the second masker condition was also random.

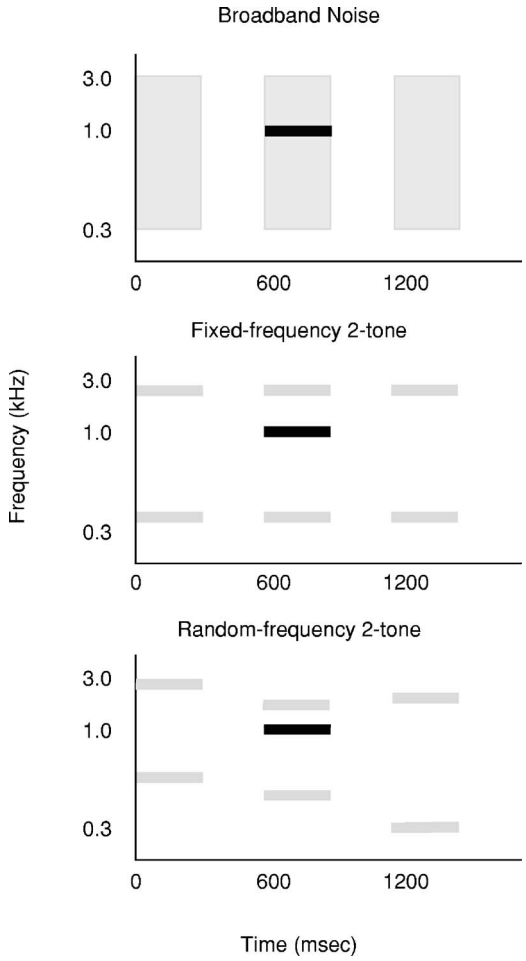


FIG. 1. Schematic illustrations of stimuli in the form of a spectrogram. Three repetitions of the continuously repeating masker (light shading) are illustrated, with the signal (bold) added to the second masker repetition. The signal, when present, was played synchronously with one repeat of the masker.

An observer-based procedure was used to assess infants' performance (Werner, 1995), and the test procedure for adults was kept as similar to that for infants as possible. Infants were tested sitting on their parent's lap. An assistant sat to the left of the parent and infant and manipulated quiet toys in order to keep the infant facing toward the midline. To prevent the assistant and the parent from hearing the tones and influencing the infant's response, both wore circumaural headphones that delivered masking sounds. Adult listeners were tested alone in the booth. To the listener's right were two mechanical toys with lights in a dark Plexiglas box. An observer sat outside of the booth and initiated trials when the listener was quiet and facing midline.

Trials were either "signals" in which the 1000-Hz tone was presented simultaneously with one burst of the repeating masker, or "no signals" in which only the masker was presented. The observer did not know which type of trial occurred and was required to decide on the trial type based on the listener's behavior within 4 s of trial onset. The listener was provided with reinforcement if the observer correctly identified a signal trial. Reinforcement was the activation and illumination of a mechanical toy. For infant listeners, any response that reliably distinguished between signals and no-signals could be used by the observer; these typically included head turns or eye movements toward the reinforcer or changes in activity. Adult listeners were instructed to raise a hand when they heard "the sound that makes the toy come on", and the observer recorded those responses that occurred during trials. Feedback was provided to the observer after every trial.

Sessions consisted of two training phases and one testing phase. In both training phases, the signal was presented at a level that was expected to be clearly audible, depending on the masker condition and based on previous studies and pilot work. The purpose of the first training phase was to demonstrate the relation between the signal and the mechanical reinforcer. Thus, the probability of a signal was 0.80 and the listener was reinforced after each signal trial, regardless of the observer's response. The first training phase ended when the observer correctly responded to four of five consecutive trials, including at least one no-signal trial. The purpose of the second training phase was to demonstrate to the listener that he needed to respond to signal trials in order to activate the mechanical toy. The probability of a signal was 0.50, and reinforcement was only provided if the observer correctly identified a signal trial. The second training phase ended when the observer achieved a hit rate of 0.80 and a false alarm rate of no greater than 0.20. Infants required an average of 38.2 (SD=13.5) trials to complete both training phases. In comparison, adults required an average of 17.8 (SD=7.3) trials to complete both training phases. There were no significant differences in the number of trials required to complete training across condition in either age group.

During testing, the probability of a signal trial was 0.75 and the probability of a no-signal trial was 0.125. Probe trials, on which the level of the target tone was the training level, were presented with a probability of 0.125. Detection thresholds were estimated adaptively using a one-up two-down algorithm (Levitt, 1971). Only signal trials affected the direction of the adaptive track and step size varied during a run following PEST rules (Taylor and Creelman, 1967). Based on pilot data, the starting level for the signal was approximately 10 dB above the expected threshold value. The initial step size was 6 dB. Testing continued until eight reversals were obtained and threshold was calculated as the average of the last six reversals. Thresholds were only accepted if the probe response rate was greater than 0.60 and the false alarm response rate was lower than 0.40. Ideally, a bias-free measure should be used to estimate sensitivity. However, informational masking data have not been collected previously in infants, and it is difficult and time consuming to obtain a psychometric function from an infant. Thus, we chose to estimate infants' sensitivity using an adaptive procedure in order to characterize the general effects of masker-frequency variability during infancy in this initial study. Limiting the acceptable false alarm rate was expected to provide some control of response bias. Additional analyses addressing this issue are described in the results.

The same experimental phases and psychophysical procedures were used to test infants and adults. Thus, from the listener's perspective, the testing conditions were the same for infants and adults: Trials were not defined and occurred

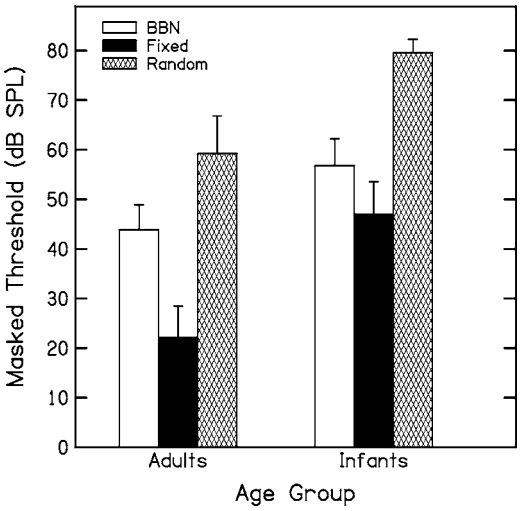


FIG. 2. Mean thresholds (±1 SD) are shown as a function of masker condition for adults and infants. The open, filled, and shaded bars represent the broadband noise, fixed two-tone, and random two-tone conditions, respectively.

at irregular intervals, and responses to signals produced activation of the mechanical toy, while responses at other times did not. It is important to note that the timing of signal presentations is variable in this procedure for all conditions. Although the introduction of temporal variability is appealing in the sense that it is probably closer to real world conditions, the effects of temporal and frequency variability might interact in such a way that the effects of frequency variability observed may not be generalizable to conditions of low temporal variability, and thus may limit the comparability of the results to those of previous studies.

Listeners were initially assigned to one masker condition. Thresholds were obtained from 13 infants in the broadband noise masker condition, 9 infants in the fixed two-tone masker condition, and 11 infants in the random two-tone masker condition. Thresholds were obtained from 11 adults in the broadband noise masker condition, 10 adults in the fixed two-tone masker condition, and 10 adults in the random 2-tone masker condition. Given the potential for large individual differences in performance, we tested listeners in two masker conditions whenever possible. A total of 17 infants and 24 adults were tested in two masker conditions. Six infants and seven adults were tested in both the broadband noise and fixed two-tone masker conditions. Four infants and seven adults were tested in both the broadband noise and random two-tone masker conditions. Seven infants and ten adults were tested in both the fixed two-tone and random two-tone masker conditions.

III. RESULTS

Two sets of analyses were undertaken, a between-subjects analysis based on the data of 64 listeners and a within-subjects analysis based on the data of 41 listeners.

A. Between-subjects analysis

Figure 2 shows the average threshold estimates in dB SPL for each age and masker condition. For listeners providing data points in two conditions, only the initial threshold was included in these averages. As expected, the average infant threshold was higher than the average adult threshold in all conditions. However, the threshold changed with condition in a similar way for the two age groups. The average threshold within each age group was lowest in the fixed two-tone condition, intermediate in the broadband noise condition, and highest in the random two-tone condition. One apparent difference between the age groups is that while the adults' threshold in the fixed two-tone condition is considerably lower than their threshold in both of the other conditions, infants' threshold in the fixed two-tone condition is not much lower than their threshold in the broadband noise condition. The average adult threshold in the fixed two-tone condition is roughly 22 dB lower than the average adult threshold in the broadband noise condition and 37 dB lower than in the random two-tone condition. The average infant threshold in the fixed two-tone condition is only 10 dB lower than the average infant threshold in the broadband noise condition and 33 dB lower than in the random two-tone condition. It should be noted that the adults' threshold in the fixed-frequency two-tone condition, about 22 dB SPL, is higher than the expected threshold for a 1000-Hz tone in quiet.

An analysis of variance (*ANOVA*) confirmed the trends observed in Fig. 2. All of the effects in the Condition X Age analysis were significant: Condition $[F(2,58)=185.98, p<0.001]$, Age $[F(1,58)=180.48, p<0.001]$, and Condition X Age $[F(2,58)=5.962, p=0.004]$. The significant Condition X Age interaction suggests that the differences between conditions are not the same in the two age groups. To explore the nature of the interaction, a one-way ANOVA was performed within each age group. This analysis revealed a significant main effect of condition for adults $[F(2,28)=86.02; p<0.001]$ and for infants $[F(2,30)=112.96; p<0.001]$. *Post hoc* testing (Scheffe, using a criterion of $p<0.05$) indicated thresholds were lower in the fixed two-tone than in the noise condition, thresholds were lower in the noise condition than in the random two-tone condition, and thresholds were lower in the fixed two-tone than in the random two-tone condition, for both infants and adults. A two-way ANOVA (Condition X Age) was then performed on each pair of conditions to identify the source of the Condition X Age interaction in the omnibus analysis. A significant Condition X Age interaction was observed for the broadband noise and random two-tone conditions $[F(1,41)=5.29, p=0.027]$, indicating a larger difference between conditions for infants than for adults. The comparison between broadband and random-frequency masking is the one made in previous studies of children, and the relatively larger increase in average threshold in the random-frequency condition among infants is consistent with the results of those studies. In contrast, there was no Condition X Age interaction for the fixed- and random-frequency two-tone conditions $[F(1,36)=1.39, p=0.25]$, providing no evidence that adding masker-frequency variability to a two-tone masker affected infants' performance any more than it affected adults'. A significant Condition X Age interaction was found for the broadband noise and fixed two-tone conditions $[F(1,39)=11.15, p=0.002]$, indicating that the difference between the fixed two-tone and the broadband noise

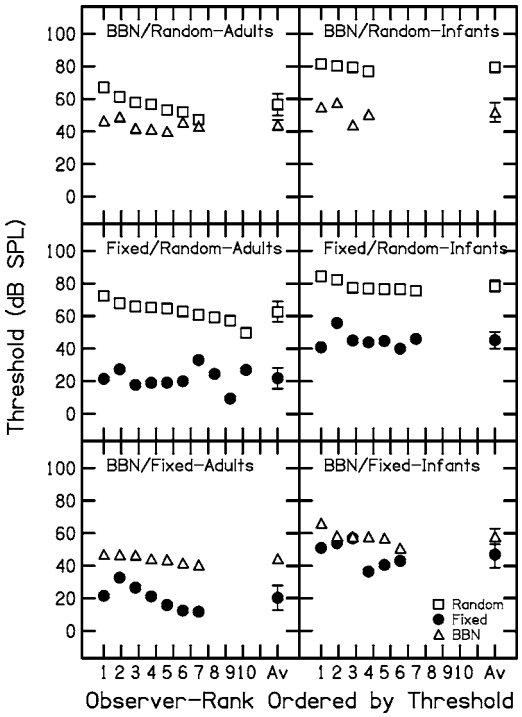


FIG. 3. Thresholds are plotted for individual adults and infants and for the average across listeners (Av). The error bars indicate ±1 SD. Panels are for listeners tested in the broadband noise and random two-tone conditions (top row), the fixed two-tone and random two-tone conditions (middle row), and the broadband noise and fixed two-tone conditions (bottom row). Adult thresholds are shown in the left panel and infant thresholds are shown in the right panel. The random two-tone, fixed two-tone, and broadband noise conditions are indicated using squares, filled circles, and triangles, respectively.

conditions was smaller for infants than for adults. This result suggests that infants are more susceptible to masking in the fixed two-tone condition than adults.

B. Within-subjects analysis

Because of large individual differences in informational masking (and among infants in all kinds of masking), it is important to determine whether the effects seen in the group comparisons are evident in the data of individual listeners. Figure 3 shows threshold estimates for individual listeners tested in two masker conditions, with adult data presented in the left panel and infant data presented in the right panel. The squares, filled circles, and triangles are for the random two-tone, fixed two-tone, and broadband noise masker conditions, respectively. The individual listeners are rank ordered from highest to lowest threshold in each panel. The average thresholds are plotted at the right of each panel. Note that the average thresholds plotted in Fig. 3 are quite similar to those of the larger groups (Fig. 2). Three-way repeated-measures *ANOVA* (Condition X Age X Order) of threshold for each condition pair showed that the same effects were significant as in the between-subjects comparison. The effect of Order was not significant in any of these analyses (all $p > 0.40$).

In most conditions, the effects observed in the individual data are similar to those in the group data. For the broadband-noise/random two-tone comparison, all four infants demonstrated larger differences in threshold than any of the adults. For the fixed two-tone/random two-tone comparison, the difference in threshold for the infants ranged from 26–44 dB compared with a range of 23–51 dB for the adults. For the broadband-noise/fixed two-tone comparison, several adults have broadband/fixed threshold differences as large as 30 dB, and no adults have threshold differences less than 15 dB. Interestingly, three of six infants look similar to adults in that their thresholds are 15–20 dB higher for the broadband-noise relative to the fixed two-tone condition. The remaining three infants show little or no difference in threshold between the broadband-noise and fixed two-tone conditions. Thus, some infants demonstrate considerable masking in the fixed two-tone condition, and some infants demonstrate about as much masking as many adults. These individual differences could not have been identified from the mean thresholds.

Note that considerable between-subjects variability was evident for both infants and adults. Infants and adults were about equally variable in the broadband noise and fixed two-tone conditions, and there was no apparent difference in variability between those two conditions. However, in the random two-tone masker condition, infants were less variable, while adults were more variable than infants and more variable than in the other masker conditions. The same trends are evident in the group data (Fig. 2).

C. Ceiling effects

The average infant threshold in the random two-tone condition was 79.6 dB SPL, with individual thresholds ranging from 75.5 to 84.4 dB SPL. Standard errors of 1.5, 2.2, and 0.8 dB were obtained for infants in the broadband noise, fixed two-tone, and random two-tone conditions, respectively. In contrast, standard errors of 1.5, 2.0, and 2.4 dB were obtained for adults in the broadband noise, fixed two-tone, and random two-tone conditions, respectively. Thus, the standard errors of infants and adults were comparable in all but the random two-tone condition. Because the range of thresholds obtained from infants was smaller than the range of thresholds obtained from adults in this condition, and because the maximum allowable level of the signal was 91 dB SPL, we examined the possibility that infant thresholds in the random two-tone condition were constrained by ceiling effects.

In fact, there were few indications in the data that many infants' thresholds in the random two-tone condition were much higher than those included in the analysis. Infants did not appear to have more difficulty completing training in the random two-tone condition relative to the other two conditions. The average number of trials required to complete training was 39, 33, and 43 for the broadband noise, fixed two-tone, and random two-tone conditions, respectively, and these were not significantly different. For infants assigned to the random two-tone masker condition, 13/15 successfully completed the two training phases and proceeded to testing. In comparison, 14/15 infants assigned to the broadband noise masker condition and 9/10 assigned to the fixed two-tone masker condition successfully completed the two training phases. The proportions of infants meeting training criteria

TABLE I. The average d's across listeners are compared for the between-subjects and within-subjects data. Standard deviations of the mean and range of d's are also shown for each set of data.

Between-subjects							
Infants				Adults			
Condition	d′	S.D.	Range	Condition	d′	S.D.	Range
BBN	0.92	0.39	0.39-1.61	BBN	1.74	0.36	0.81-2.18
Fixed	0.88	0.53	0.30-1.90	Fixed	1.91	0.15	1.68-2.14
Random	1.21	0.56	0.44-2.12	Random	1.81	0.36	1.22-2.46
Within-subjects							
BBN	1.06	0.54	0.40-1.82	BBN	1.75	0.22	1.52-2.09
Fixed	0.81	0.24	0.51-1.23	Fixed	1.82	0.24	1.39-2.06
BBN	1.20	0.90	0.56-1.61	BBN	1.79	0.39	0.92-2.14
Random	1.22	0.52	0.87-2.12	Random	1.94	0.29	1.64-2.46
Fixed	0.84	0.32	0.48-1.44	Fixed	1.87	0.18	1.65-2.14
Random	0.94	0.46	0.44-1.64	Random	1.68	0.31	1.22-2.15

who also completed testing were 11/13 for the random two-tone masker condition, 13/14 for the broadband noise masker condition, and 9/9 for the fixed two-tone masker condition. The average range of the reversals used to calculate infants' thresholds was similar in the three conditions: 4.6, 4.1, and 4.0 dB for the broadband noise, fixed two-tone, and random two-tone conditions, respectively. Only one infant's data were excluded from analysis because the next stimulus level called for by the adaptive algorithm exceeded 91 dB SPL. That infant was in the random two-tone condition, but if many infants' thresholds in the random two-tone condition actually were higher than 91 dB SPL, we would have expected more than one infant to be excluded for this reason. In summary, there is little evidence of a ceiling effect in the random two-tone masker condition for infants.

D. Response bias

In the current study, listeners were tested using a yes-no procedure in which only responses on signal trials affected the value of threshold. Thus, while the threshold should consistently indicate the stimulus level yielding a 71% hit rate, it may not indicate equivalent sensitivity if there were differences between ages or conditions in response bias. Furthermore, the average false alarm rate in the test phase of the procedure was 0.26 for infants, while it was 0.03 for adults.

To examine this issue, d' near threshold was calculated for each listener and experimental condition. Because threshold was estimated adaptively, the d' estimates were based only on trials after the adaptive track had stabilized (± 1 SD of the reversal points for each listener). Because the level is fairly constant for these trials, the expectation is that the listener's sensitivity and response bias are also fairly constant. A value of 0.5 was added to all cells in all conditions due to the high number of listeners having a false alarm rate of 0 (Snodgrass and Corwin, 1988).

Table I provides average estimates of d' for the group data and for listeners tested in two masker conditions. For all conditions, infants produce a d' of about 1, a value roughly

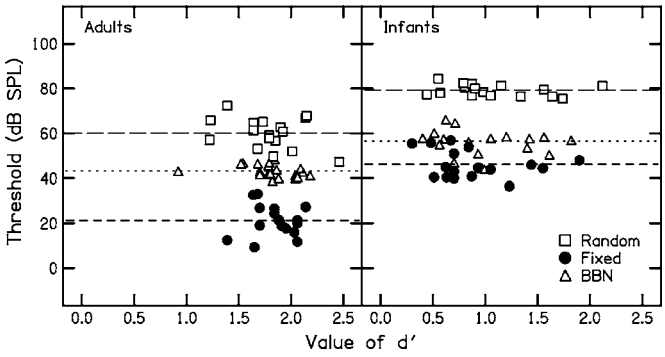


FIG. 4. d' values are plotted against threshold for individual listeners. Panels are for adults (left panel) and infants (right panel). Both the between-subjects and within-subjects data are included. The random two-tone, fixed two-tone, and broadband noise conditions are indicated using squares, filled circles, and triangles, respectively. The dashed horizontal lines indicate group mean thresholds for the random two-tone, broadband noise, and fixed two-tone conditions, from top to bottom in each panel.

consistent with 70% correct responses for an unbiased observer. In contrast, the average d' estimates for adults, 1.7–1.9, were consistently higher than those for infants, suggesting that at the estimated threshold adults are operating at the equivalent of 80% correct for an unbiased observer. Thus, it is likely that the overall difference between infants and adults in sensitivity is greater than the age difference in thresholds would suggest. There is no evidence, however, of systematic differences in d' across condition within either age group. This conclusion is supported by Fig. 4, where individual thresholds are plotted against d' by condition. While d' is variable, the range of d' estimates is similar across condition within each age group. In contrast, a clear relation between condition and threshold is evident for both the infant and adult data, and threshold differences are similar regardless of d'. Thus, the between-condition differences within age groups would not be greatly affected by differences in response bias. The trends in d' observed in Fig. 4 were confirmed in a Condition X Age ANOVA. The main effect of Age was significant $[F(2,58)=66.15, p<0.0001]$, indicating that d' estimates differed between infants and adults. Unlike the ANOVA performed on threshold, however, neither the main effect of Condition nor the Age X Condition interaction was significant (both $ps>0.4$).

IV. DISCUSSION

The major question addressed by this study was how masker-frequency variability affects infants' detection of a fixed-frequency pure tone. The results indicate that infants have higher masked thresholds than adults under conditions of masker-frequency variability. Further, the difference between infants' threshold in the presence of a broadband noise masker and their threshold in the presence of a random-frequency two-tone masker is greater than the difference seen in adults. These results were evident in between-group comparisons, as well as in the data for individual infants and adults tested in both conditions. This result matches one prediction offered in the introduction that informational masking effects might be very large during infancy, as they appear to be during the preschool period. However, the substantial

between-subjects differences that characterize preschoolers' performance in the presence of masker-frequency variability were not evident in the infant data. Further, there was no overlap between the range of infants' thresholds and the range of adults' thresholds in the random two-tone condition.

In addition to the elevated masked thresholds observed under conditions of masker-frequency variability, both infants and adults demonstrated masking in the presence of a fixed-frequency two-tone masker. The amount of masking in this condition appeared, on average, higher for infants than adults. For some individual infants tested in both the fixed two-tone and broadband noise conditions, however, the amount of masking seen in the fixed two-tone condition, relative to the broadband noise condition, was within the range observed in adults.

Finally, thresholds in the random two-tone masker condition were higher than those in the fixed two-tone masker condition, for both infants and adults. The difference between these conditions should reflect the effect of masker-frequency variability in the presumed absence of energetic masking. Defining the effect in this way, there was no significant difference between infants and adults in the effect of masker-frequency variability. The random-fixed difference for all seven infants tested in both conditions fell well within the range of differences observed in adults. This result does not match well with any of our *a priori* predictions. The introduction of masker-frequency variability, in the presumed absence of energetic masking, was clearly detrimental to infants' performance, but there was no evidence that masker-frequency variability was more detrimental to infants' performance than to adults'.

A. Masking in conditions of minimal acoustic variability

Thresholds obtained here in the presence of intermittent broadband noise are comparable to infant and adult data reported in the literature for continuous broadband noise maskers using the same observer-based procedure (Bargones *et al.*, 1995). In addition, thresholds in noise for adults in the current study are similar to those expected for trained adults using more conventional two-interval forced-choice (2IFC) procedures with broadband noise maskers gated during observation intervals (e.g., Neff and Dethlefs, 1995).

The finding that infants had more difficulty than adults detecting a 1000-Hz tone in the presence of a broadband noise is consistent with previous investigations. Several observations suggest that the mechanisms responsible for infants' immature detection in noise are similar to the mechanisms responsible for informational masking in adults. There is evidence that infants monitor the output of many auditory filters during tone-in-noise detection tasks, whereas adults place little weight on filters remote in frequency from the tone (Werner and Boike, 2001; Bargones *et al.*, 1995; Bargones and Werner, 1994). As a result, thresholds are elevated because the signal-to-noise ratio is effectively decreased.

The infant-adult difference in threshold in broadband noise, however, is roughly 13 dB here, compared to a difference of 8 dB reported by Bargones *et al.* (1995). Because Bargones *et al.* estimated thresholds from a bias-free measure of sensitivity, and because adults were more conservative listeners than infants in the current study, we would expect that the infant-adult difference might be smaller in the current study, not larger. The basis for the discrepancy between studies is not clear, but the mode of masker presentation may have had a greater effect on infants' pure-tone detection than adults (Werner *et al.*, 2006). Bargones *et al.* used a continuous noise masker while we used a masker that was presented intermittently and that was gated on and off with the signal tone.

The fixed-frequency two-tone masking condition was included in the present experiment as a second comparison condition with minimal stimulus variability but with little or no possibility of energetic masking. We expected thresholds would be similar to those previously reported for a 1000-Hz signal in quiet, at least for the adults. This prediction was not met. The average adult threshold obtained in the fixed two-tone condition was 22.1 dB SPL (SD=6.4). In contrast, Werner *et al.* (2004) reported an average adult threshold of 10.6 dB SPL (SD=3.8) for a 1000-Hz pure tone in quiet using the same software, hardware, procedure, and transducer used here. Thus, the average adult threshold was approximately 11 dB higher in the fixed-frequency two-tone condition than the expected threshold in quiet. We did not, however, obtain thresholds in quiet in the current study. Because threshold in quiet can be variable across subjects, we tested an additional three adults in quiet as well as in the fixed-frequency two-tone masker condition. Condition order was randomized. Individual thresholds obtained for the three subjects in the fixed two-tone condition were 28.1, 19.3, and 14 dB SPL (mean=20.5 dB SPL). Thresholds in quiet for the same three subjects were 21.8, 1.6, and 11.4 dB SPL, respectively (mean=11.6 dB SPL). The average threshold obtained in the fixed-frequency two-tone condition for the three subjects was elevated by 10.4 dB relative to threshold in quiet. This estimate is similar to the 11 dB difference estimated from the group average adult data. Notice that the amount of masking in these three listeners ranged from less than 3 dB to nearly 18 dB, as is typical of informational masking. In a recent study, we confirmed that while many adult listeners demonstrate 3–5 dB of masking in the fixed-frequency two-tone masker condition, some exhibit 10–15 dB of masking (Werner and Leibold, 2005).

The group data suggest that both infants' and adults' detection was affected by the presence of a fixed-frequency two-tone masker remote from the signal frequency. However, the average infant's poor performance in this condition was striking. Average threshold was 22 dB higher in the broadband noise compared to the fixed two-tone condition for adults, but only 10 dB higher for infants. In addition, the average infant threshold measured in the presence of the fixed two-tone masker was elevated by approximately 18 dB relative to the average threshold in quiet estimated by Werner *et al.* (2004) with the same software, hardware, procedure, and transducer used here. This is similar to the observation of Werner and Bargones (1991) that infants' detection thresholds were elevated by about 10 dB when a

remote-frequency band of noise was simultaneously presented with the signal. It would appear, however, that even more masking is observed when the remote-frequency masker is a tonal complex rather than noise.

For listeners tested in both the fixed two-tone and broadband noise masker conditions, the average effects are consistent with the effects observed in the group data. An examination of individual thresholds, however, revealed substantial variability in performance for infants tested in both conditions. Although one-half of the infants had higher thresholds in the broadband noise relative to the fixed two-tone conditions, thresholds for the remaining infants were similar in the fixed two-tone and broadband noise conditions. In contrast, all of the adult listeners tested in both the broadband noise and fixed two-tone conditions had lower thresholds in the fixed two-tone relative to the broadband noise condition. In other words, when energetic masking was reduced, all adults and some infants exhibited a "release" from masking, but some infants exhibited little or no improvement in threshold. Similarly, Werner and Bargones (1991) observed 0–30 dB of remote-frequency masking in infants, but only 0–7 dB in adults. The range of performance for infants in the fixed two-tone condition here is also similar to that reported in the literature for preschoolers under conditions of masker-frequency variability (e.g., Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003). Note, however, that the infants and adults whose fixed two-tone thresholds were 15–20 dB lower than their broadband noise thresholds were probably experiencing some masking in the fixed two-tone masker condition.

There is no apparent peripheral explanation for the masking observed in the presence of the fixed two-tone masker. Several other studies have also reported that thresholds for some adults in conditions with reduced masker-frequency variability (e.g., masker frequency fixed across intervals on each trial or across a block of trials) remain elevated above what would be predicted from energy-based detection (e.g., Alexander and Lutfi, 2004; Neff and Callaghan, 1988, Neff and Dethlefs, 1995; Richards and Neff, 2004). In the current study, the fixed-frequency masker consisted of pure tones at 581 and 2920 Hz, remote from the frequency region of the signal (Moore and Glasberg, 1996) and unlikely to produce substantial energetic masking. Further, at least two lines of evidence argue that the masking produced by the fixed frequency tonal masker should be considered informational masking. First, both the current and previous studies showed large individual differences in performance in fixed-frequency tonal masking conditions, as they do in random-frequency masker conditions. For example, Alexander and Lutfi (2004) reported thresholds ranging from 14.5 to 52.3 dB SPL in a group of normal-hearing adults asked to detect a 2000-Hz pure tone in the presence of a fixed-frequency ten-component masker used to estimate energetic masking (their $p=1.0$ condition). In contrast, listeners typically have similar and stable thresholds with broadband noise maskers. Second, Leibold *et al.* (2005) recently showed that temporal manipulations of the masker that reduced informational masking under conditions of masker-frequency variability also reduced adults' masking by fixed-remote-frequency maskers in a 2IFC task.

Thus, it appears that both infants and adults are susceptible to informational masking in the absence of masker-frequency variability. Although the mechanisms responsible for informational masking in this condition are unclear, Durlach *et al.* (2003a) have argued that stimulus uncertainty is not required to produce informational masking. Instead, it has been suggested that informational masking is largely determined by how much the signal and masker "sound alike" (Durlach *et al.*, 2003a; 2003b). Alexander and Lutfi (2004) have suggested that the poor performance of some listeners in the presence of a fixed-frequency ten-tone masker may reflect a perceptual grouping effect similar to that described by Kidd *et al.* (1994; 2001; 2002) in variable masker conditions. Alternatively, uncertainty might be required to produce informational masking as the CoRE model suggests, but the source of the uncertainty might not be in the acoustic stimulus. For example, uncertainty might be created by variability in the internal representation of the masker, or listeners might have an imperfect memory for the masker frequency. That infants are generally more susceptible than adults to informational masking in the absence of stimulus variability could be accounted for by immaturities in any or all of these mechanisms.

B. Masking in conditions of masker-frequency variability

Consistent with previous studies, masker-frequency variability had a detrimental effect on adults' detection performance (e.g., Neff and Green, 1987). However, the amount of informational masking measured for adults in the current study was larger than in previous studies of informational masking. For example, the average adult threshold ranged from 47.3 to 72.4 dB SPL in the current random two-tone condition whereas Neff and Dethlefs (1995) reported a threshold range of 5 to 64 dB SPL from their trained adult subjects using two-tone, random-frequency maskers. This could be due to the greater temporal variability in our experiment (discussed in Sec. C below).

Masker-frequency variability was clearly detrimental to infants as well as adults. Whether we judge infants to be more susceptible to masker-frequency variability than adults, however, depends on the comparison that is made. When the effect of masker-frequency variability is estimated as the elevation in threshold in the random two-tone relative to the broadband noise condition, the effect appears to be larger for infants than adults. Note that this comparison is the one used in most previous studies of informational masking, and infants' apparent increase in susceptibility to masking in the random two-tone condition is consistent with previous studies that have reported more masking for children than adults under conditions of masker-frequency variability (e.g., Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003). For example, thresholds for 7/8 preschoolers tested by Oh *et al.* (2001) in the presence of a random-frequency two-tone masker were elevated by 20 dB or more relative to thresholds for the same children in the presence of broadband noise. In contrast, none of the adults tested by Oh *et al.*

had higher thresholds in the random two-tone condition than in the broadband noise condition.

A different conclusion might be reached if the effect of varying the frequency content of the masker is estimated as the elevation in threshold in the random two-tone relative to the fixed two-tone masker condition. The increase in threshold associated with masker-frequency variability is similar for infants and adults in this comparison, providing no evidence that infants are more susceptible to masker-frequency variability than adults. Although only seven infants and ten adults were tested in both conditions, these differences are consistent across both infant and adult listeners. This finding raises an important question: What is the appropriate baseline condition for estimating the effect of masker-frequency variability? Effects of masker-frequency variability are often reported as the difference in threshold in the presence of a random-frequency multitone complex compared to that in broadband noise or to that in quiet. However, a logical reference for measuring effects of spectral variability is a masker that is matched to the random frequency masker in all respects except spectral variability. A single multitone masker sample approaches that condition (as in the minimal-uncertainty procedures discussed by Watson *et al.*, 1976). We are currently examining whether the effect of masker frequency variability is greater in children than in adults, when a fixed-frequency masker condition is the reference condition.

C. Limitations

The number of subjects tested in two masker conditions in this study was small and limited our ability to interpret these data. However, it is encouraging that the effects evident in between-group comparisons were also generally evident in the limited within-subject comparisons. The results indicate the questions that might be addressed in future studies employing a within-subjects design.

Infant-adult threshold differences are determined by age differences in sensitivity, in response bias and in the accuracy with which listener responses can be recorded. A particular limitation of the current study is the apparent difference in response bias between infants and adults. The age difference in threshold reported here likely underestimates the infant-adult difference in sensitivity. Psychometric functions have been reported for infants' and adults' detection of a 1000-Hz tone in continuous broadband noise using the current procedures (Bargones *et al.*, 1995). If the psychometric functions for infants and adults here are consistent with those of Bargones *et al.*, then the adult threshold would be 2 dB lower in each condition than reported here. However, given likely differences between the slope of the psychometric function for detection in a random-frequency masker and that for detection in a broadband noise masker (Durlach *et al.*, 2005; Lutfi *et al.*, 2003a) and the lack of psychometric function data for infants in any multitone masker condition, it is difficult to estimate the extent to which infants' thresholds differ from adults' in the two-tone masking conditions. Fortunately, response bias did not vary across masking conditions, making within-age-group comparisons meaningful.

A further limitation of these data is the possibility of an interaction between the spectral characteristics of the stimuli and the temporal variability inherent in our "infant" procedure. The observer-based procedure differs from most psychoacoustic procedures in the degree of temporal variability; that is, observations intervals are typically defined in psychoacoustic experiments, while in the observer-based procedure they are not. It is possible that this difference exacerbated the effects of spectral variability on masked thresholds. That thresholds for adults tested in the random two-tone condition were higher than reported in previous studies is consistent with this possibility. In a follow-up study conducted in our lab (Leibold and Werner, 2003), we attempted to separate the effects of temporal and spectral variability on adults' pure-tone detection. Five adults were tested using the same procedures described in the current study, except for the addition of a condition in which listeners initiated each trial, and a visual cue marked the masker burst with which the signal might coincide. For all listeners, performance was better when the listening interval was defined, with improvements in threshold of 10–20 dB. Because procedural differences apparently lead to only small differences in threshold for a 1000-Hz tone masked by broadband noise, these findings suggest a potentially interesting interaction between temporal and frequency variability, at least for adult listeners.

A third potential limitation of these data is that infants' thresholds in the random two-tone condition may have been constrained by a ceiling effect. The bulk of evidence does not indicate a ceiling effect. If this effect is present, however, infants' thresholds in the random two-tone condition would be higher than observed here, and the effects of spectral variability on infants' pure-tone detection greater than the current data would suggest.

V. SUMMARY AND CONCLUSIONS

Thresholds of both infants and adults were substantially elevated in the presence of a random-frequency masker. Whether there are developmental changes in the effect of spectral variability is a matter of interpretation and will only be settled by additional research. Infants, as predicted on the basis of previous studies, exhibited substantial masking in the presence of a fixed-frequency tonal masker. Adults also exhibited masking in the fixed-frequency tonal masker condition, although on average, not to the extent observed in infants. An argument can be made that this masking in the absence of spectral variability is informational masking. Finally, several observations suggest that the effects of spectral variability on detection interact with those of temporal variability. Understanding the development of listening abilities is challenging, but it may shed light on the mechanisms responsible for informational masking.

ACKNOWLEDGMENTS

This work was supported by funding from the NIH (R01 DC00396, P30 DC04661, F31 DC06122, T32 DC00013). The authors are grateful to Donna Neff and Walt Jesteadt for useful discussions and comments.

[1]Given the wide frequency separation between the two masker components, masking by combination tones is not expected.

Alexander, J. M., and Lutfi, R. A. (**2004**). "Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights," J. Acoust. Soc. Am. **116**, 2234–2247.

Allen, P. A., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

Bargones, J. Y., and Werner, L. A. (**1994**). "Adults listen selectively; infants do not," Psychol. Sci. **5**, 170–174.

Bargones, J. Y., Werner, L. A., and Marean, G. C. (**1995**). "Infant psychometric functions for detection: Mechanisms of immature sensitivity," J. Acoust. Soc. Am. **98**, 99–111.

Dai, H., Scharf, B., and Buus, S. (**1991**). "Effective attenuation of signals in noise under focused attention," J. Acoust. Soc. Am. **89**, 2837–2842.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G., Jr. (**2005**). "Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers," J. Acoust. Soc. Am. **118**, 2482–2497.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003b**). "Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–56.

Greenberg, G. Z., and Larkin, W. D. (**1968**). "Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: the probe-signal method," J. Acoust. Soc. Am. **44**, 1513–123.

Hall, J. W. III, Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.

Kidd, G., Jr., 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.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2001**). "Informational masking in listeners with sensorineural hearing loss," J. Assoc. Res. Otolaryngol. **3**, 107–119.

Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (**2002**). "Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, G., Jr., Mason, C. R., 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.

Leibold, L. J., Neff, D. L., and Jesteadt, W. (**2005**). "Effects of reduced spectral uncertainty and masker fringes with multitonal maskers," J. Acoust. Soc. Am. **118**, 1893.

Leibold, L. J., and Werner, L. A. (**2003**). "Infants' detection in the presence of masker uncertainty," J. Acoust. Soc. Am. **113**, 2208.

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

Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758.

Lutfi, R. A., Kistler, D. J., Callaghan, M. R., and Wightman, F. L. (**2003a**). "Psychometric functions for informational masking," J. Acoust. Soc. Am. **114**, 3273–3282.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003b**). "One factor underlies individual differences in auditory informational masking within and across age groups," Percept. Psychophys. **65**, 396–406.

Moore, B. C. J., and Glasberg, B. R. (**1996**). "A revision of Zwicker's loudness model," Acta Acust. **82**, 335–345.

Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.

Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.

Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.

Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.

Snodgrass, J. G., and Corwin, J. (**1988**). "Pragmatics of measuring recognition memory: Applications to dementia and amnesia," J. Exp. Psychol. Gen. **117**, 34–50.

Taylor, M. M., and Creelman, C. D. (**1967**). "PEST: Efficient estimates on probability functions," J. Acoust. Soc. Am. **41**, 782–787.

Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbassat, C. A. (**1975**). "Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals," J. Acoust. Soc. Am. **57**, 1175–1185.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). "Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty," J. Acoust. Soc. Am. **60**, 1176–1186.

Werner, L. A. (**1995**). "Observer-based approaches to human infant psychoacoustics," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhauser Verlag, Boston), pp. 135–146.

Werner, L. A., and Bargones, J. Y. (**1991**). "Sources of auditory masking in infants: Distraction effects," Percept. Psychophys. **50**, 405–412.

Werner, L. A., and Boike, K. (**2001**). "Infants' sensitivity to broadband noise," J. Acoust. Soc. Am. **109**, 2103–2111.

Werner, L. A., Jeon, H., and Kopyar, B. (**2006**). "Update on infants' increment detection in tones and noise," *Proceedings of the 29th MidWinter Meeting of the Association for Research in Otolaryngology*, 257.

Werner, L. A., and Leibold, L. J. (**2005**). "Thresholds for a tone masked by constant, remote-frequency maskers," J. Acoust. Soc. Am. **118**, 1893.

Werner, L. A., Olsen, S. E., and Holmer, N. M. (**2004**). "The contribution of middle ear function to infants' pure-tone sensitivity," *Abstracts of the 27th MidWinter Research Meeting*, Association for Research in Otolaryngology, p. 931.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.

Wightman, F. L., and Kistler, D. J. (**2005**). "Informational masking of speech in children: Effects of ipsilateral and contralateral distractors," J. Acoust. Soc. Am. **118**, 3164–3176.

Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects

Matthew J. Goupell[a] and William M. Hartmann
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 2 May 2005; revised 28 March 2006; accepted 5 April 2006)

One-hundred left-right noise-pairs were generated, all with a fixed value of long-term interaural coherence, 0.9922. The noises had a center frequency of 500 Hz, a bandwidth of 14 Hz, and a duration of 500 ms. Listeners were required to discriminate between these slightly incoherent noises and diotic noises, with a coherence of 1.0. It was found that the value of interaural coherence itself was an inadequate predictor of discrimination. Instead, incoherence was much more readily detected for those noise-pairs with the largest fluctuations in interaural phase or level differences (as measured by the standard deviations). One-hundred noise-pairs with the same value of coherence, 0.9922, and geometric mean frequency of 500 Hz were also generated for bandwidths of 108 and 2394 Hz. It was found that for increasing bandwidth, fluctuations in interaural differences varied less between different noise-pairs and that detection performance varied less as well. The results suggest that incoherence detection is based on the size of interaural fluctuations and that the value of coherence itself predicts performance only in the wideband limit. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2200147]

PACS number(s): 43.66.Pn, 43.66.Ba, 43.66.Qp [AK] Pages: 3971–3986

I. INTRODUCTION

Interaural coherence is a measure of the similarity of signals in a listener's two ears. It is derived from the interaural cross-correlation function, $\gamma(\delta)$, which is a function of the interaural lag δ,

$$\gamma(\delta) = \frac{\int_0^T x_L(t)x_R(t+\delta)dt}{\sqrt{\int_0^T x_L^2(t_1)dt_1 \int_0^T x_R^2(t_2)dt_2}}, \quad (1)$$

where x_R is the signal in the right ear, and x_L is the signal in the left. The cross-correlation is bounded by $-1 \leqslant \gamma \leqslant 1$.

With respect to perception, interest normally centers on the peak of $\gamma(\delta)$. The value of δ for which the peak occurs is the relevant interaural time difference (ITD) cue for the location of the sound image. This value of δ was given a place representation in the famous binaural model by Jeffress (1948). The height of the peak is thought to determine the compactness of the image. If the sounds in the two ears are identical except for an interaural delay and/or a fixed level difference then γ assumes its maximum value of 1, and the image is expected to be maximally compact. If the height of the peak is less than 1, the image is broader or more diffuse (Barron, 1983; Blauert and Lindemann, 1986).

The height of the cross-correlation peak is known as the coherence. To be psychologically relevant, the peak must occur in a range of lag values, δ, that the binaural system can use for localization, and this requirement limits the number of peaks in the cross-correlation function for which the concept of coherence is meaningful, though the quantitative limits are somewhat uncertain. As applied in architectural acoustics, the coherence is called the IACC (interaural cross-correlation), and it refers to the height of $\gamma(\delta)$ for δ in the range $-1 \leqslant \delta \leqslant 1$ ms (Beranek, 2004). In recent years this objective architectural measure has been divided into two measures. One is the apparent source width (ASW), based on coherence within 80 ms of the onset of a sound (Barron and Marshall, 1981). The other is listener envelopment (LEV), determined by the coherence of later arriving sound, as measured after 80 ms (Bradley and Soulodre, 1995; Barron, 2004). Together, the ASW and LEV greatly influence the spatial impression of a sound in a room. Normally, the architectural measurements are made with directional microphone techniques and not with artificial heads.

Perceptual aspects of cross-correlation and interaural coherence have been studied by psychoacousticians, usually with bandpass filtered noise as a stimulus. Noise provides an abstraction of real-world sounds that is devoid of meaningful information, and it affords many opportunities for parametric variations. Using broadband noise, Pollack and Trittipoe (1959a, b) found thresholds for changes in cross-correlation as a function of the reference correlation, either 1.0 (i.e., No) or -1.0 (i.e., Nπ). They explored the effects of duration, sound level, frequency range, and interaural level difference.

Listeners are particularly sensitive to deviations from a reference correlation of 1.0. Using narrowband noise, Gabriel and Colburn (1981) found that listeners could easily distinguish between noise with a coherence of 1.0 and noise with a coherence of 0.99. They also reported the somewhat counterintuitive result that as the bandwidth of the noise increases, the just-noticeable difference (jnd) also increases. One might have expected the jnd to decrease instead given that a wider bandwidth generally offers the listener more information.

[a]Electronic mail: goupellm@msu.edu

A reference coherence of 1.0 is also of interest in connection with the masking level difference (MLD). Wilbanks and Whitmore (1967) and Koehnke *et al.* (1986) concluded that the threshold signal-to-noise ratio in the NoSπ condition is essentially determined by the ability to detect the incoherence introduced by the out-of-phase signal.

The present article is also concerned with incoherence detection starting with perfectly coherent noise as a reference. Its working hypothesis is based on the suspicion that the extreme sensitivity shown by listeners to small amounts of incoherence is not properly described by the long-term cross-correlation function. The reason is that when a small amount of incoherence is added to an otherwise perfectly coherent noise, the image of the noise acquires lateral fluctuations that are not present when the coherence is 1.0. The hypothesis continues with the observation that whereas long-term coherence is a measure that is averaged over time, the fluctuations that are hypothesized to be at the basis of coherence discrimination are dynamic. Thus, although the coherence measure may be a mathematically useful characterization of the similarity or dissimilarity of signals in the two ears, this measure may not be the most perceptually relevant characterization. Instead, it is possible that some measure that specifically considers fluctuations is better. A similar point of view with respect to the MLD was taken by Breebaart and Kohlrausch (2001) and in unpublished work by Isabelle and Colburn (2003). The rest of this article describes experiments that test this hypothesis.

II. EXPERIMENT 1: INCOHERENCE AND INTERAURAL FLUCTUATIONS

The purpose of Experiment 1 was to determine whether the height of the peak of the cross-correlation function (the coherence) adequately describes incoherence detection given a reference coherence of 1.0. The experiments employed reproducible noises, as have been used in MLD experiments by Gilkey *et al.* (1985), Isabelle and Colburn (1991), and Evilsizer *et al.* (2002). The noise stimuli, consisting of left- and right-ear channels, are referred to as "noise-pairs." In Experiment 1, the different noise-pairs all had the same value of long-term interaural coherence. If the coherence is an adequate measure of perception, then all the noise-pairs will have approximately the same detectability. That is, all pairs will be about equally distinguishable from perfectly coherent noise.

A. Stimuli

A collection of 100 noise-pairs with reproducible amplitudes and phases was created for Experiment 1. The process began with two waveforms, A and B, written as a sum of cosines in the form

$$x_A(t) = \sum_{n=1}^{N} C_n^A \cos(\omega_n t + \phi_n^A) \tag{2}$$

and

$$x_B(t) = \sum_{n=1}^{N} C_n^B \cos(\omega_n t + \phi_n^B), \tag{3}$$

where C and ϕ are the amplitudes and phases of the spectral components. The narrowband noises were generated with components having random phases over a frequency range of 490–510 Hz and with a frequency spacing of 2 Hz. Components between 495 and 505 Hz had equal amplitudes of unity. Frequencies below 495 Hz had a raised-cosine window applied to the amplitude spectrum of the form

$$C_n^A = C_n^B = \sin^2\left[\frac{\pi(f_n - 490)}{10}\right] \quad \text{for } 490 \leqslant f_n \leqslant 495 \tag{4}$$

to minimize spectral edge effects in the noises. The amplitudes of components from 505 to 510 Hz were similarly windowed. The 3-dB bandwidth of the noise was therefore 14 Hz.

For each noise in a collection of 100 reproducible noise-pairs, the B noise (x_B) was orthogonalized to the A noise (x_A) by the Gram-Schmidt orthogonalization procedure. The orthogonalized B noise is here denoted as x_B'. This was done to ensure that signals were uncorrelated and that the final value of the cross-correlation after mixing would be precise. The orthogonalization procedure is outlined by Culling *et al.* (2001).[1]

The two perfectly uncorrelated noises were then mixed, with mixing factor α, to create the final left and right (L and R) noise-pairs to be sent to the listeners,

$$\begin{aligned} x_L &= x_A, \\ x_R &= \sqrt{(1-\alpha^2)}x_A + \alpha x_B'. \end{aligned} \tag{5}$$

The coherence, as defined by Eq. (1), is

$$\rho = \sqrt{1-\alpha^2}. \tag{6}$$

Because the mixing factor used in all of the experiments was $\alpha = 0.125$, the interaural coherence of all the noise-pairs was $\rho = 0.9922$.

After mixing, each noise was given a time interval shape with a total duration of 500 ms. A temporal window, $s(t)$, was applied such that there were raised-cosine edges with rise/fall times of 30 ms and a full-on duration of 440 ms. Consequently, the stimuli were made with both spectral and temporal windows. Both had raised-cosine edges. The application of the temporal window could change the coherence of the noise; therefore, the value of the computed stimulus waveform coherence was measured after the window was applied so that $\rho = 0.9922 \pm 0.0001$. Noise-pairs that did not meet this criterion were rejected. For the 100 noise-pairs accepted in this experiment, 875 were rejected.

It should be noted that the coherence value of 0.9922 is a long-term value computed over the 500-ms stimulus. It differs from the short-term coherence—typically averaged over tens of milliseconds—in some models of binaural detection (e.g., Kollmeier and Gilkey, 1990). The long-term coherence is the relevant *stimulus* cross-correlation in incoherence detection experiments, such as those by Gabriel and

Colburn (1981), and in MLD experiments that are described in terms of signal-to-noise ratios.

To determine the time-dependent interaural phase difference (IPD) and interaural level difference (ILD), the analytic signals were found. By eliminating the negative frequencies, the analytic signal for either x_L or x_R is

$$\hat{x}(t) = s(t)\sum_{n=1}^{N} C_n \exp[i(\omega_n t + \phi_n)], \tag{7}$$

where $s(t)$ is the temporal window, C_n and ϕ_n are left or right amplitudes and phases as required, and i is the square root of -1.

By Euler's relation, the analytic signal becomes

$$\hat{x}(t) = s(t)\sum_{n=1}^{N} C_n[\cos(\omega_n t + \phi_n) + i\,\sin(\omega_n t + \phi_n)]$$
$$= F(t) + iG(t). \tag{8}$$

The phase and envelope of the analytic signal as functions of time are

$$\phi(t) = \arg[G(t), F(t)] \tag{9}$$

and

$$E(t) = \sqrt{F^2(t) + G^2(t)}, \tag{10}$$

where the arg function is the arctangent with possible quadrant correction. The time-dependent IPD (rad) and ILD (dB) of the analytic signal are then defined as

$$\Delta\Phi(t) = \phi_R(t) - \phi_L(t) \tag{11}$$

and

$$\Delta L(t) = 20\log_{10}\left[\frac{E_R(t)}{E_L(t)}\right]. \tag{12}$$

Equation (11) yields a positive value of $\Delta\Phi(t)$ for signals that lead in the right ear. Similarly, Eq. (12) gives a positive value of $\Delta L(t)$ for a signal that has a larger level in the right ear. The interaural phase, $\Delta\Phi(t)$, was required to remain in the range between $-\pi$ and π rad at every point in time, and was corrected by adding or subtracting 2π when necessary. Because of the narrow bandwidth, the IPD is proportional to the ITD, which can be calculated by dividing the IPD by the center angular frequency of $2\pi \times 500$.

B. Signals

The IPD fluctuation over time is defined as

$$s_t[\Delta\Phi] = \sqrt{\frac{1}{T}\int_0^T [\Delta\Phi(t) - \overline{\Delta\Phi}]^2 dt}, \tag{13}$$

where T is the duration of the noise, 500 ms, $\overline{\Delta\Phi}$ is the mean IPD computed over time (normally very close to zero), and the t subscript indicates a standard deviation computed over time. The ILD fluctuation over time is defined as

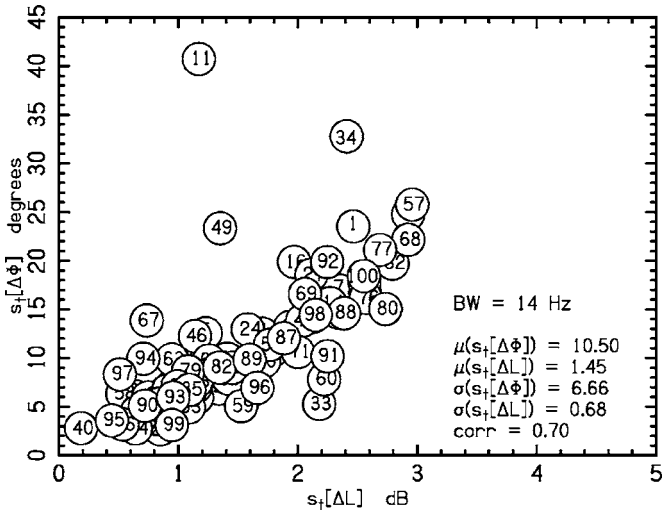


FIG. 1. Fluctuations of IPD vs fluctuations of ILD for the collection of 100 reproducible noise-pairs having a 14-Hz bandwidth, as used in Experiment 1. Each noise-pair is labeled by a serial number indicating only the order of creation. The means, standard deviations, and the correlation between $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ (denoted by corr) are reported.

$$s_t[\Delta L] = \sqrt{\frac{1}{T}\int_0^T [\Delta L(t) - \overline{\Delta L}]^2 dt}. \tag{14}$$

Figure 1 shows $s_t[\Delta\Phi]$ plotted against $s_t[\Delta L]$ for the 14-Hz bandwidth noise-pairs. The mean, standard deviation, and correlation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ are reported in Fig. 1. It is interesting to see that the correlation between phase fluctuations and level fluctuations across different noise-pairs is fairly strong, 0.70. To determine whether the statistics of our 100-noise ensembles were typical, we computed similar statistics for ensembles of 5000. These computations are described in the Appendix, which shows that a correlation of 0.70 is somewhat low compared to a typical value for this bandwidth, namely 0.80.

To perform Experiment 1, the five noise-pairs with the greatest fluctuations in IPD and the five noise-pairs with the smallest fluctuations in IPD were selected to form a *phase set* of ten noise-pairs. Similarly, the ten noise-pairs with the greatest and smallest fluctuations in ILD were selected to form a *level set*.

The two channels of noise were computed and downloaded by a Tucker-Davis AP2 array processor (System II) and converted from digital signals by 16-bit DACs (DD1). The buffer size was 4000 samples per channel and the sample rate was 8 ksps; thus the duration of the noises was 500 ms. The noise was lowpass filtered with a carrier frequency of 4 kHz and a −115 dB/octave roll off. The noises were presented at 70 dB SPL±3 dB with levels determined by programmable attenuators (PA4, prior to lowpass filtering) operating in parallel on the two channels. The levels of the two channels were equal and were randomized over a range of ±3 dB for each of the three intervals within a trial in order to discourage the listener from trying to use level cues to perform the task.

TABLE I. The p-values from a two-sample one-tailed t-test for phase and level sets with different bandwidths—Experiments 1, 2, and 3.

	14 Hz		108 Hz		2394 Hz	
Listener	Phase	Level	Phase	Level	Phase	Level
D	<0.001	<0.001	0.024	0.007	0.907	0.375
M	0.002	0.002	0.019	0.047	0.981	0.080
P	0.011	<0.001	0.187	0.187	0.683	0.654
W	0.001	<0.001	0.002	0.026	0.947	0.866

C. Procedure

Listeners were tested individually, seated in a double-walled sound attenuating room and using Sennheiser HD414 headphones. Six runs were devoted to listening to a set of ten reproducible noise-pairs. A noise-pair could be presented either incoherently–the dichotic presentation of x_L and x_R–or it could be presented coherently–the diotic presentation of x_L. A run consisted of 60 trials, where each of the ten noise-pairs in a set was presented incoherently a total of six times. Thus, listeners heard an individual noise-pair incoherently a total of 36 times (six runs times six presentations per run).

On each trial the listener heard a three-interval sequence. The first interval was the standard interval, which was always a coherent noise. The second interval was randomly chosen to be either incoherent or coherent. The third interval was the opposite of the second (e.g., if the second interval was coherent, the third interval was incoherent). The two coherent presentations were randomly selected from the remaining nine reproducible noises in the set except that they were required to be different from the incoherent "odd" interval and to be different from one another. The interinterval duration was 150 ms.

D. Data collection

Listeners used a four-button response-box to make decisions. Four buttons were used so that listeners could respond to the correct interval with a confidence estimate. The buttons from left to right were 2!, 2, 3, and 3!, representing confident second interval, second interval, third interval, and confident third interval respectively. Listeners were instructed to use a confident response only if there was no uncertainty as to which interval was incoherent. If more than one incorrect confident response occurred during a run, the run was terminated immediately, and the listener was obliged to begin again. After a decision was made by the listener, the next trial began following an intertrial duration of 900 ms.

There were two reasons to introduce the confidence measure in this experiment. The first was that for a given coherence, it proved to be noticeably easier to detect incoherence in some stimuli. Thus, extra weighting was wanted for identifying obviously incoherent sounds. The second was to increase the "dynamic range" of the experiment. The dynamic range problem arose because it was necessary to use the same waveforms, with a coherence of $\rho=0.9922$, for all the listeners, and it was further desired to use the same coherence for waveforms with different bandwidths. However, some listeners were better at the task than others, as shown by the percentage of correct (P_c) responses, and some bandwidths led to better performance than others. The confidence measure increased the dynamic range by preventing ceiling effects for the most successful listeners and easiest bandwidths.

The data collection procedure kept track of both the percentage of correct responses, which ignored the confidence estimate (e.g., responses of 2 and 2! were not treated differently), and a confidence adjusted score (CAS). The CAS is defined as the number of times the listener responded correctly plus the number of times that the listener was confident about the correct response. Since an individual noise-pair was heard 36 times, it was possible for a listener to get a score of 72 if the listener was able to respond correctly and confidently for all 36 presentations. This article will report only the CAS. In comparison with P_c, the use of CAS improved interlistener correlation, moved p-values of t-tests to greater significance, and improved the agreement that was achievable by models of binaural processing. Further justification of this technique together with the P_c results are in the first author's thesis (Goupell, 2005).[2]

E. Listeners

This article employed four male listeners, D, M, P, and W. Listeners D, M, and P were between the ages of 20 and 30 and had normal hearing according to standard audiometric tests and histories. Listener W was 65 and had a mild bilateral hearing loss, but only at frequencies four octaves above the center frequency used in this experiment. Listeners M and W were the authors.

F. Results

Figure 2(a) shows the CAS values for the selected phase set in Experiment 1. The five smallest $s_t[\Delta\Phi]$ noise-pairs are to the left of the vertical dashed line. The five largest $s_t[\Delta\Phi]$ noise-pairs are to the right of the dashed line. The dashed line thus represents a gap of 90 unused noise-pairs. All four listeners usually show a greater CAS for the five largest $s_t[\Delta\Phi]$ noise-pairs compared to the five smallest. In a two-sample t-test, all four differences were significant at the 0.02 level. The individual p-values are shown in the left-most block of Table I. Figure 2(a) also shows that listeners tend to agree about which noise-pairs are easy and which are difficult. The left-most block of Table II shows the correlations between listeners; all of them are above 0.77.

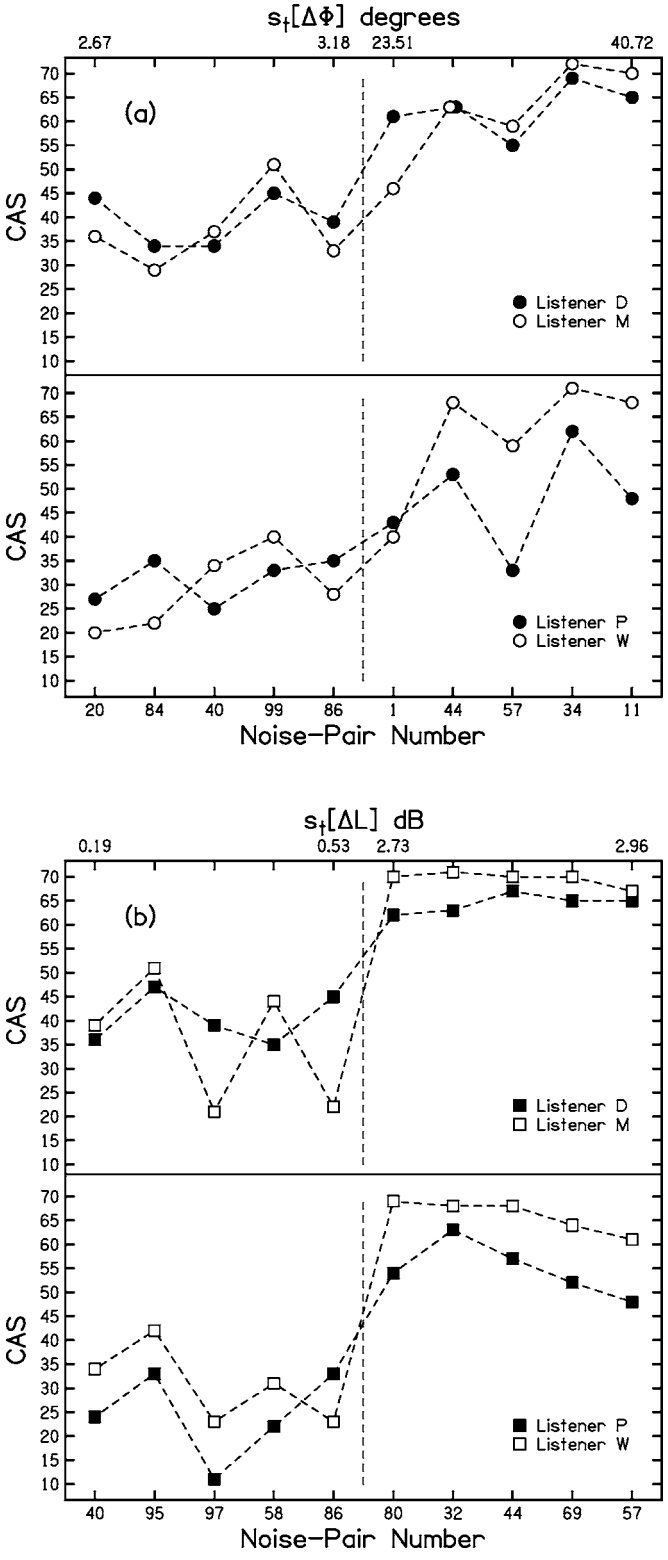


FIG. 2. (a) The confidence adjusted scores (CAS) for four listeners for the *phase set* of Experiment 1, the 14-Hz noise-pairs. The noise-pairs were chosen to have the five smallest and the five largest values of $s_t[\Delta\Phi]$ in the collection of 100 noise-pairs. The noise-pairs are rank ordered by increasing $s_t[\Delta\Phi]$ along the horizontal axis. The vertical dashed line represents 90 unused noise-pairs. The CAS values are higher for noise-pairs with the largest $s_t[\Delta\Phi]$ than for noise-pairs with the smallest $s_t[\Delta\Phi]$ for all listeners. The plots of CAS scores vs noise-pair serial number show a large measure of agreement among listeners. (b) The CAS for four listeners for the *level set* of Experiment 1, the 14-Hz noise-pairs. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs.

Figure 2(b) shows the CAS values for the level set for Experiment 1. Figure 2(b) exhibits results similar to Fig. 2(a) for the phase set. As shown by Tables I and II, the p-values are all significant at the 0.02 level, and the correlations between the individual listeners are all greater than or equal to 0.86.

The percent-correct values were computed to compare with the CAS. In terms of P_c, seven of eight t-tests were significant at the 0.05 level, and two were significant at the 0.02 level. The average interlistener correlation in P_c was 0.56 for the phase set and 0.64 for the level set. Therefore the P_c results also show significant differences among the noise-pairs, but the P_c results are less robust than the CAS results due to ceiling effects.

G. Discussion

Experiment 1 shows that the peak of the cross-correlation function is not an adequate predictor of the detectability of incoherence for narrowband noise. Instead, the fluctuations in interaural phase and interaural level as measured by the standard deviation are clearly related to incoherence detection.

The interlistener correlation of 0.77 or greater indicates that listeners agree in detail about the kinds of fluctuations that are easy or hard to detect. This correlation can also be seen in Figs. 2(a) and 2(b). The correlation between listeners is clear, even when the detection results are out of line with expectation based on interaural fluctuations. For example, Fig. 2(a) shows that the incoherence in noise-pair #57 was difficult to detect compared to other large-fluctuation noise-pairs of the phase set, and Fig. 2(b) shows that incoherence in noise-pair #95 was easy to detect compared to the small-fluctuation noise-pairs of the level set.

The anomalous results obtained with noise-pairs such as #57 and #95 indicate that the phase sets and level sets were not assembled using optimum criteria. For example, when the envelope is small, dramatic phase fluctuations can occur because the arg function in Eq. (9) becomes unstable. However, when the envelope is small, such phase fluctuations would be hard to detect. Nevertheless these fluctuations are not discounted in the criteria. A serious attempt to optimize criteria will be the subject of a following article. For the present, it is enough to note that the criteria as applied represent the phase and level fluctuations well enough to lead to highly significant differences.

Last, it should be reported that some noise-pairs [e.g., #97 in Fig. 2(b)] led to values of percent correct that were well below chance. Through informal listening experiments, it was found that these noise-pairs had very little roughness in their waveforms. It seems possible that when the fluctuations of IPD and ILD are barely detectable, listeners may sometimes mistake roughness for incoherence. They would then be reluctant to say that a particularly smooth sounding noise-pair is incoherent, which could lead to a P_c less than 50%. (See Sec. VI.)

TABLE II. The correlations between listeners for phase and levels sets with three different bandwidths—Experiments 1, 2, and 3.

Listener pair	14 Hz		108 Hz		2394 Hz	
	Phase	Level	Phase	Level	Phase	Level
D-M	0.90	0.86	0.42	0.54	0.57	0.24
D-P	0.85	0.93	0.05	−0.04	−0.06	0.14
D-W	0.87	0.93	0.67	0.58	0.43	0.13
M-P	0.77	0.89	0.31	−0.03	0.17	0.04
M-W	0.97	0.97	0.33	0.30	0.38	−0.68
P-W	0.79	0.93	0.49	0.31	0.03	−0.49

III. EXPERIMENT 2: INCREASED NOISE BANDWIDTH

Experiment 2 was identical to Experiment 1 except that the bandwidth was increased from 14 to 108 Hz. The value of the coherence remained fixed at 0.9922.

A. Method

As for Experiment 1, the geometric mean frequency was 500 Hz and the spectral spacing was 2 Hz. The spectral shaping again had 5-Hz edges so that components 444–449 and 555–560 Hz were shaped with raised-cosine functions, and components between 449 and 555 Hz had unity amplitude. The sample rate was 8 ksps and the analog output was low-pass filtered with a 4-kHz corner frequency.

Figure 3 shows $s_t[\Delta\Phi]$ vs $s_t[\Delta L]$ for the 108-Hz bandwidth noise-pairs. The black dots represent the noise-pairs with 14-Hz bandwidth from Experiment 1 for comparison. The means of the distributions of fluctuations increased only slightly when compared to the 14-Hz bandwidth waveforms. The insensitivity of the mean fluctuation to the bandwidth has been seen before, e.g., Fig. 3 of Breebaart and Kohlrausch (2001). By contrast, the standard deviations across different waveforms decreased. The standard deviation of the phase fluctuations decreased from 6.66° to 3.72°. The standard deviation of the level fluctuations decreased from 0.68 to 0.40 dB. The correlation between $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, evaluated over the ensembles of 100 waveforms, was essentially the same, 0.73 compared to 0.70. The Appendix shows that a value of 0.73 is in line with expectation based on statistics for a large number of noise-pairs.

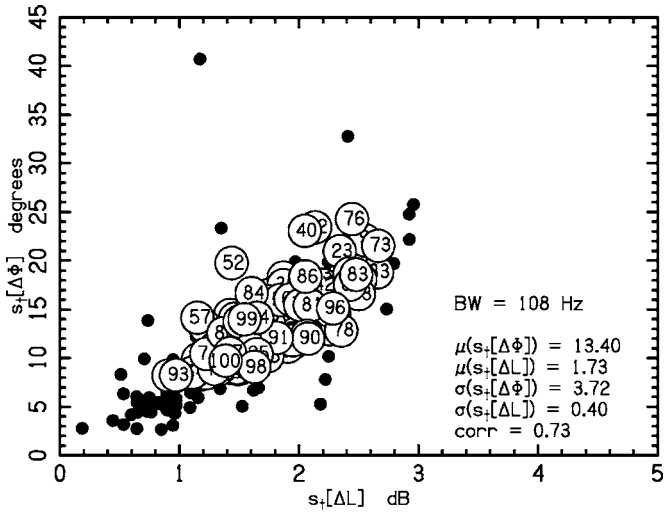


FIG. 3. Fluctuations of IPD vs fluctuations of ILD for the 100 reproducible noise-pairs with a 108-Hz bandwidth. Each noise-pair is labeled by a serial number. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The closed dots represent the noise-pairs with 14-Hz bandwidth from Experiment 1, as shown in Fig. 1. The means remained about the same as in Fig. 1, but the standard deviations and IPD-ILD correlation decreased.

B. Results

Figures 4(a) and 4(b) show the CAS values for the four listeners for the phase and level sets for the 108-Hz noise-pairs of Experiment 2. Table I shows that three of the four t-tests from the phase set were significant at the 0.05 level. That means that the differences between the CAS values for the five largest fluctuations and the CAS values for the five smallest fluctuations were significant for three listeners. Also, three of the four t-tests from the level set led to differences that were significant at the 0.05 level. In both cases, it was Listener P who did not show a significant difference. Further, as shown in Table II, the correlation between listeners was smaller for the wider bandwidth for all listener pairs. The correlation between listeners was 0.38 on average for the phase set and 0.28 on average for the level set. Listener P showed negative correlations with Listeners D and M.

C. Discussion

The standard deviation of the fluctuations computed across the 100 different waveforms of the ensemble decreased by more than 40% when the bandwidth was increased from 14 to 108 Hz. Since these fluctuations were found to correlate with incoherence detection from Experiment 1, it was not surprising to find that there was less variation in the listeners' ability to detect incoherence at the wider bandwidth.

The bandwidth of 108 Hz may be of special interest because this bandwidth approximately corresponds to a critical bandwidth at 500 Hz, and critical band noise has often been used in binaural experiments. For instance, Koehnke *et al.* (1986) used a noise with a bandwidth of 114 Hz centered on 500 Hz, and Evilsizer *et al.* (2002) used a bandwidth of 100 Hz. Our experiments found that the tests comparing detection performance with the size of the phase and level fluctuations led to a significant difference at the 0.05 level on six of eight of the tests. This can be compared to all of the tests

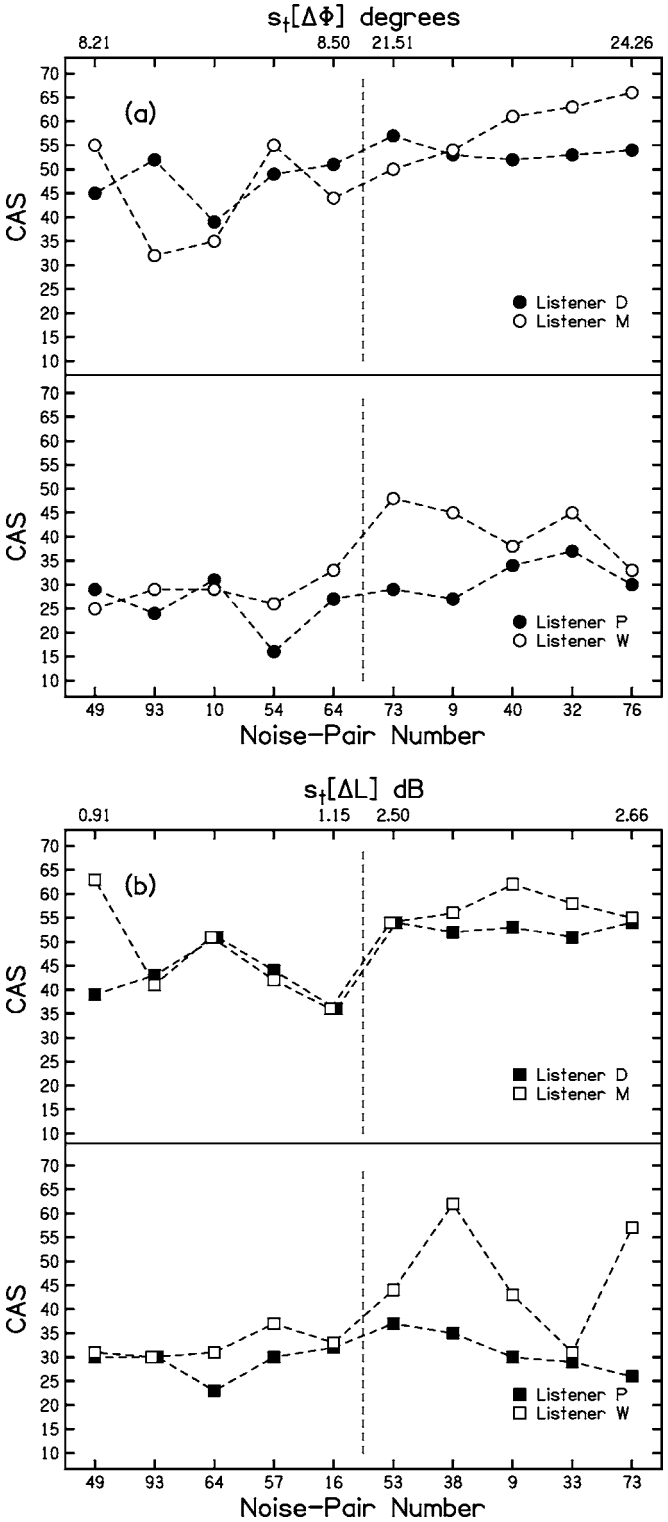


FIG. 4. (a)The CAS for the phase set of Experiment 2, the 108-Hz noise-pairs. The noise-pairs to the left and right of the dashed vertical line were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noise-pairs. On average, the CAS is smaller than for the 14-Hz noise-pairs shown in Fig. 2(a). The difference between the left and right noise-pairs is less dramatic than in Fig. 2(a). Also, the plots of CAS vs noise-pair serial number show less agreement among listeners than in Fig. 2(a). (b) The CAS for the level set of Experiment 2, the 108-Hz noise-pairs. The noise-pairs to the left and right of the dashed vertical line were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs. On average, the CAS is smaller than for the 14-Hz noise-pairs as shown in Fig. 2(b). The difference between the left and right noise-pairs is less dramatic than in Fig. 2(b). Also, the plots of CAS vs noise-pair serial number show less agreement among listeners than in Fig. 2(b).

being significant at the 0.02 level in Experiment 1. It was expected that if the bandwidth were further increased the variation in fluctuations would continue to decrease and the incoherence detection performance would be approximately the same for all the waveforms in the ensemble. In that case, the ability to detect incoherence would be only a function of the incoherence measure itself. That expectation led to Experiment 3.

IV. EXPERIMENT 3: WIDE BAND

A. Method

Experiment 3 was identical to Experiments 1 and 2 except that the bandwidth was increased to 2394 Hz. The geometric mean frequency of 500 Hz and the coherence of 0.9922 remained the same. Spectral components between 105 and 2495 Hz had unity amplitude and components in the ranges of 100–105 and 2495–2500 Hz were shaped by the raised-cosine spectral envelope. The sample rate and lowpass corner frequency were again 8 ksps and 4 kHz.

For Experiment 3, with a band more than four octaves wide, several features of the narrower-band stimuli in Experiments 1 and 2 no longer apply. Interaural phase differences are no longer equivalent to interaural time differences, and the temporal variations can no longer be neatly separated into fine structure and envelope. The procedures for selecting waveforms in Experiment 3 followed those for Experiments 1 and 2 in that noise-pairs were selected according to their IPDs (not ITDs), and the mathematical definition of the envelope was formally extended to the wider bandwidth without change. Because binaural processing is expected to occur within tuned channels, a selection of noise-pairs according to criteria based on the entire bandwidth does not have an evident perceptual significance. However, following the procedure of Experiments 1 and 2 did lead to uniformity of treatment.

Figure 5 shows the distribution of interaural phase and level fluctuations for one-hundred 2394-Hz bandwidth noises. The corresponding values for Experiments 1 and 2 are shown by the small and large black dots, respectively. The means of the distribution remained about the same as in Experiment 2 in that the ensemble average phase fluctuation was about 13° and the level fluctuation was approximately 1.7 dB. However, the standard deviations of the phase and level fluctuations decreased dramatically when the bandwidth was increased to 2394 Hz, as shown by the σ values in Fig. 5, respectively 1.07° and 0.11 dB. The correlation between level and phase fluctuations decreased to 0.27. The dramatic decrease in the variability of interaural fluctuations approaches the wideband limit where fluctuations are entirely determined by the coherence with negligible variation for individual waveforms.

B. Results

The results of the incoherence detection experiments, expressed as CAS values, are shown in Figs. 6(a) and 6(b) for the phase and level sets respectively. The CAS values in the figures are remarkably flat compared to Figs. 2(a), 2(b), 4(a), and 4(b). With four listeners and two sets there were

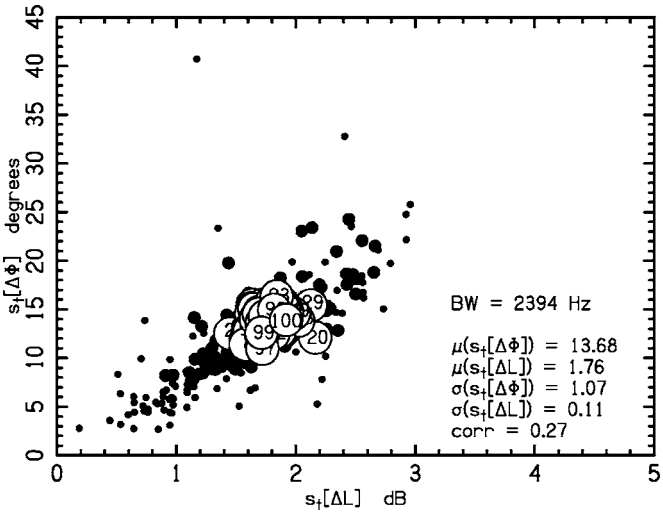


FIG. 5. Fluctuations of IPD vs fluctuations of ILD for the 100 reproducible noise-pairs with a 2394-Hz bandwidth. Each noise-pair is labeled by a serial number. The means, standard deviations, and IPD-ILD correlation of the distributions are reported. The small closed dots represent the 14-Hz noise-pairs from Experiment 1 and the larger closed dots represent the 108-Hz noise-pairs from Experiment 2, as shown in Figs. 1 and 3. The means are about the same as those in Figs. 1 and 2 but the standard deviations and correlation are smaller.

eight possible significance tests, and Table I shows that none of them led to a significant difference at the 0.05 level. Further, Table II shows that the interlistener correlations were smaller on average than for the other two bandwidths.

C. Discussion

The stimuli for Experiments 1, 2, and 3 show clear trends in the ranges of fluctuations over waveforms. The broadband stimuli of Experiment 3 approach the wideband limit. The incoherence detection results for Experiments 1, 2, and 3 also show clear trends. For narrow bands, detection depends on the details of interaural fluctuations, and it is not possible to predict detection performance if one knows only the value of coherence. As the bandwidth increases, the coherence becomes a better predictor of detection performance. In addition to the wideband limit for the stimulus, one can imagine a wideband *perceptual* limit in which incoherence detection is entirely determined by the coherence with negligible variation for individual waveforms.

The wideband perceptual limit was apparently reached for the bandwidth of 2394 Hz because no systematic differences appeared between those noise-pairs with the largest fluctuations and those noise-pairs with the smallest. However, this evidence is suspect because it is not clear that those large and small fluctuations, which were computed over the entire band of noise, are perceptually relevant. Instead, following binaural theory, one would expect that fluctuations computed over a critical band would be relevant. For instance, the MLD experiments by van de Par and Kohlrausch (1999) indicated that listeners detect an $S\pi$ signal in noise by monitoring a critical band. This difficulty has two parts. The first is in the interpretation of the data from Experiment 3. The second is in understanding the nature of the wideband perceptual limit.

If listeners monitor a critical band, in agreement with binaural theories, then our choices of noise-pairs in Experiment 3 were incorrectly made and there is no reason to expect better detection performance for noise-pairs to the right of the vertical dashed line compared to those on the left. Of course, given a bandwidth of 2394 Hz there is no way to know which critical band listeners might monitor, but it is usual to assume that listeners learn to monitor a band that is useful to them in performing the task. Therefore, it is expected that listeners would monitor a band somewhere in the 500-Hz neighborhood, where IPD fluctuations are of greatest use. Consequently the fluctuations in a critical band would resemble those for the 108-Hz bandwidth in Experiment 2. Possibly the range of fluctuations might be somewhat larger than in Experiment 2 because critical bands near 500 Hz are somewhat narrower than 108 Hz. In that case, the uniformity of performance across the 17 noise-pairs of Experiment 3 poses a problem. Whether or not the noise-pairs were well chosen, the odds are good that the relevant fluctuations, whatever they might be, would be uncharacteristically strong in at least one of the 17 noise-pairs. Thus it would be expected that at least one pair would stand out from the rest, but there is no evidence of that experimentally.

A second difficulty posed by our data for a model incorporating critical band listening arises from the overall level of detection performance. Listeners performed much less well for the wide band in Experiment 3 compared to the 108-Hz band of Experiment 2. Comparing CAS values in Figs. 4(a) and 6(a) (phase sets) and comparing values in Figs. 4(b) and 6(b) (level sets) show that the most successful listeners, D and M, had higher scores even for the most difficult noise-pairs at 108 Hz than for *any* noise-pairs with wide bands. Listeners P and W also scored consistently better at 108-Hz for the level sets. If listeners were able to take advantage of the slower and potentially larger fluctuations in a critical-band portion of the wide band one would have expected that some listener would have scored well for some one of the noise-pairs, contrary to the results of Experiment 3.

The data of Experiment 3 are more easily understood from an alternative view which says that in a broadband incoherence detection experiment, listeners cannot monitor a critical band but must monitor a much wider band. There is evidence in favor of this view in the detection results of Gabriel and Colburn (1981) where incoherence jnds generally increased for increasing bandwidth and continued to increase as the bandwidth went from 1000 to 4500 Hz. There is informal evidence in favor of the wider analysis bandwidth in the fact that the spatial fluctuations seem to the authors to be faster for the 2394-Hz bandwidth than for 108 Hz.

If the operative perceptual bandwidth were to be as large as the stimulus bandwidth, then our choice of noise-pairs would be correct and Experiment 3 would have demonstrated the wideband perceptual limit. Although it is hard to argue for a perceptual bandwidth as large as the stimulus bandwidth, the data do suggest that the perceptually relevant frequency range extends beyond a critical band. A similar conclusion from a MLD experiment was reached by Evilsizer *et al.* (2002), who found that $NoS\pi$ detection was *not*

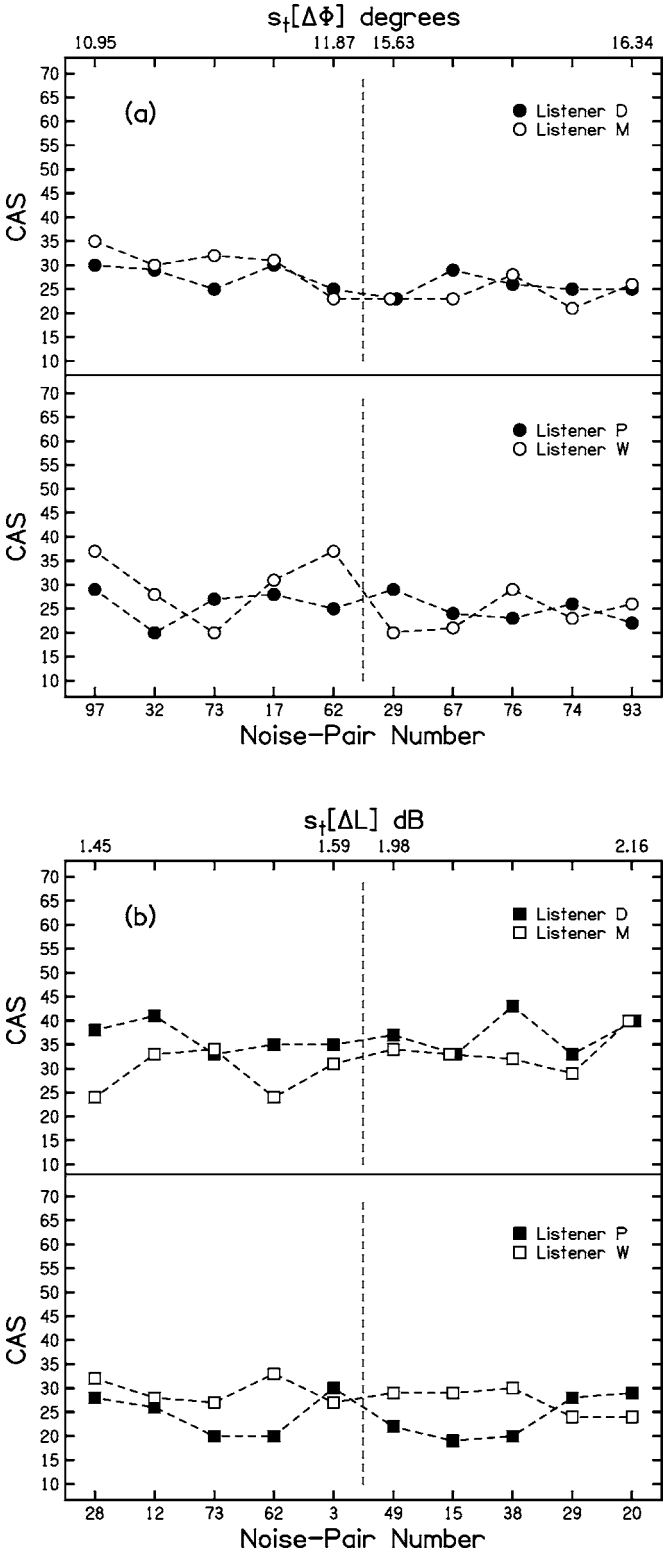


FIG. 6. (a) The CAS for the phase set of Experiment 3, the 2394-Hz noise-pairs. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noise-pairs. Compared to Figs. 2(a) and 4(a), the differences between CAS values for different noise-pairs are less dramatic and the correlation among individual listeners is smaller. (b) The CAS for the level set of Experiment 3, the 2394-Hz noise-pairs. The noise-pairs were chosen to have the smallest and largest $s_t[\Delta L]$ in the collection of 100 noise-pairs. Compared to Figs. 2(b) and 4(b), the differences between CAS values for different noise-pairs are less dramatic and the correlation among individual listeners is smaller.

solely determined by the masker spectrum within a critical band centered on the target.

The wideband perceptual limit is defined as the condition in which incoherence detection performance is predictable from the coherence alone. As noted previously, the odds are that Experiment 3 actually reached that limit. There remains the problem, however, of understanding what that limit represents. If the perceptual bandwidth is as large as the stimulus bandwidth then the wideband perceptual limit is not different in character from the wideband limit for the stimulus. The perceptual limit is reached when there are so many spectral components in the stimulus that, for a given coherence, the sizes of fluctuations are about the same for all different noises.

If the perceptual bandwidth is as small as a critical band, large fluctuations are available to the listener at a peripheral level, but for some reason listeners cannot take advantage of them in detecting incoherence. Perhaps the wideband limit should be thought of as a multiband limit wherein fluctuations in some bands serve as distractors that detract from the detection of fluctuations in other bands. Incoherence detection in broadband noise raises questions of optimum filter choice similar to questions that arise in MLD experiments with respect to off-frequency listening strategies. However, in the case of incoherence detection, there would seem to be no systematic approach to these questions because what corresponds to the signal varies with individual waveforms.

V. EXPERIMENT 4: THE ROLE OF BANDWIDTH

Experiments 1, 2, and 3 demonstrated that as the bandwidth increases two effects occur. First, the ranges of fluctuations in IPD and ILD become narrower. Second, the ability of listeners to detect incoherence depends less on the individual noise-pairs and is better determined by the value of coherence itself. According to our hypothesis, the second effect is the direct result of the first, and the main effect of a variation in bandwidth is to alter the distributions of interaural variances. Experiment 4 was designed to test this idea.

A. Method

To test the hypothesis, we assembled a subset of noise-pairs from a new collection of 1000 noise-pairs with a bandwidth of 14 Hz to make a "matched set" whose members were selected to best match the fluctuations in the noise-pairs from Experiment 2, which had a bandwidth of 108 Hz. Experiment 2 included 20 noise-pairs, ten for the phase set and ten for the level set, as determined by the five largest and five smallest fluctuations. However, phase and level fluctuations tend to be correlated and five of the noise-pairs were common to the phase and level sets. Therefore, there were only 15 distinct noise-pairs in Experiment 2. For each of these 108-Hz bandwidth noise-pairs a 14-Hz bandwidth noise-pair was selected that best matched the fluctuations in phase and level. The selection is illustrated by the 15 open and filled circles in Fig. 7. Phase and level sets using the matched noise-pairs formed the stimuli for Experiment 4, which was otherwise identical to the other experiments of this article.

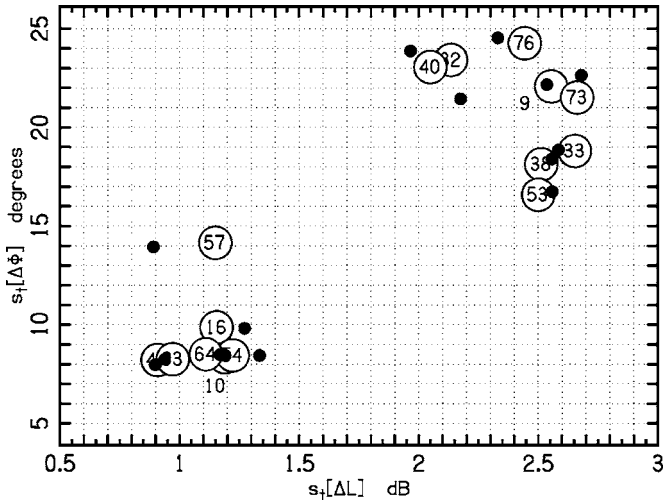


FIG. 7. The fluctuations in IPD vs the fluctuations in ILD for 15 noise-pairs used in Experiment 2 (bandwidth of 108 Hz) are shown by open circles. The best matching noise-pairs from Experiment 4 (bandwidth of 14 Hz) are shown by closed circles. These noise-pairs are called the "matched set." Incoherence detection results for these corresponding pairs were compared to determine the role of bandwidth *per se* when fluctuations were held constant.

B. Results

According to our working hypothesis, the detection scores (CAS) for the matched 14-Hz sets from Experiment 4 ought to be identical to the detection scores in the 108-Hz sets from Experiment 2. The results of the comparison are shown in Figs. 8(a) and 8(b) for the phase and level sets, respectively. The phase set data in Fig. 8(a) show that the values of CAS are comparable for the two bandwidths when the IPD fluctuations are matched for Listeners D and M. However, for Listeners P and W performance is better for the 14-Hz bandwidth than for the 108-Hz bandwidth.

The level set data in Fig. 8(b) show comparable CAS values for the two bandwidths for all the listeners, though Listeners P and W still tend to show better performance at the smaller bandwidth—better on 14 of 20 possible comparisons. Therefore, the raw data, shown in Fig. 8, offer modest support for the hypothesis that bandwidth should be unimportant if the sizes of fluctuations are matched.

The hypothesis can be further tested by examining the relative detectability of the noise-pairs with the largest interaural fluctuations versus the noise-pairs with the smallest interaural fluctuations. These are, respectively, to the right and to the left of the vertical dashed line in Fig. 8. A t-test of the hypothesis that CAS scores are higher for the five noise-pairs on the right led to the *p*-values in Table III. There, it can be seen that three of eight *p*-values are significant at the 0.05 level for the 14-Hz matched sets, and that six of eight *p*-values are significant at that level for the targeted 108-Hz sets. By comparison, all eight *p*-values were significant at the 0.02 level for the 14-Hz bandwidth in Experiment 1 and none of the *p*-values were significant even at the 0.05 level for the wide bandwidth in Experiment 3. Thus, the matched phase and level sets appear to have approximately matched the relative difference in performance between the noise-

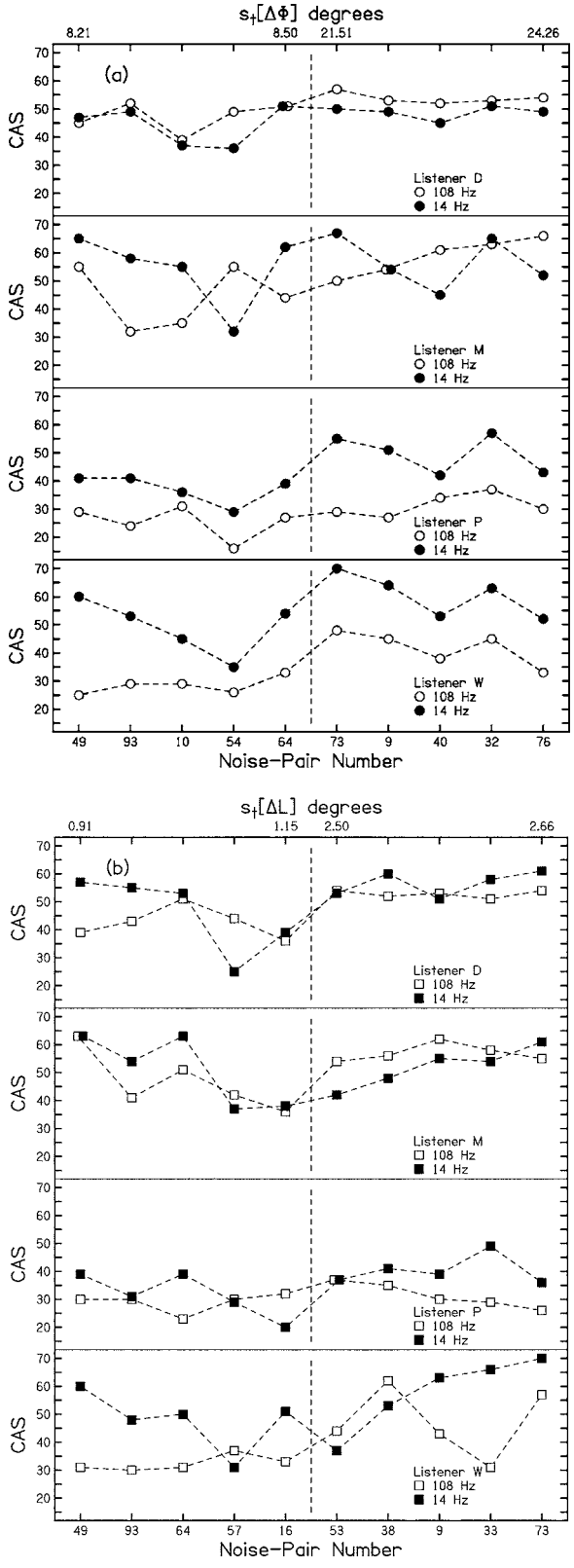


FIG. 8. (a) The CAS for Experiment 4, using the matched phase set selected from the 14-Hz noise-pairs are shown by closed circles. They are plotted together with the data from Experiment 2, the 108-Hz noise-pairs, shown by open circles, repeated from Fig. 4(a). (b) The CAS for Experiment 4, using the matched level set selected from the 14-Hz noise-pairs are shown by closed squares. They are plotted together with the data from Experiment 2, the 108-Hz noise-pairs, shown by open squares, repeated from Fig. 4(b).

TABLE III. The p-values from a two-sample one-tailed t-test for the 108-Hz phase and level sets in Experiment 2 (repeated from Table I) and for the matched 14-Hz phase and level sets in Experiment 4.

	14 Hz		108 Hz	
Listener	Phase	Level	Phase	Level
D	0.103	0.077	0.024	0.007
M	0.384	0.315	0.019	0.047
P	0.006	0.038	0.187	0.187
W	0.042	0.117	0.002	0.026

pairs in the Experiment 2 sets, consistent with the hypothesis that the size of interaural fluctuations determines the detection of incoherence.

However, there are differences between the results of Experiments 2 and 4. As noted above, there is a tendency for listeners P and W to score better on the 14-Hz bandwidth sets (Experiment 4). More impressive, a comparison of the interlistener correlations in Table IV shows that all 12 of the correlations are higher for the 14-Hz matched sets than for the 108-Hz bandwidth sets. Averaged over listener pairs and over phase and level sets, the interlistener correlation was 0.64 (sd=0.11) for the 14-Hz bandwidth and only 0.33 (sd =0.23) for the 108-Hz bandwidth.

C. Discussion

Experiment 4 attempted to construct a set of noise-pairs with a 14-Hz bandwidth that would lead to the same patterns of detection performance that had been seen in Experiment 2, which used noise-pairs with a 108-Hz bandwidth. This was done by best-matching the values of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ of the Experiment 2 noise-pairs with noise-pairs having a 14-Hz bandwidth. Experiment 4 was partially successful in that the number of significant p-values was comparable in Experiments 4 and 2, and quite different from Experiments 1 and 3. It is clear that the fluctuation matching procedure caused the 14-Hz bandwidth noise-pairs of Experiment 4 to behave more like the 108-Hz noise-pairs of Experiment 2 than like 14-Hz noise-pairs of Experiment 1.

Although the values of CAS for the 108-Hz bandwidth (Experiment 2) and the 14-Hz bandwidth (Experiment 4) were interleaved for two of the listeners, the values of CAS were higher for the 14-Hz bandwidth for the other two, and interlistener correlations were higher overall for the 14-Hz bandwidth. Thus, there seems to be an advantage for the smaller bandwidth, whether the fluctuations are large or small. An advantage for smaller bandwidth was also found by Gabriel and Colburn (1981) where the jnd for detecting incoherence decreased with decreasing bandwidth. The impressive listener correlation for the 14-Hz bandwidth compared to 108-Hz may find a straightforward explanation in terms of the number of instances of salient fluctuations. For a 14-Hz bandwidth, fluctuations tend to be slow and within a duration of 500 ms there are relatively few strong fluctuations. Because of the absence of many alternatives, listeners are likely to detect these fluctuations in similar ways. For a 108-Hz bandwidth, the number of fluctuations is about eight times greater and different listeners may use different fluctuations to perform the task.

TABLE IV. The correlations between listeners for the 108-Hz phase and level sets of Experiment 2 (repeated from Table II) and for the matched 14-Hz phase and level sets of Experiment 4.

	14 Hz		108 Hz	
Listener pair	Phase	Level	Phase	Level
D-M	0.68	0.71	0.42	0.54
D-P	0.60	0.66	0.05	−0.04
D-W	0.75	0.68	0.67	0.58
M-P	0.66	0.56	0.31	−0.03
M-W	0.80	0.69	0.33	0.30
P-W	0.91	0.46	0.49	0.31

VI. EXPERIMENT 5: MONAURAL CUES

During the course of our numerical study of interaural fluctuations we compared the interaural fluctuations in noise-pairs with the envelope fluctuations in the left-ear signal itself. (Left and right signals were so similar that it did not matter which we chose.) *A priori* it seemed possible that noise-pairs with especially large (small) interaural fluctuations might often be derived from individual noise tokens with especially large (small) envelope fluctuations.

To make the comparison we calculated the envelope fluctuation,

$$s_t[E] = \sqrt{\frac{1}{T}\int_0^T [E(t) - \bar{E}]^2 dt}, \tag{15}$$

where $E(t)$ is the envelope defined in Eq. (10), and $\bar{E}$ is the average over the time of the stimulus. Calculations with 1000 noise-pairs with 14-Hz bandwidth–stochastically identical to the pairs of Experiment 1–showed a positive correlation between envelope fluctuations and interaural fluctuations. Specifically, the correlation between $s_t[E]$ and $s_t[\Delta L]$ was 0.48. The correlation between $s_t[E]$ and $s_t[\Delta\Phi]$ was 0.43. The latter correlation would be hard to understand were it not for the strong correlation between the standard deviations of interaural phase and level differences, as shown in Figs. 1, 3, and 5.

We next made the same calculations for the 20 noise-pairs actually used in Experiment 1. The interaural phase fluctuations and the interaural level fluctuations correlated with the monaural envelope fluctuations at levels of 0.59 and 0.65, respectively. Evidently, those waveforms with interaural fluctuations that are especially large or small particularly owe their binaural character to the envelope of the original generating noise token.

Given the positive correlation between interaural and monaural fluctuations, it seemed possible that there might be information in the monaural signals that was used by listeners in performing the experiments of this article. Because of the evident importance of fluctuations in Experiment 1, our attention centered on the stimuli used there, with a bandwidth of 14 Hz.

A. Diotic experiment method

Experiment 5 was identical to Experiment 1 except for the important difference that the left-ear signal of Experiment 1 was the signal for both ears in Experiment 5. In a second difference, the listeners in Experiment 5 had all completed Experiments 1–4 and therefore were highly experienced. The listeners were given three-interval sequences as before and were asked to apply the same strategy that they had used in the previous experiments. There was reason to believe that this approach might be successful because all the listeners volunteered that in the previous experiments they based their decisions on a sense of width, choosing the interval–either two or three–with the larger width. It seemed possible that the sense of roughness or other "action" associated with a diotic stimulus having large fluctuations could be associated with a sense of width. Consequently, we expected that each trial of Experiment 5 would constitute a comparison between apparent "widths" for a particular identified noise and a different, randomly-chosen, noise from the set of ten. The stimulus sets were the phase set and level set from Experiment 1.

B. Results

Listeners made a negligible number of "confident" responses. Apparently the strong sensation of width elicited by some of the dichotic stimuli did not occur with any of the diotic noises. Therefore, the CAS had little value and results of Experiment 5 were analyzed in terms of the percentage of the trials on which a given noise was selected over other noises. The statistic will be called P_s, percent selected. It can be compared with P_c, the percent correct in the dichotic experiment.

1. Large fluctuation comparison

Particular interest centered on the five noise-pairs for which the interaural fluctuations were the greatest. By examining the scores from the diotic experiment we expected to gain insight into the role that envelope fluctuations may have played in the dichotic experiment. The average values of P_s for those five noise-pairs for listeners D, M, P, and W were, respectively: for the phase set(%): 54, 61, 54, 63; for the level set (%): 63, 58, 72, 70.

Evidently, in the diotic experiment listeners chose the noises that had led to the largest interaural fluctuations clearly more than half the time. These numbers can be compared with the values of P_c in the dichotic experiment (Experiment 1) which averaged 88%.

2. Agreement between listeners

The agreement among the listeners was assessed by comparing values of P_s against noise serial number for listeners taken in pairs. Interlistener correlations for D-M, D-P, D-W, M-P, M-W, and P-W were as follows: for the phase set: 0.59, −0.38, 0.38, −0.22, 0.88, 0.17; for the level set: 0.70, 0.53, 0.76, 0.67, 0.89, 0.76.

The strongest correlation was between M and W. Listener P was responsible for the only negative correlations–both in the phase set. Correlations were clearly larger in the level set than in the phase set. The strong correlation indicates that listeners tended to agree about which fluctuations were salient.

3. Comparison with envelope fluctuation

A comparison between the listener selection of noises and fluctuation was assessed by comparing P_s with $s_t[E]$ as a function of the noise-pair serial number. Correlations for listeners D, M, P, and W were, respectively: for the phase set: 0.38, 0.84, −0.03, 0.88; for the level set: 0.85, 0.66, 0.65, 0.70.

Again, P is responsible for the only negative correlation. The positive correlation indicates that the choices that listeners make can be predicted based on the physical envelope fluctuations, as measured by the standard deviation of the envelope over time, especially for the level set.

4. Comparison with Experiment 1

A comparison between the results of the corresponding diotic and dichotic experiments was made by comparing P_s on Experiment 5 with P_c on Experiment 1, both as functions of the noise-pair serial number. Correlations for listeners D, M, P, and W were, respectively: for the phase set: 0.24, 0.53, −0.07, 0.44; for the level set: 0.82, 0.66, 0.81, 0.80.

Again, correlations are larger for the level set.

C. Discussion

The above-mentioned correlations are fairly impressive, with some exceptions for listener P. These include the correlations between P_s in Experiment 5 and P_c in Experiment 1 as well as the correlations between P_s and monaural and dichotic fluctuations. There are several possible interpretations of these correlations.

Possibly the correlation between P_c and P_s scores only represents a chain of stimulus circumstances. For a narrow bandwidth like 14 Hz, every kind of stimulus fluctuation seems to correlate with every other kind. Interaural phase fluctuations correlate with interaural level fluctuations and both correlate with noise envelope fluctuations. In a dichotic experiment probing the detection of interaural incoherence listeners attend to the interaural fluctuations. In a diotic experiment probing an evaluation of roughness or other stimulus action listeners attend to the envelope fluctuations. The results of the two experiments, as functions of the stimulus serial number, are similar because the interaural and monaural fluctuations behave similarly with respect to serial number.

Alternatively, it is possible that the correlation between P_c and P_s scores arises because listeners in a dichotic experiment are misled by monaural envelope fluctuations that are particularly large or particularly small. Given the enormous difference in the average P_c and P_s values for the five noise-pairs with large fluctuations, as noted in Subsection 1, it seems highly unlikely that monaural fluctuations *per se* contribute to listener judgments in the dichotic experiment when the detection of interaural incoherence is easy. But when detection of interaural incoherence is difficult, or impossible, the cues from monaural envelope fluctuations (or the lack of

them) may influence judgments and are probably responsible for P_c values in a dichotic experiment that are less than chance.

VII. DISCUSSION AND CONCLUSION

Listeners are sensitive to small amounts of interaural incoherence in an otherwise diotic noise. Given a diotic noise as a comparison, listeners can detect a coherence change of 0.01, i.e., they are sensitive to the difference between 1.00 and 0.99 (Gabriel and Colburn, 1981). The goal of the present article was to understand the origin of this remarkable sensitivity.

A. Detection of incoherence

Experiment 1 selected stimuli from an ensemble of 100 reproducible left-right noise-pairs, all of which had a bandwidth of 14 Hz and an interaural coherence of 0.9922. It was found that those pairs that had a large fluctuation in IPD or large fluctuation in ILD were much more readily recognized as not perfectly coherent compared to pairs with small fluctuations. This result led to the conclusion that, for bandwidths as narrow as 14 Hz, the interaural coherence is not an adequate predictor of the ability to detect incoherence. Instead, the size of the interaural fluctuations tends to dominate.

Experiments 2 and 3 progressively increased the bandwidth and found that the ranges of fluctuations in IPD and ILD among different noise-pairs in an ensemble decreased with increasing bandwidth (see the Appendix). This observation led to the expectation that the detectability of incoherence would exhibit less variation for different noise-pairs with these wider bandwidths. Detection experiments similar to Experiment 1 showed an increasing uniformity in detectability for the incoherence in noises with increasing bandwidths as expected.

It was conjectured that the only reason that detection performance for 14-Hz bandwidth was different from performance for 108-Hz bandwidth was that more extreme values of average interaural fluctuations (both very small and very large) were available in the ensemble with the narrow bandwidth. In Experiment 4, a comparison was made between performance on noise-pairs with 108-Hz bandwidth and performance on a matched set of noise-pairs with 14-Hz bandwidth. Noise-pairs in the matched set were selected to have approximately the same interaural fluctuations as the pairs of the 108-Hz set. The comparison showed that detectability differences among different noise-pairs in the 14-Hz matched set were reduced to about the same level as for the 108-Hz noises-pairs, consistent with the conjecture. The overall performance on the 14-Hz matched set was approximately equal on the 108-Hz set for two of the listeners; it was consistently higher for the other two listeners. These two results from the comparison suggest that differences in interaural fluctuations are responsible for differences in the detectability of incoherence for different noises, but that for some listeners fluctuations of comparable size are more easily detected when the bandwidth is narrow. A possible explanation for the advantage of narrow bands is that fluctuations are slower. The role of fluctuation speed will be addressed in a following article.

Our conclusions differ from those of Breebaart and Kohlrausch (2001) who dismissed a specific role for IPD and ILD fluctuations in binaural detection because the distributions of those fluctuations, with or without a signal, failed to show a bandwidth dependence. By contrast their $N\rho S\pi$ detection experiments did show such a dependence. We agree that the $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ have mean values, averaged across waveforms, that are insensitive to bandwidth, as noted in Figs. 1, 3, and 5, and as shown in the widths of the distributions plotted by Breebaart and Kohlrausch (2001). However, the *variations* in $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ among different waveforms depend strongly on bandwidth, and these variations are responsible for the large detectability differences that we are reporting.

B. Caveats—duration and bandwidth

All the stimuli used in this article had a duration of 500 ms. The stimuli were all constructed to have an interaural coherence of 0.9922 computed over this duration. Because different noise-pairs exhibited very different detectabilities, this article reached a main conclusion that the coherence measure is inadequate to predict detection for narrow bands. A problem that arises with this conclusion is that one expects that the temporal analysis windows that are appropriate for binaural perception are considerably shorter than 500 ms. Therefore, it is possible that, after all, the coherence is a perfectly adequate statistic to predict incoherence detection, but that the coherence must be calculated over the correct (shorter) time interval(s). This point of view would say that it was a mistake to compute a fixed coherence over an interval as long as 500 ms and expect it to be perceptually valid because the coherence varies from one momentary analysis window to the next.

There are two responses to the above criticism. First, we have performed model calculations based on a running short-term cross-correlation, averaged over temporal windows ranging from 0 to 150 ms. The calculated results were compared with the results of incoherence detection experiments with 100 randomly chosen reproducible noises with bandwidths of 14 Hz and durations of 500 ms. It was found that short-term coherence is inadequate to account for the detectability of incoherence. Second, we have done a series of experiments with 14-Hz bandwidth and progressively shorter durations. Coherence values calculated over those short durations were again required to be 0.992. Stimuli were again selected based on large or small fluctuations over the duration, and listeners did discrimination experiments similar to Experiment 1. It was found that the results for durations of 100 and 50 ms were the same as the results for 500 ms in that noise-pairs with large fluctuations led to CAS values that were significantly greater compared to noise-pairs with the smallest fluctuations. Both these responses to the duration criticism appear in Goupell's thesis (2005) and will be discussed in articles to follow.

A second criticism pertains to the wideband noise-pairs

investigated in Experiment 3. Specific noise-pairs were chosen with large or small fluctuations as computed over the entire bandwidth. However, it is expected that the perceptually relevant bands would be determined by auditory filters similar to critical bands. If so, then the basis for choosing the pairs was faulty and it would not be surprising to find that there is no significant difference in CAS values for the different pairs. However, as noted in Sec. IV C, a model that begins with critical-band filtering runs into difficulties in explaining both the poor performance and the uniformity of performance in Experiment 3. It would become necessary to add some mechanism of cross-band interference to explain the data.

C. Binaural processing

The psychoacoustical literature often connects binaural capabilities, particularly binaural release from masking, with interaural cross-correlation (Wilbanks and Whitmore, 1967; Domnitz and Colburn, 1976; Koehnke *et al.*, 1986). Wilbanks (1971) cited historical articles by Cherry, Licklider, and Jeffress "…supporting the notion that the binaural system is, logically speaking, a correlational detector."

The experiments of the present article agree with that conclusion, but only as a statistical approximation that gains validity as the bandwidth becomes larger. A correlational model adequately reflects the fluctuations only in the wideband limit. Wideband noise signals tend to be ergodic wherein the statistical properties of an ensemble of noises become manifest in any given noise sample as that particular noise evolves in time. Then the fluctuations in any given sample of wideband noise are not much different from those in any other sample, and the size of the fluctuations appears to be a simple function of the coherence. The transition to that limit was seen in the stimuli for Experiments 1–3 where, with increasing bandwidth, the variance among different noises of IPD and ILD decreases.

The wideband limit for signals has a psychological parallel in the wideband perceptual limit, where incoherence detection becomes a simple function of coherence. The wideband limit is approached but not reached for a bandwidth of 108 Hz, which is close to a one-third octave or critical bandwidth for 500 Hz. With a bandwidth that wide, a psychoacoustical experiment might easily miss the inhomogeneity of noises unless it were specifically designed to look for it.

For a narrow bandwidth, such as 14 Hz, each individual noise has so few spectral components that the interaural properties of an individual noise-pair can differ greatly from the ensemble mean properties. Improbable variations of this kind are responsible for the fact that a small amount of incoherence may be difficult to detect for one sample of noise but easy to detect for a another one, as reported in this article. It is likely that this effect was present in many historically important studies of binaural effects in narrow bands. In their coherence discrimination experiment Gabriel and Colburn (1981) found the just-noticeable difference to decrease with decreasing bandwidth given a reference coherence of 1.0. The decrease may result from particularly favorable noise samples that occur for narrow bands.

Parallels to incoherence detection appear in MLD experiments in connection with the detection of a low-frequency signal in the $\mathrm{NoS}\pi$ condition. Van de Par and Kohlrausch (1999) found that as the bandwidth decreased, while remaining narrower than a critical band, threshold signal-to-overall-noise ratios remained approximately constant. The results of listeners D and M in Experiments 1 and 2 seem similar in that CAS values do not change much as the bandwidth changes from 14 to 108 Hz. By contrast, Zurek and Durlach (1987) found that the threshold signal to noise ratio decreases as the bandwidth decreases, a result that seems consistent with listeners P and W in Experiments 1 and 2 presented here. Zurek and Durlach (1987) interpreted the decrease in threshold that they observed as the result of binaural sluggishness and a consequent advantage for the relatively slow fluctuations in narrow bands. The results of our Experiment 4, using stimuli that matched the IPD and ILD fluctuations but not the fluctuation rate, support this interpretation too, at least for listeners P and W. However, our computational modeling for overall results (Goupell, 2005) does not support a role of binaural sluggishness. Instead, it supports a binaural system that is capable of responding to rapid fluctuations from a stable diotic standard stimulus as suggested by Hall *et al.* (1998) and references cited therein. For instance, Green (1966) found essentially no difference in binaural advantage for signals that were 1 s long and 10 ms long, and what difference there was suggested better binaural processing for 10 ms.

Although there are parallels between interaural incoherence detection and MLD tasks, there are also important differences. An $\mathrm{NoS}\pi$ stimulus with a sine signal and noise masker may include both static and dynamic interaural cues. An interaural incoherence detection experiment, as presented here, involves dynamic fluctuations only. Results can be different too. Using multiplied noise maskers and sine signals Breebaart *et al.* (1999) found that cross-correlation accounted for their signal detection data better than various measures of interaural fluctuations, a result that is entirely contradictory to our experience with incoherence detection. Further, the experimental data obtained by Breebaart *et al.* were best fitted by a model that used the energy in the binaural difference signal as a decision variable. For stimuli such as ours, constructed from orthogonalized waveforms all having the same coherence, the energy in the difference signal is the same in all waveforms. Such a model would predict no difference in a listener's ability to detect the incoherence in any of our noise-pairs, contrary to our experiments.

In a MLD experiment the cues for detecting the signal are reported to be different depending on the bandwidth. For narrow bands an $\mathrm{S}\pi$ signal contributes a width to the binaural image. For wide bands an $\mathrm{S}\pi$ signal increases the tone-like strength (Evilsizer *et al.*, 2002). By contrast, in an incoherence detection experiment only the width cue occurs, whatever the bandwidth, though the rate of the fluctuations that establish the width does vary with bandwidth.

Experimental evidence in favor of this bandwidth effect on cues is that in a MLD experiment, performance for NoSo is correlated with performance for $\mathrm{NoS}\pi$ for wide bands, but the correlation is significantly less for narrow bands (Evil-

sizer *et al.*, 2002.) (It should be noted that the bands called "narrow" by Evilsizer *et al.* are 100 Hz in width, equivalent to our "critical-band" noise-pairs.) See also Gilkey *et al.* (1985) and Isabelle and Colburn (1991) for similar evidence, as described by Evilsizer *et al.* Presumably this difference in correlation arises because listeners are using the tonal cue for wide bands, which is similar for So and Sπ, but are using the width cue for narrow bands, which has a binaural contribution only for Sπ.

Similarly, Evilsizer *et al.* found that interlistener correlation was strong for NoSo and also for wideband experiments for any combination of interaural noise and signal phases. Interlistener correlations dropped and became negative for NoSπ when the band was narrow (100 Hz). This result is consistent with the idea that the tonal cue is similarly detected by different listeners, but that there are marked individual differences when the cue becomes an image width for narrow bands (Bernstein *et al.*, 1998). The situation for incoherence detection is just the opposite. Interlistener correlations are strongest for narrow bands where the width cues vary greatly among different noise-pairs, and interlistener correlations are weak for wide bands where the width cues are similar for different pairs. A corollary to the above argument is that Evilsizer *et al.* would have found better correlation among listeners for Sπ in narrow bands had they used bands as narrow as ours.

In a MLD experiment with noise bands that are not narrow, a listener might listen in different subbands where the values of interaural coherence will be very different. In an incoherence detection experiment the stimulus is simpler because the coherence is spectrally homogeneous and advantageous sub-band listening possibilities are limited to special cases for individual noise samples as deviations from the ensemble mean statistics. In our view, the incoherence detection experiment has the advantage that it extracts the essential binaural element from a MLD experiment and exposes it for observation.

Experiment 1 for narrow-band stimuli clearly showed the importance of the dynamic fluctuations in interaural parameters, in contrast to the coherence measure of interaural differences. Experiment 1 gained its power from a selection of stimuli based on values of the standard deviation of IPD and ILD over time. These measures were plausible guesses about what is important to the binaural system, but there is no reason to expect them to be optimum measures of interaural fluctuations. In fact, it is clear that they are not optimum. Stimulus #57 in the 14-Hz phase set had one of the largest phase fluctuations and yet Fig. 2(a) shows that all four listeners had a relatively difficult time with it. Another article, to follow, explores alternative models for the detection of incoherence in narrowband noises and attempts to determine the best characterization of fluctuations from a perceptual point of view.

ACKNOWLEDGMENTS

The authors are grateful to Dr. H. S. Colburn and Dr. C. Trahiotis for useful discussions about coherence. Dr. N. I. Durlach and Dr. J. F. Culling provided important comments as did Associate Editor Dr. A. Kohlrausch. This work was supported in part by the National Institute on Deafness and Other Communicative Disorders, Grant No. DC00181.

APPENDIX: VARIATION AS FUNCTION OF BANDWIDTH

The mean values and standard deviations of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ computed across noise-pairs are given in in the text blocks of Figs. 1, 3, and 5 for a collection of 100 noise-pairs and for bandwidths of 14, 108, and 2394 Hz respectively. It can be seen that as bandwidth increases, the standard deviation decreases. The purpose of this appendix is to provide more precise values of the means and standard deviations and to fit a functional form to the variation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ as a function of bandwidth.

For each value of bandwidth, 5000 noise-pairs were generated using a mixture of two orthogonalized noises as described in Experiments 1–3. These noise-pairs included the same spectral envelope and temporal windowing as in the experiments. The results are given in Table V.

A power regression was used to find a power law describing the variation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ as a function of bandwidth. The power regression was of the form $y=mx^p$ where the bandwidth is the x-variable and the standard deviation of $s_t[\Delta\Phi]$ or $s_t[\Delta L]$ is the y-variable. The line of best fit was

$$y = 18.6x^{-0.37} \tag{A1}$$

for $s_t[\Delta\Phi]$ and

$$y = 1.93x^{-0.36} \tag{A2}$$

for $s_t[\Delta L]$. Both fits described over 99% of the variance of the points. In summary, the standard deviations of the IPD and ILD computed over 500 ms follow distributions that have standard deviations computed over 5000 different noise-pairs that vary approximately the as inverse cubed-root of the bandwidth.

TABLE V. Values of the mean and standard deviation of $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ for noise-pairs with three bandwidths: 14, 108, and 2394 Hz. Correlation between the standard deviations is also given. Each value is based on 5000 noise-pairs.

BW (Hz)	$\mu(s_t[\Delta\Phi])$ (deg)	$\mu(s_t[\Delta L])$ (dB)	$\sigma(s_t[\Delta\Phi])$ (deg)	$\sigma(s_t[\Delta L])$ (dB)	Correlation
14	10.78	1.45	6.95	0.77	0.80
108	13.19	1.71	3.40	0.35	0.73
2394	13.56	1.75	1.04	0.12	0.41

[1]In the description of the orthogonalization procedure by Culling *et al.* (2001), the denominator of Eq. (A4) is missing a factor of N.

[2]In the CAS measure, a confident response is given a weight of 1, the same weight as a correct response. The P_c measure corresponds to a confidence weight of zero. Goupell's thesis (2005) studied the effect of varying the weight and showed that interlistener correlation is insensitive to the value of the weight in the vicinity of a weight of 1. Further, the results of significance tests do not change as the weight increases past 1. This insensitivity tends to validate the CAS technique as applied here.

Barron, M. (**1983**). "Objective measures of spatial impression in concert halls," Proceedings of the Sixth International Congress on Acoustics, Paris, Vol. **7**, pp. 105–108.
Barron, M. (**2004**). "The current status of spatial impression in concert halls," Proceedings of the 18th International Congress on Acoustics, Kyoto, Th2.B1.1, Vol. **IV**, pp. 2449–2452.
Barron, M. and Marshall, A. H. (**1981**). "Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure," J. Sound Vib. **77**, 211–232.
Beranek, L. L. (**2004**). *Concert Halls and Opera Houses—Music, Acoustics, and Architecture* (Springer, New York).
Bernstein, L. R., Trahiotis, C., and Hyde, E. L. (**1998**). "Inter-individual differences in binaural detection of low-frequency tonal signals masked by narrowband or broad-band noise," J. Acoust. Soc. Am. **103**, 2069–2078.
Blauert, J., and Lindemann, W. (**1986**). "Auditory spaciousness: Some further psychoacoustic analyses," J. Acoust. Soc. Am. **80**, 533–542.
Bradley, J. S., and Soulodre, G. A. (**1995**). "Objective measures of listener envelopment," J. Acoust. Soc. Am. **98**, 2590–2597.
Breebaart, J., and Kohlrausch, A. (**2001**). "The influence of interaural stimulus uncertainty on binaural signal detection," J. Acoust. Soc. Am. **109**, 331–345.
Breebaart, J., van de Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.
Culling, J. F., Colburn, H. S., and Spurchise, M. (**2001**). "Interaural correlation sensitivity," J. Acoust. Soc. Am. **110**, 1020–1028.
Domnitz, R. H., and Colburn, H. S. (**1976**). "Analysis of binaural detection models for dependence on interaural target parameters," J. Acoust. Soc. Am. **59**, 598–601.
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.
Gabriel, K. J., and Colburn, H. S. (**1981**). "Interaural correlation discrimination. I. Bandwidth and level dependence," J. Acoust. Soc. Am. **69**, 1394–1401.
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.
Goupell, M. J. (**2005**). "The use of interaural parameters during incoherence detection in reproducible noise," Ph.D. dissertation, Michigan State University.
Green, D. M. (**1966**). "Interaural phase effects in the masking of signals of different durations," J. Acoust. Soc. Am. **39**, 720–724.
Hall, J. W., Grose, J. H., and Hartmann, W. M. (**1998**). "The masking level difference in low-noise noise," J. Acoust. Soc. Am. **103**, 2573–2577.
Isabelle, S. K., and Colburn, H. S. (**1991**). "Detection of tones in reproducible narrow-band noise," J. Acoust. Soc. Am. **89**, 352–359.
Isabelle, S. K., and Colburn, H. S. (**2003**). "Binaural detection of tones masked by reproducible noise: Experiments and models," J. Acoust. Soc. Am. (submitted).
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. **41**, 35–49.
Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). "Performance in several binaural-interaction experiments," J. Acoust. Soc. Am. **79**, 1558–1562.
Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.
Pollack, I., and Trittipoe, W. J. (**1959a**). "Binaural listening and interaural noise cross correlation," J. Acoust. Soc. Am. **31**, 1250–1252.
Pollack, I., and Trittipoe, W. J. (**1959b**). "Internal noise correlation: Examination of variables," J. Acoust. Soc. Am. **31**, 1616–1618.
van de Par, S., and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," J. Acoust. Soc. Am. **106**, 1940–1947.
Wilbanks, W. A. (**1971**). "Detection of a narrow-band noise as a function of the interaural correlation of both signal and masker," J. Acoust. Soc. Am. **49**, 1814–1817.
Wilbanks, W. A., and Whitmore, J. K. (**1967**). "Detection of monaural signals as a function of interaural noise correlation and signal frequency," J. Acoust. Soc. Am. **43**, 785–797.
Zurek, P. M., and Durlach, N. I. (**1987**). "Masker-bandwidth dependence in homophasic and antiphasic tone detection," J. Acoust. Soc. Am. **81**, 459–464.

Simulation of vocal fold impact pressures with a self-oscillating finite-element model

Chao Tao, Jack J. Jiang,[a] and Yu Zhang
Department of Surgery, Division of Otolaryngology Head and Neck Surgery, University of Wisconsin Medical School, Madison, Wisconsin 53792-7375

(Received 8 July 2005; revised 20 March 2006; accepted 27 March 2006)

Vocal fold impact pressures were studied using a self-oscillating finite-element model capable of simulating vocal fold vibration and airflow. The calculated airflow pressure is applied on the vocal fold as the driving force. The airflow region is then adjusted according to the calculated vocal fold displacement. The interaction between airflow and the vocal folds produces a self-oscillating solution. Lung pressures between 0.2 and 2.5 kPa were used to drive this self-oscillating model. The spatial distribution of the impact pressure was studied. Studies revealed that the tissue collision during phonation produces a very large impact pressure which correlates with the lung pressure and glottal width. Larger lung pressure and a narrower glottal width increase the impact pressure. The impact pressure was found to be roughly the square root of lung pressure. In the inferior-superior direction, the maximum impact pressure is related to the narrowest glottis. In the anterior-posteriorfirection, the greatest impact pressure appears at the midpoint of the vocal fold. The match between our numerical simulations and clinical observations suggests that this self-oscillating finite-element model might be valuable for predicting mechanical trauma of the vocal folds. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2197798]

PACS number(s): 43.70.Aj, 43.70.Bk, 43.70.Jt [BHS] Pages: 3987–3994

I. INTRODUCTION

The studies of vocal fold biomechanics not only give an insight into voice production but also provide important information about laryngeal pathology development.[1–3] A previous study has revealed that abnormal stress distribution in the vocal folds is one reason for tissue damage.[4] Furthermore, the contact force produced by vocal fold collision increases the risk of tissue damage.[3,5,6] Therefore, the study of vocal fold collision is important for understanding the mechanism of phonation and determining potential vocal fold pathological development. Many experimental and theoretical works have been done to study the collision dynamics of the vocal folds. Jiang and Titze,[5] Scherer and Titze,[8] and Hess *et al.*[7] have experimentally measured the vocal fold impact stress of collision during phonation. Theoretical works have also been done to reveal the dynamics of vocal fold oscillation. Lumped mass models have been broadly used to study speech synthesis,[9] simulate the body-cover interaction in vocal folds,[10] predict vocal fold chaotic vibration,[11,12] and model vocal fold nodules.[13] Some simulation results have been proven to fit the measurement results. In these lumped mass models, vocal fold impact was described by activating a stiff spring when a mass crosses the glottal midline.[9–13]

Recently, finite-element models (FEM) have been employed to study vocal fold vibrations.[4,6,14–20] FEMs are numerical methods used to solve governing equations of continuum mechanics. A FEM has been developed to predict the pathological conditions associated with the stress distribution in a vocal fold with nodules.[4] The FEM has also been used to simulate the vocal fold bulging, and it was found that vocal fold bulging increases glottal flow resistance but decreases the mean glottal volume velocity.[17] Chaotic vibration was discovered in the FEM with the low transverse Young's modulus of the vocal fold cover.[19] Using the FEM that includes false vocal folds and laryngeal ventricles, it was found that these laryngeal structures modify the pressure distribution along the glottal walls.[20] The impact stress was simulated by the FEM with a high spatial and temporal resolution.[6] A detailed mathematical description of vocal fold FEMs can be found in Refs. 14–16 and 18. As the size of the elements becomes smaller, the solution of the FEM approaches the actual vocal fold vibration. The advantage of the FEM is the ability to deal with complex boundaries and provide more spatial information about the object than lumped mass models. However, some previous studies of FEM just considered the tissue mechanics with quasistatic boundary condition, where the preset period pressure or static pressure was applied on the vocal fold instead of the complex fluid-solid interaction between air flow and the vocal fold.[4,6] Vocal fold vibration is actually a self-oscillating dynamic process combining tissue elasticity and aerodynamics. Three main forces cause the closure of the vocal folds: (1) elastic recoil of the deformed tissue; (2) decreased intraglottal pressure due to the high velocity of the air stream; and (3) decreased subglottal pressure due to air escaping from the subglottal reservoir.[6,21] Therefore, a self-oscillating model might better describe real phonation conditions[17,18] and real collision processes.

In this paper, a self-oscillating finite-element vocal fold model will be presented to study vocal fold impact under natural phonation conditions. This model is comprised of a FEM depicting the vocal folds and a FEM describing airflow.

[a]Electronic mail: Jiang@surgery.wise.edu

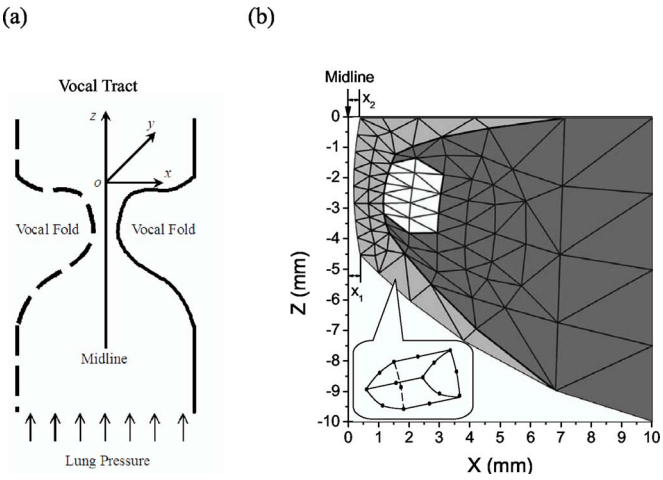


FIG. 1. The self-oscillating vocal fold model. (a) The schematic diagram of the self-oscillating vocal fold model. The dashed contour denotes that the left part is the mirror image of the right part because the vocal fold and airflow is assumed to be symmetric about the midline. The solid contour is the computational domain. (b) One cross section of the finite-element model of the vocal fold, where the light gray region represents the cover, the dark gray region represents the vocal fold body, and the white region represents the ligament. x_1 is the inferior, glottal half width, and x_2 is the superior glottal half width. The callout located at the left lower corner presents the structure of a 15 node, 3D high order prism-shaped element.

Based on this model, the vocal fold impact process was studied. The impact pressures during collision were examined. The effect of lung pressure on impact pressure and vibratory amplitude were investigated. Furthermore, the spatial characteristics of the vocal folds were studied and the spatial distribution of impact pressure is given. The effect of medial surface shape on the impact pressure was also studied. This paper is organized as follows: In Sec. II, the self-oscillating finite-element model is introduced; in Sec. III; the model is used to study the collision dynamics of vocal folds; in Sec. IV, the conclusions are drawn.

II. THE SELF-OSCILLATING FINITE-ELEMENT MODEL

Our self-oscillating model contained three parts: (1) an airflow region was used to simulate the airflow in the glottis and drive the vocal fold; (2) one rigid plane was defined in the midline ($x=0$) to simulate the interaction between two vocal folds[6,22] (3) a three dimensional (3D) vocal model was used to simulate the vibration of the vocal fold. The schematic diagram of our self-oscillating model is given in Fig. 1(a). When solving the oscillation of the vocal fold, the airflow model and the vocal fold model were alternately solved. In step 1, airflow was calculated, and the calculated airflow pressure was applied as the driving force on the surface of the vocal fold through a fluid-solid interaction surface. In step 2, the movement of the vocal fold under the driving force of the airflow was solved. The airflow region was then adjusted according to the calculated vocal fold displacement. The finite-element analysis software ANSYS was used to solve the airflow-vocal fold interaction by repeating the above two steps until a convergence of solution was reached in each time step. The dynamic solution of the vocal folds was obtained.

In this self-oscillating finite-element model, the airflow was assumed to be a single phase incompressible Newtonian laminar fluid governed by the laws of conservation of mass and momentum. These laws are expressed in terms of partial differential equations that are made discrete by using a finite-element-based technique. The fluid properties, such as density and viscosity, were set as air properties with a temperature of 37 °C. In order to reduce the model complexity and computational time, the acoustic wave propagation was not included in the airflow model. Moreover, as in previous works,[23,24] the airflow was assumed to be symmetric about the midline ($x=0$) in this model. The symmetry condition was applied on the midline plane. That is, the velocity component normal to the rigid plane was set to zero and all other degrees of freedom were unspecified. Moving wall condition was applied on the fluid-solid interaction boundary. The flow velocity on the wall surface was specified as the velocity of a vocal fold surface. During vocal fold collision, the airflow nodes at the collision point were uncoupled with the vocal fold surface and fixed at the midline. The flow velocity at the collision point was specified as zero, i.e., the airflow was interrupted at the collision point. One 1-cm subglottal tube and one 2-cm supraglottal tube were attached to simulate the upstream flow and the downstream flow. The model was driven by pressure input at the inlet of the subglottal tube, where the pressure is the lung pressure. A zero pressure boundary condition was set at the outlet of the supraglottal tube.

Because the contact of symmetrical vocal folds occurred at the midline between two vocal folds, a rigid plane was defined to investigate the impact effect at the midline. Interaction between the vocal fold and a rigid surface in the middle of the glottis represented the interaction between one vocal fold and the opposing vocal fold.[6] When the surface nodes of vocal fold cross the glottal midline, impact occurred and a contact pressure was applied on the vocal fold surface nodes to prevent them from further penetrating through the midline. The contact pressure was normalized to the rigid surface and its value was increased with the penetration distance. The contact pressure was calculated by using the augmented Lagrangian method.[22]

A 3D finite element model was present to describe the vibration of the vocal folds. In the anterior-posterior direction, the vocal fold is symmetric. Each cross-sectional plane of vocal fold has the same geometry and tissue properties. Figure 1(b) shows one cross section of the finite element model of the vocal fold. The model is defined in Cartesian coordinates. The x dimension defines the lateral distance from the glottal midline (the plane $x=0$), the y dimension (the direction normal to the paper surface) defines the anterior-posterior distance, and the z dimension represents vocal fold movement in the vertical direction. x_1 and x_2 are the inferior and superior glottal half widths, respectively. In this finite-element vocal fold model, the vocal fold dimensions are 1.0 cm in depth (x direction), 1.6 cm long (y direction), and 0.45 cm in thickness (z direction). The inferior and superior glottal half widths were set $x_1=x_2=0.3$ mm in the following numerical calculations. In Fig. 1(b), the light gray region represents the cover, the dark gray region represents

TABLE I. The input parameters used in the finite-element model of the vocal fold.

Transverse Young's modulus of the body	4 kPa
Transverse Young's modulus of the cover and ligament	2 kPa
Longitudinal shear modulus of the body	12 kPa
Longitudinal shear modulus of the cover	10 kPa
Longitudinal shear modulus of the ligament	40 kPa
Viscosity of the body, cover, and ligament	0.5 Pa s

the vocal-fold body, and the white region represents the ligament. A linear elastic model with linear damping was used to model the vocal fold tissue and the linear (infinitesimal) deformation was considered. The input tissue parameters used in the finite-element model of the vocal fold are given in Table I.[18,19]

The above 3D vocal fold model was meshed by a second-order prism-shaped element which is defined by 15 nodes. This prism-shaped element can tolerate irregular shapes without much loss of accuracy. It has compatible displacement shapes and is well suited to model curved boundaries. Figure 1 gives the geometry of the prism-shaped element at the left-lower corner. In the anterior-posterior direction, the 3D geometry of the vocal fold was divided into 16 thin layers in order to simplify the model without sacrificing the accuracy. Because experimental observations showed that the vibration of vocal fold takes place in a single plane only, the movement in each layer can be reduced to lateral and vertical motions.[18] For each layer, the geometry and the mechanical tissue properties are the same. Each layer contains 142 prism-shaped finite elements [see Fig. 1(b)]. The whole vocal fold finite-element model contains 2272 elements with 6547 nodes.

The finite-element model, including the airflow region, contact surface, and vocal fold, was built and numerically solved by using the finite-element software ANSYS. In the time domain, the model was integrated with a step length of 50 μs.[18]

III. SIMULATING VOCAL IMPACT

A. Code validation

In this section, before the above self-oscillating model is used to study the impact pressure of the vocal folds, the airflow pressure and the nodal displacement of the vocal folds are given as the basic output of the self-oscillating finite-element model. Figure 2 gives some typical wave forms simulated by the finite-element model with a lung pressure of 1.0 kPa. Curve (a) corresponds to the subglottal pressure (P_s), curve (b) corresponds to the intraglottal airflow pressure (P_{air}), and curve (c) corresponds to the glottal width (G_w) between two vocal folds. The intraglottal airflow pressure P_{air} is the airflow pressure recorded at the node located at $x=0$ mm, $z=-1.68$ mm. From the results presented in Fig. 2, it can be seen that G_w is an approximate triangular wave form. The closing stage ($G_w=0$) of the glottal cycle is due to the impact of the vocal folds. During the closing stage of the glottal cycle, the vocal folds make contact at the midpoint of the glottis thus cutting off the airflow.

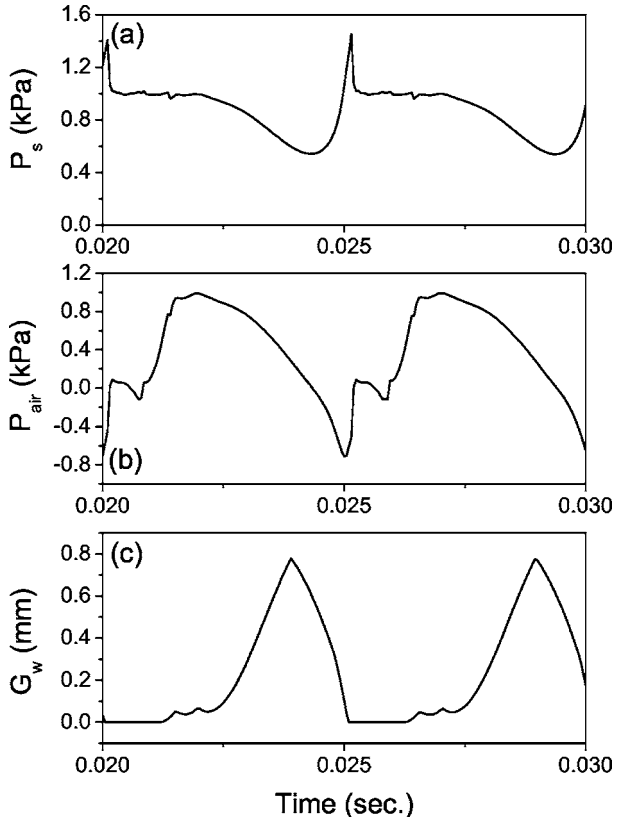


FIG. 2. Glottal wave forms of simulation at a lung pressure of 1.0 kPa. (a) The wave form of the subglottal pressure (P_s). (b) The wave form of the intraglottal air pressure (P_{air}). (c) The wave form of the glottal width (G_w) between the two vocal folds.

The subglottal pressure (P_s) then reaches its maximum value because of vocal fold collision and pushes the vocal folds apart. After the glottis is opened, air escapes from the subglottis, and the subglottal pressure decreases. At the same time, the intraglottal airflow pressure (P_{air}) is decreased due to Bernoulli's effect as air flows at a high velocity through the glottis. The two above factors provide an aerodynamic driving force. This aerodynamic force and the elastic recoil of the deformed tissue close the vocal fold again, and a new cycle is begun. The oscillating of the vocal fold system is sustained. All these wave forms are reasonable when compared with experimental observations and other computer models.[18,25]

Figure 3 gives the surface nodal displacement wave forms at the midpoint ($y=0$ mm) of the vocal fold, where curves (a), (b), and (c) correspond to the position $z=0$ mm, $z=-1.68$ mm, and $z=-3.38$ mm, respectively. Near the superior vocal fold [Figs. 3(a) and 3(b)], the combined effect of the tissue's deformation and airflow pressure generates a higher amplitude vibration than that at $z=-3.38$ mm [Fig. 3(c)]. The displacement wave form of the node at $z=-1.68$ mm [Fig. 3(b)] is sinusoidal-like in shape, but near the midline ($x=0$ mm), the wave form is deformed and the symmetry of the wave form is destroyed because the strong impact compresses the vocal folds. Far from the superior vocal fold [Fig. 3(c)], the collision does not occur and the nodal displacement wave form is symmetrical. Moreover, the phase of the nodal displacement wave form near the interior

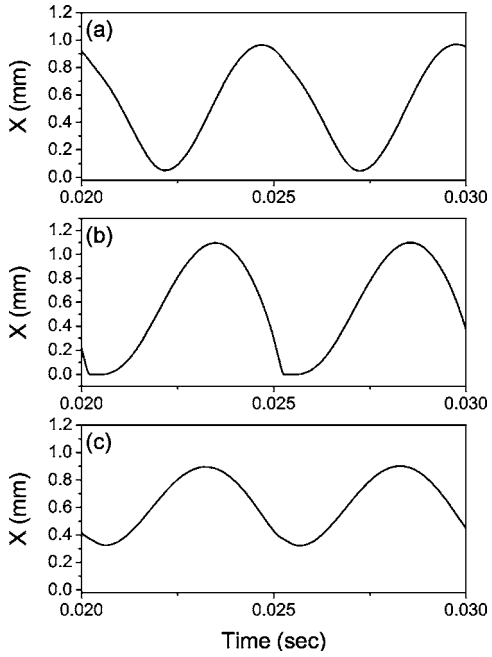


FIG. 3. The surface nodal displacement wave forms at the midpoint ($y=0$ mm) of the vocal fold, where (a), (b), and (c) correspond to nodes $z=0$ mm, $z=-1.68$ mm, and $z=-3.38$ mm, respectively.

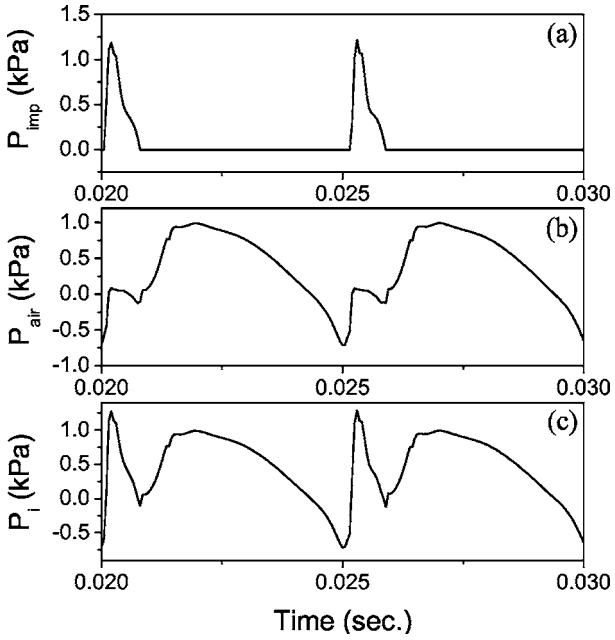


FIG. 4. Impact pressure, intraglottal airflow pressure, and intraglottal pressure as functions of time. The lung pressure is 1.0 kPa. (a) The wave form of impact pressure (P_{imp}). (b) The wave form of air flow pressure (P_{air}). (c) The wave form of intraglottal pressure (P_i), where the intraglottal pressure is the sum of the impact pressure and air flow pressure, i.e., $P_i=P_{\text{imp}}+P_{\text{air}}$.

position is in advance of that near the superior position because the mucosal wave spreads along the medial surface of the vocal fold mucosa.[21] These wave forms agree with previous experimental observations.[27]

Collision kinematics is also a useful output of our FEM. When collision occurs, it is thought that the contact force between two vocal folds may cause tissue damage.[29–33] Therefore, the study of vocal fold collision may provide information concerning laryngeal pathologies such as vocal nodules and polyps.[5] Figure 4 shows impact pressure (P_{imp}), intraglottal airflow pressure (P_{air}), and intraglottal pressure (P_i) as functions of time where the lung pressure is $P_L=1.0$ kPa. The intraglottal pressure is a combined effect of the impact pressure and the intraglottal airflow pressure. Thus intraglottal pressure can be expressed as the sum of the above factors, that is, $P_i=P_{\text{imp}}+P_{\text{air}}$. Figure 4(a) shows that the impact pressure appears at the moment that the wave form of the nodal displacement is compressed, and that the peak value of impact pressure exceeds 1.2 kPa in this model. As a comparison, Fig. 5 gives a typical intraglottal pressure wave form measured in previous experiments.[5] It can be seen that the intraglottal pressure P_i predicted by the FEM agrees with excised larynx experimental observations very well. Therefore, this FEM can effectively study the vocal fold vibration and collision.

B. Lung pressure

Lung pressure affects phonation production and vocal fold vibration; therefore, the effect of lung pressure on collision dynamics should be investigated. Sixteen lung pressures between 0.2 and 2.5 kPa were used to drive the self-oscillating model. Figure 6(a) gives the relationship between the vibration amplitude and lung pressure. Figure 6(b) presents the relationship between the peak value of impact pressure and the lung pressure. The impact pressure on the fourth element from the top left edge is given. The circled labels show the results predicted by our FEM. It is seen that vibration amplitude and impact pressure depend on lung pressure. In the beginning, the vocal fold does not vibrate when lung pressure is low ($P_L<0.29$ kPa). Then when the lung pressure reaches its phonation threshold pressure (PTP, PTP $\approx$0.29 kPa in this model), Hopf bifurcation occurs and the vocal fold begins vibrating. However, collision does not occur in this condition because the vibration amplitude is very, small. When lung pressure reaches 0.36 kPa, collision happens. The lowest lung pressure causing vocal fold collision is defined as the impact threshold pressure (ITP). In this FEM, ITP is estimated as 0.36 kPa, which is a little higher than PTP$\approx$0.29 kPa.

Using a small-amplitude approximation, Titze found the vibration amplitude of the vocal folds to be roughly the square root of the lung pressure.[26] Jiang and Titze predicted

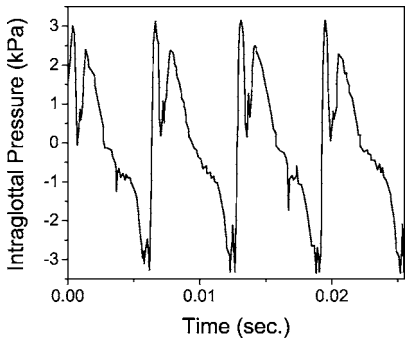


FIG. 5. The intraglottal pressure wave form measured in experiments.5 This result comes from Fig. 9 in the work of Jiang and Titze, J. Voice **8**, 132–144 (1994).

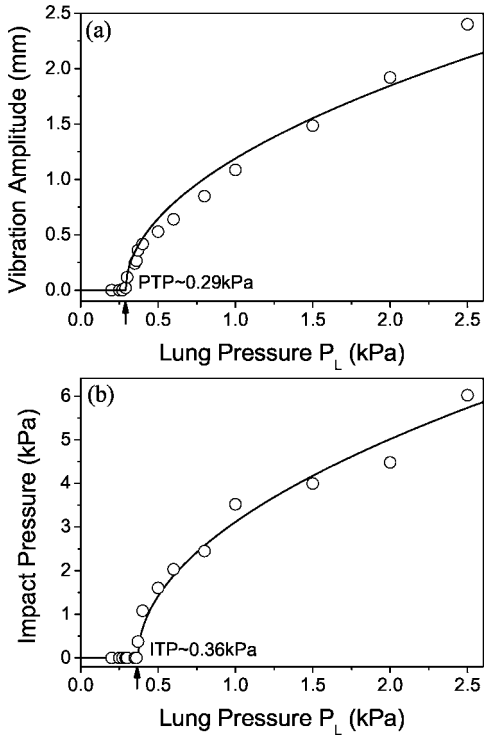


FIG. 6. The influence of lung pressure on the vibration amplitude and the impact pressure. (a) The vibration amplitude as a function of lung pressure. The phonation threshold pressure (PTP) of this model is approximately 0.29 kPa. (b) The impact pressure as a function of lung pressure. The impact threshold pressure (ITP) of this model is approximately 0.36 kPa.

that the relation between vibration amplitude and impact pressure is linear.[5] Based on these two studies, the impact pressure is expected to be the square root of lung pressure. A square root curve is therefore used to fit the relationship between impact pressure and lung pressure from FEM results

$$A = \alpha(P_L - P_{\mathrm{PTP}})^{1/2}, \tag{1}$$

$$P_{\mathrm{imp}} = \beta(P_L - P_{\mathrm{ITP}})^{1/2}, \tag{2}$$

where A is vibration amplitude, P_{PTP} is phonation threshold pressure, P_{ITP} is impact threshold pressure, α and β are two constants which are determined from FEM results using the least-square approximation. Figure 6 presents the fitted results as solid curves. It is shown that Eqs. (1) and (2) roughly approximate the FEM simulation results. However, it is noticeable that the square root relationship is fitted best when lung pressure is close to P_{PTP} and P_{ITP}. This is because the small-amplitude theory is most effective when lung pressure is low. When lung pressure is high, the error of small-amplitude approximation increases.

C. Spatial distribution

In addition to a temporal curve and the peak value of impact pressures, the finite-element model can also provide a spatial distribution of impact pressures. This is an advantage of the finite-element model which is lacking in a lumped-mass model. Figures 7–9 give the spatial distribution of impact pressure.

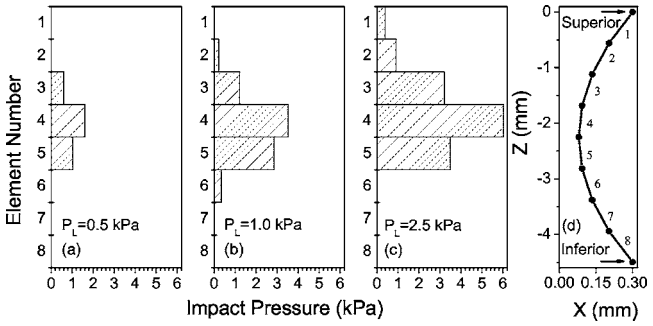


FIG. 7. The impact pressure distribution on the superior-inferior glottal direction (z direction) with varying lung pressure (P_L): (a) P_L=0.5 kPa. (b) P_L=1.0 kPa. (c) P_L=2.5 kPa. (d) Medial surface shape of vocal fold. The surface is divided into eight elements and these elements are numbered 1 to 8 from superior glottis to the inferior glottis. (a)–(c) present the impact pressure of each element.

Figure 7 shows the impact pressure distribution in the superior-inferior glottal direction (z direction), where (d) is the medial surface shape of the vocal fold. The surface is divided into eight elements and these elements are numbered 1 to 8 from the superior glottal position to the inferior glottal position. (a), (b), and (c) correspond to the lung pressures of 0.5, 1.0, and 2.5 kPa, respectively. The y axis in Figs. 7(a)–7(c) represents the element index shown in Fig. 7(d), and the heights of columns in Figs. 7(a)–7(c) are the impact pressure on each element. Because higher lung pressure will produce a stronger aerodynamic driving force, a higher lung pressure not only generates a stronger impact pressure, but also expands the impact region in the inferior-superior direction, as shown in Fig. 7.

The prephonatory glottal shape is another important factor that could influence the impact pressure distribution. Fig-

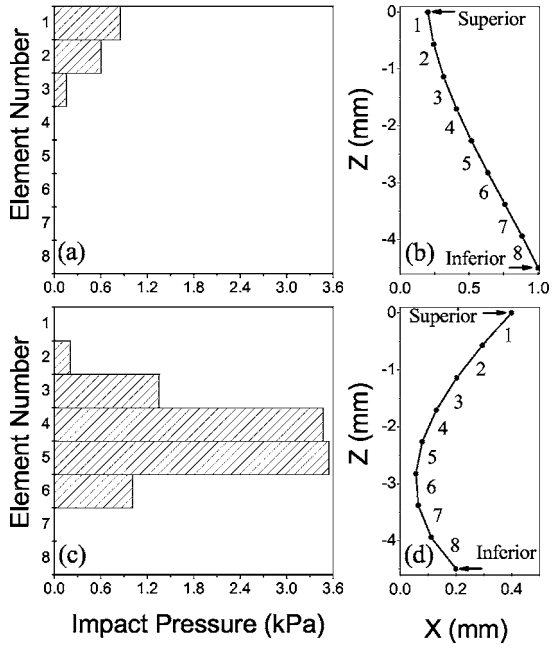


FIG. 8. The effect of the medial surface shape on the impact pressure distribution in the superior-inferior direction (z direction), where the lung pressure is 1.0 kPa. (a) A typical impact pressure distribution of a convergent glottis shown in (b). (c) A typical impact pressure distribution of a divergent glottis shown in (d).

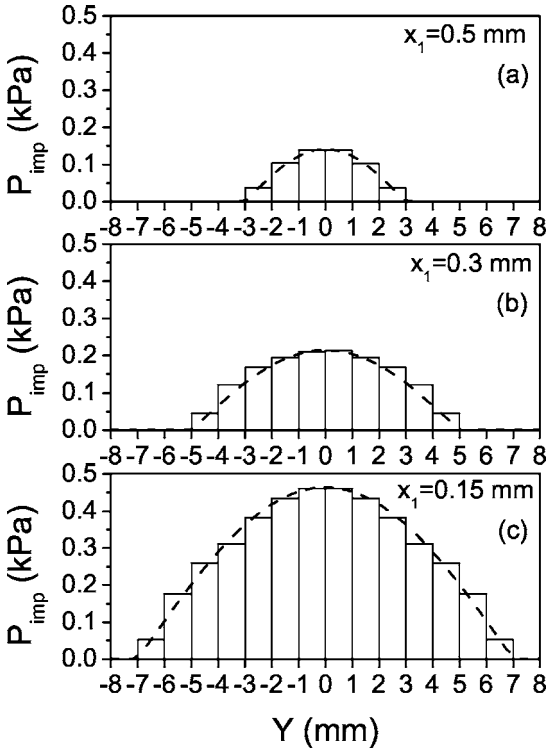


FIG. 9. The impact pressure distribution in the anterior-posterior direction (y direction) with a lung pressure of 1.0 kPa, where P_{imp} is detected from element 2 of each layer. The columns represent the results of the finite-element model and the dashed lines represent the theoretical results predicted by Eq. (2). (a) The prephonatory glottal half width is 0.5 mm. (b) The prephonatory glottal half width is 0.3 mm. (c) The prephonatory glottal half width is 0.15 mm.

ure 8 presents the impact pressure distributions with two different glottal surface shapes but with the same lung pressure $P_L=1.0$ kPa, where (a) and (c) are impact pressure distributions of a convergent glottis and a divergent glottis. Their corresponding surface shape is drawn in (b) and (d), respectively. It is seen that the distribution of impact pressure depends on the prephonatory glottal shape. The maximum value of impact pressure appears at element 5 for a divergent glottis but at element 1 for a convergent glottis. The position of strongest impact is usually associated with the narrowest glottis.

The pressure distribution in the anterior-posterior direction also strongly depends on the prephonatory glottal width. Titze and Jiang's work[5,26] demonstrated that the collision pressure resembles a half-sinusoid distribution in the anterior-posterior direction

$$P_{\text{imp}}(y) = P_m \sin\left(\frac{\pi y}{L} + \frac{\pi}{2}\right) \quad |y| < \frac{L}{2}, \tag{3}$$

where L is the vocal fold length, and P_m is the impact pressure at $y=0$ mm. Equation 3 is based on the assumption that there is no gap between two vocal folds, and the beginning of impact is at the moment of peak velocity. However, there usually exists a prephonatory glottal width between the two vocal folds. It can be expected that when the prephonatory glottal width is wide, impact can only occur at the region where the tissues have enough vibration amplitude. Figure 9 gives the impact pressure distribution in the anterior-posterior direction (y direction) with the lung pressure $P_L=1.0$ kPa, where P_{imp} was detected from element 2 [see Fig. 7(d)] of each layer. It is seen that the impact is significant near the midpoint of the vocal fold in the y direction. The maximum value of impact pressure appears near the midpoint. The impact pressure is zero near the two ends of the vocal folds (the anterior position and posterior position $y=\pm 8.0$ mm) because the anterior and posterior boundaries are held fixed. Figures 9(a)–9(c) correspond to prephonatory glottal half widths $x_2=0.5, 0.3$, and 0.15 mm, respectively. An equation is improved from Eq. (3) to fit the FEM results while considering the existence of a prephonatory glottal gap

$$P_{\text{imp}}(y) = \begin{cases} P_m \sin\left(\dfrac{\pi y}{L'} + \dfrac{\pi}{2}\right), & |y| \leqslant \dfrac{L'}{2} \\ 0 & \dfrac{L}{2} \geqslant |y| > \dfrac{L'}{2} \end{cases}, \tag{4}$$

where L' is the length of the vocal fold involved in impact. It can be expected that both the length L' and P_m in Eq. (4) depend on the size of the prephonatory glottal width. L' and P_m increase with decreasing glottal width. For the sake of comparison, the results fitted by Eq. 4 are also given with dashed lines in Fig. 9. It is seen that the FEM results can be fitted by Eq. (4) very well. The length of L' is related to the prephonatory glottal width. When prephonatory glottal width is decreased to 0.15 mm, the vocal folds almost collide with each other along the whole length of the vocal fold. Furthermore, the impact pressure roughly takes a half-sinusoid distribution which roughly fits the theoretical prediction of Eq. (3). However, when the glottal half width is increased to 0.5 mm, the impact is concentrated in a very limited region of the vocal fold near the midpoint. The impact pressure is much smaller than the condition with a narrow glottal width.

The simulation results in Fig. 9 verify that the length L' in Eq. (4) depends on the size of the prephonatory glottal width. L' increases with decreasing prephonatory glottal width. When the glottal width approaches zero, L' will approach the vocal fold length L, Eq. (4) will approach Eq. (3), and the impact pressure will take approximately a half-sinusoid distribution in the anterior-posterior direction. Furthermore, impact pressure is also related to glottal width. Narrower glottal width generates stronger collisions whereas wider glottal gaps decrease the contact pressure.

IV. DISCUSSION AND CONCLUSION

In this paper, vocal fold impact was studied by using a self-oscillating finite-element model. Compared with a lumped mass model, the self-oscillating finite-element model can simulate a complex boundary condition and provide more spatial information. Compared with a quasistatic finite-element model, the self-oscillating model is a more realistic approximation of normal phonation conditions.

Using the self-oscillating finite-element model, the collision dynamics of vocal folds with a normal vibration condition were studied. The subglottal pressure, the intraglottal pressure, the glottal width, and the nodal displacement on the

vocal fold surface were given first. Simulation results show that there are two peaks in each cycle of the wave form of the intraglottal pressure. One peak is due to the variation of airflow pressure, and the other is generated by the impact of vocal folds. The output of the FEM is consistent with experimental observations. The self-oscillating finite-element model can effectively represent the process of vocal fold vibration. Then the relationship between lung pressure and impact pressure was investigated. A higher lung pressure will generate higher vibration amplitude and produce stronger impact. The impact pressure is roughly the square root of lung pressure.

Furthermore, the spatial characteristics of the impact pressure distribution were studied. The impact pressure does not take a uniform distribution in the spatial domain and strong impact is only located in a limited region. In the superior-inferior direction, the impact pressure distribution strongly depends on the surface shape of the vocal fold, and the strongest impact is usually close to the narrowest glottis. In the anterior-posterior direction, it was found that the impact pressure distribution depends on the glottal width. A wide glottal width limits the impact region and decreases the impact pressure, but a narrow glottal width expands the impact region and increases the impact pressure. When the glottal width approaches zero, the impact pressure takes a half-sinusoid distribution. However, when the glottal width is wide, the impact pressure region is reduced to a local region near the midpoint of vocal fold.

Clinical observations assert that loud voice and hyperadduction may cause injurious abuse and misuse of the voice.[34] Previous studies have also shown that vocal nodules are caused by phonatory trauma, and trauma occurs after prolonged voice use with phonation types involving high impact pressure.[28–33] The nodule often occurs at the middle of membranous vocal folds.[32,33] In our simulation, the peak value of impact pressure also occurs at the midpoint of membranous vocal folds. The match between clinical observations and our simulation supports these clinical observations. The fact that the peak impact pressure occurs at the midpoint of the vocal fold may be one reason for nodule development. Furthermore, our simulation reveals that low lung pressures and wide glottal half widths can effectively decrease the peak impact pressure, which supports voice rest and soft speech as the key facts in recovering. It seems that low lung pressures and wide glottal half widths may prevent the development of vocal nodules. This also sets up a fundamental in treating vocal nodules in speech training to reduce vocal stress by the use of a breathy voice.

In this paper, the results came from a symmetric vocal fold model. The laryngeal pathologies such as paralysis, edema, and adduction/abduction could cause the vocal fold asymmetry, including the left-right asymmetry and the anterior-posterior asymmetry. Furthermore, even for a symmetric vocal fold, the turbulence and Coanda effect in the glottis could cause unequal pressure distributions between the left and right vocal fold surfaces.[35,36] Therefore, the assumption of the symmetric vocal fold can be well satisfied only when the lung pressure is not significantly high.[37] It can be expected that both the asymmetry of the vocal folds and airflow could influence the impact pressure distribution. The influence of asymmetry on the impact stress will be of interest in further investigations. In addition, polyps, nodules, and vocal fold scars could change the local tissue properties of the vocal folds and further affect the impact stress distribution. The self-oscillating model presented in this study can be extended to model the pathological vocal fold and should be considered for use in further work to study the impact stress of a pathological larynx.

ACKNOWLEDGMENTS

This study was supported by NSF of China (No. 30328029) and NIH Grant (No. 1-RO1DC006019 and No. 1-RO1DC05522) from the National Institute of Deafness and other Communication Disorders.

[1]E. B. Holmberg, R. E. Hillman, and J. S. Perkell, "Glottal air-flow and transglottal air-pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am., **84**, 511–529 (1988).
[2]S. M. Zeitels, "Phonosurgery—past, present, and future," Operative technique in Otolaryngol.-Head Neck Surg., **9**, 179 (1998).
[3]I. R. Titze, "Mechanical stress in phonation,"J. Voice, **8**, 99–105 (1994).
[4]J. J. Jiang, C. E. Diaz, and D. G. Hanson, "Finite element modeling of vocal fold vibration in normal phonation and hyper functional dysphonia: implications for the pathogenesis of vocal nodules," Ann. Otol. Rhinol. Laryngol. Suppl., **107**, 603–610 (1998).
[5]J. J. Jiang and I. R. Titze, "Measurement of Vocal Fold Intraglottal Pressure and Impact Stress," J. Voice, **8**, 132–144 (1994).
[6]H. E. Gunter, "A mechanical model of vocal-fold collision with high spatial and temporal resolution," J. Acoust. Soc. Am., **113**, 994–1000 (2003).
[7]M. M. Hess, K. Verdolini, W. Bierhals, U. Mansmann, and M. Gross, "Endolaryngeal contact pressure," J. Voice, **12**, 50–67 (1998).
[8]R. C. Scherer and I. R. Titze, "Vocal fold contact stress during phonation," J. Acoust. Soc. Am., **71** (suppl 1), S55 (1982).
[9]K. Ishizaka and J. L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J., **51**, 1233–1268 (1972).
[10]B. H. Story and I. R. Titze, "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am., **97**, 1249–1260 (1995).
[11]J. J. Jiang, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," J. Acoust. Soc. Am., **110**, 2120–2128 (2001).
[12]J. J. Jiang and Y. Zhang, "Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds," J. Acoust. Soc. Am., **112**, 2127–2133 (2002).
[13]Y. Zhang and J. J. Jiang, "Chaotic vibrations of a vocal fold model with a unilateral polyp," J. Acoust. Soc. Am., **115**, 1266–1269 (2004).
[14]F. Alipour and I. R. Titze, "Simulation of particle trajectories of vocal fold tissue," in *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, 1985), pp. 183–190.
[15]F. Alipour and I. R. Titze, "A finite element simulation of vocal fold vibration," in Proceedings of the Fourteenth Annual Northeast-Bioengineering Conference, edited by J. R. LaCourse (Durham, NH, 1988), pp. 186–189.
[16]F. Alipour and I. R. Titze, "Combined simulation of airflow and vocal fold vibrations," in *Vocal Fold Physiology, Controlling Complexity & Chaos*, edited by P. Davis and N. Fletcher (Singular Publishing Group, San Diego, 1996), pp. 17–29.
[17]F. Alipour and R. C. Scherer, "Vocal fold bulging effects on phonation using a biophysical computer model," J. Voice, **14**, 470–483 (2000).
[18]F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am., **108**, 3003–3012 (2000).
[19]D. A. Berry, H. Herzel, I. R. Titze, and K. Krischer, "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am., **95**, 3595–3604 (1994).
[20]M. de O. Rosa, J. C. Pereira, M. Grellet, and A. Alwan, "A contribution to simulating a three-dimensional larynx model using the finite-element method," J. Acoust. Soc. Am., **114**, 2893–2905 (2003).
[21]J. J. Jiang, E. Lin, and D. G. Hanson, "Voice disorders and phonosurgery,

I. Vocal fold physiology," Otolaryngol. Clin. North Am., **33**, 699–718 (2000).
[22]J. C. Simo and T. A. Laursen, "An Augmented Lagrangian Treatment of Contact Problems Involving Friction," Comput. Struct., **42**, 97–116 (1992).
[23]C.-G. Guo and R. C. Scherer, "Finite element simulation of glottal flow and pressure," J. Acoust. Soc. Am., **94**, 688–700 (1993).
[24]F. Alipour and R. C. Scherer, "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am., **116**, 1710–1719 (2004).
[25]I. R. Titze, "Parameterization of the glottal area, glottal flow, and vocal fold contact area," J. Acoust. Soc. Am., **75**, 570–580 (1984).
[26]I. R. Titze, "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am., **85**, 901–906 (1989).
[27]J. J. Jiang, C-I. B. Chang, J. R. Raviv, S. Gupta, F. M. Banzali, and D. G. Hanson, "Quantitative Study of Mucosal Wave Via Videokymography in Canine Larynges, "Laryngoscope, **110**, 1567–1573 (2000).
[28]A. Sonninen, P. H. Damste, J. Jol, and J. Fokkens, "On vocal strain," Folia Phoniatr., **24**, 321–336 (1972).
[29]O. Kleinsasser, "Pathogenesis of vocal fold polyps," Ann. Otol. Rhinol. Laryngol., **91**, 378–381 (1982).
[30]T. Gillman, "Treatise on collagen," in *Biology of Collagen*, edited by B. S. Gould (Academic Press, London, 1968), Vol. **2**, p. 334.
[31]S. D. Gray, I. R. Titze, and R. P. Lusk, "Electron microscopy of hyperphonated canine vocal folds," J. Voice, **1**, 109–115 (1987).
[32]K. Nagata, S. Kurita, S. Yasumoto, T. Maeda, H. Kawasaki, and M. Hirano, "Vocal fold polyps and nodules, a 10 year review of 1156 patients," Auris Nasus Larynx, **suppl 10**: S27–35 (1983).
[33]H. G Rubin and I. Lehrhoff, "Pathogenesis and treatment of vocal nodules," J. Speech Hear Disord., **27**, 150–161 (1962).
[34]J. C. Stemple, L. E. Glaze, and B. G. Klaben, *Clinical Voice Pathology Theory and Management*, Third Edition (Singular, San Diego, 2000).
[35]X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, Y. Auregan, "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am., **96**, 3416–3431 (1994).
[36]F. Alipour and R. C. Scherer, "Pressure and velocity profiles in a static mechanical hemilarynx model," J. Acoust. Soc. Am., **112**, 2996–3003 (2002).
[37]D. Shinwari, R. C. Scherer, K. J. DeWitt, and A. A. Afjeh, "Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees," J. Acoust. Soc. Am., **113**, 487–497 (2003).

Broadband sound generation by confined pulsating jets in a mechanical model of the human larynx

Zhaoyan Zhang[a] and Luc G. Mongeau
Ray W. Herrick Laboratories, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

(Received 18 July 2005; revised 17 March 2005; accepted 19 March 2006)

Experiments were performed to study the production of broadband sound in confined pulsating jets through orifices with a time-varying area. The goal was to better understand broadband sound generation at the human glottis during voicing. The broadband component was extracted from measured sound signals by the elimination of the periodic component through ensemble averaging. Comparisons were made between the probability density functions of the broadband sound in pulsating jets and of comparable stationary jets. The results indicate that the quasi-steady approximation may be valid for the broadband component when the turbulence is well established and the turbulence kinetic energy is comparatively large. A wavelet analysis of the broadband sound showed that random sound production was modulated at the driving frequency. Two distinct sound production peaks were observed during one cycle, presumably associated firstly with jet formation and secondly with flow deceleration during orifice closing. Most high-frequency sound was produced during the closing phase. Deviations from quasi-steady behavior were observed. As the driving frequency increased, sound production during the opening phase was reduced, possibly due to the shorter time available for turbulence to develop. These results may be useful for better quality voice synthesis. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195268]

PACS number(s): 43.70.Bk, 43.28.Ra, 43.70.Aj, 43.28.Py [AL] Pages: 3995–4005

I. INTRODUCTION

Voice production involves sound radiation from pulsating jet flows through the glottis. Air flow from the lung through the subglottal portion of the airway is modulated by quasi-periodic oscillations of the vocal folds. The geometry of the vocal folds varies over one oscillation cycle. The shape of the glottis is in turn divergent, uniform, and convergent. The flow within the glottis during phonation is generally laminar. Turbulence develops immediately downstream the glottis, as a turbulent jet plume develops within the vocal tract. Flow instabilities and large scale vortex shedding may be present, especially when the orifice adopts a divergent shape.

Various sound generating mechanisms contribute to voice production (Zhang *et al.*, 2002b). The three major sound sources are (1) a dipole source consisting of the net unsteady force exerted by the vocal folds walls on the surrounding fluid along the direction of sound propagation in the vocal tract; (2) quadrupole sources due to the components of the fluctuating Reynolds stresses along the direction of sound propagation and associated with flow turbulence within the jet downstream of the glottis in the vocal tract; and (3) a possible monopole source due to flow displaced by the vocal folds walls as they vibrate, characterized by volume changes of the tissue. The dipole source may be further subdivided into two components: (1) a periodic component caused by the flow pulsation through the glottis and (2) a random component caused by turbulent fluctuating pressures acting on the superior vocal folds walls. The periodic component of the dipole sound dominates the radiated sound field and can be accurately modeled using the quasi-steady approximation over most of the duty cycle (Zhang *et al.*, 2002b). The random parts of the dipole sound and the quadrupole sound are random and broadband in nature, and are of smaller amplitude. These components of the radiated sound, however, dominate at high frequencies and contribute to voice quality. In unvoiced sound production, these two components constitute the dominant source of sound (Howe and McGowan, 2005; Krane, 2005). A better understanding of this contribution could help develop better models for speech synthesis and assist in the diagnosis of pathological voice conditions.

The production of broadband sound from air flow through the larynx has been sparsely investigated. In early experiments, Meyer-Eppler (1953) identified a critical Reynolds number below which there was no significant turbulent sound generation, and postulated that the radiated sound pressure increased from that point on with the square of the Reynolds number. Sound production by turbulence in a rigid pipe was investigated by Davies and Ffowcs Williams (1968), who found that the sound power from large-scale turbulence scales with the sixth power of the jet centerline velocity regardless of frequency. Sound power from small-scale turbulence at low frequencies was found to increase with the sixth power of the jet centerline velocity. At high frequency (above the cut-on frequency of the first transverse acoustic mode in the tube), sound power was found to increase with the eighth power of the jet velocity, consistent with the well-known power scaling law of free subsonic jets.

[a]Present address: 31-24 Rehabilitation Center, School of Medicine, UCLA, 1000 Veteran Ave., Los Angeles, CA 90095. Electronic mail: zyzhang@ucla.edu

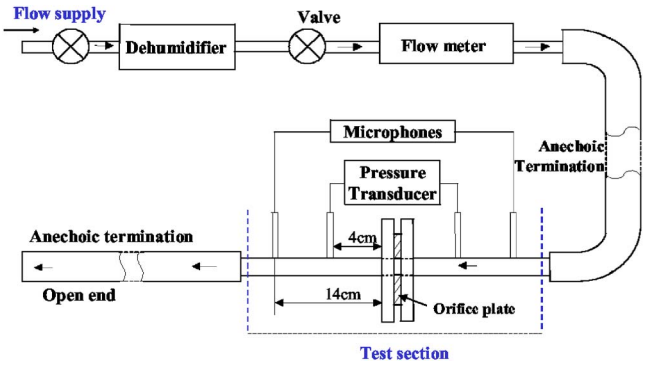


FIG. 1. (Color online) Schematic of the experimental apparatus.

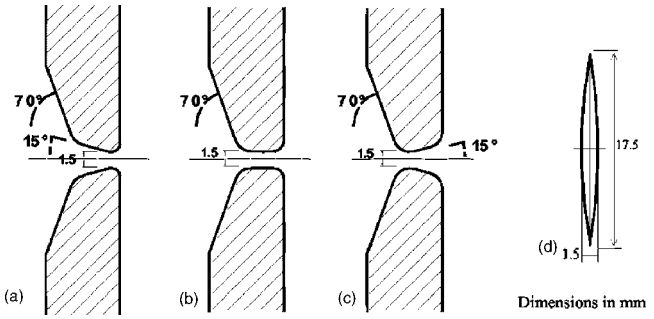


FIG. 2. Coronal cross section of the three orifice passages: (a) convergent, (b) uniform, and (c) divergent. The flow direction is from left (inferior) to right (superior). (d) Geometry of the glottis in the axial plane of minimum constriction.

Recently, sound generation by stationary jet flows through both circular and glottis-shaped fixed orifices in rigid-wall tubes was investigated by Zhang *et al.* (2002a, 2004). Sound power scaling laws that were consistent with those of Davies and Ffowcs Williams were established. The radiated sound power was found to obey a U^6A^2 scaling law at frequencies below the cut-on frequency of higher order acoustic modes in the tubes, where U is the jet centerline velocity and A is the orifice area. Both the quadrupole and dipole sources were found to contribute to sound radiation to varying degrees, depending on the orifice geometry.

Broadband sound radiation in pulsating jet flows has not been studied in detail. The quasi-steady approximation, which has been verified for the periodic component of the voiced sound (Zhang *et al.*, 2000b), may not be valid for the broadband sound component. Production of turbulence was found to be affected by flow acceleration and deceleration (Iguchi *et al.*, 1990). Dabiri and Gharib (2005) studied the transient formation of a vortex ring in a starting flow through nozzles with temporally variable exit diameters. They observed many trends in the flow development for which intrinsically transient mechanisms must be considered. Sound radiation from turbulent pulsating flows is expected to be similarly affected by flow unsteadiness. The combined effects of area modulation and dynamic pressure drops on broadband sound production by pulsating jets have not been investigated before. This question is vital to achieve high-fidelity voice synthesis. In the present study, efforts were made to verify the quasi-steady approximation for the random component of the sound radiated by confined pulsating jets. The probability density function (PDF) of the broadband sound radiated from the pulsating jets was compared to that from comparable stationary jets. Wavelet analysis was used to further investigate the time and frequency characteristics of the broadband sound.

II. EXPERIMENTAL SETUP

A schematic of the experimental facility is shown in Fig. 1. The apparatus was similar to that used in a previous study (Zhang *et al.*, 2002b). A rubber orifice plate was installed between two rectangular tubes. A generic orifice shape was designed with dimensions similar to those of the human glottis. Three orifice profiles (convergent, uniform, and divergent) were used (Fig. 2), each an idealization of the profile of the glottis along a coronal plane mid-way through the vocal folds during phonation. The rubber orifice plate was connected to, and driven by, two electro-dynamic shakers to induce oscillations. The shakers produced displacements sufficiently large to achieve complete closure at frequencies up to 150 Hz. Anechoic terminations were connected to the rectangular test tubes both upstream and downstream to minimize possible sound reflections. The design and performance of the anechoic terminations were described in detail in Zhang *et al.* (2002b). To accurately measure the orifice area, a photoelectric sensor and a light source were installed within the tube walls on opposite sides of the orifice. The photoelectric sensor signal was calibrated by establishing a regression between the output signal of the sensor and the area of the orifice for various shaker plunger positions measured using digital pictures and image processing software (Zhang *et al.*, 2002b). The calibration was repeated before and after the experiments and was verified to be stable during the course of the experiments. The regression between the light sensor output and the orifice area was used for subsequent data processing.

Sound pressures were recorded at two positions located symmetrically upstream and downstream of the orifice, 14 cm away from the orifice discharge plane. The time-averaged volumetric flow rate and the time-averaged pressure gradient across the orifice were measured. A HP356XA data acquisition system was used for the sound pressure measurements. The sampling rate was 32 768 Hz.

III. METHODOLOGY

A. Extraction of the broadband sound component

To decompose the pressure signals into broadband and periodic components, the broadband component was assumed to be random with a zero mean value. The periodic contributions were assumed to be related to the flow pulsation and deterministic. One hundred records of the sound pressure time history at one fixed observation position were recorded. Each record was 1 s long. A trigger signal was used such that each record started at the same time during the orifice oscillation cycle. The periodic component was extracted through synchronous ensemble average. One hundred

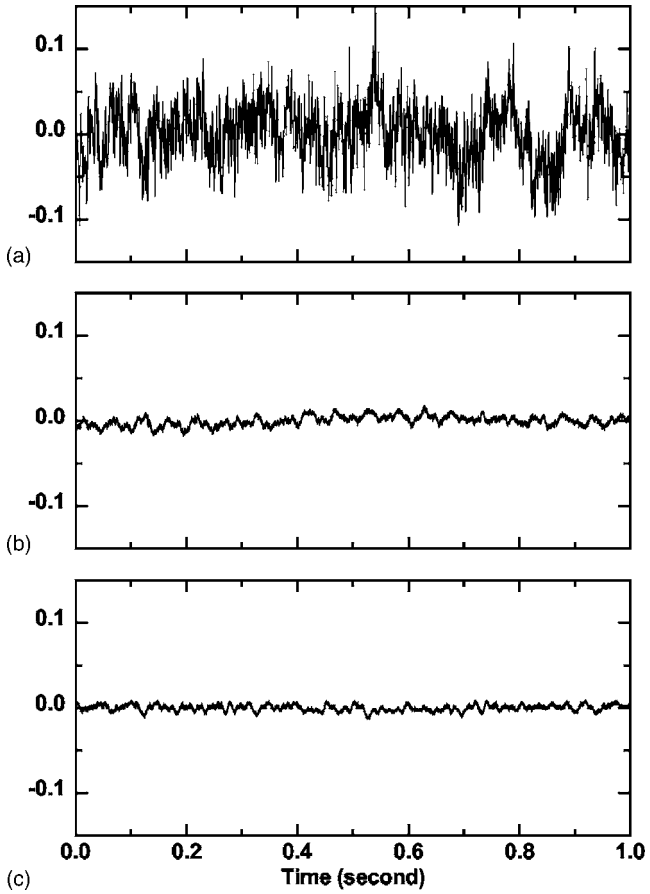


FIG. 3. Sound pressure versus time (a) with flow, periodic component removed; (b) without flow and shaker running and connected to the orifice plate; and (c) measured background noise, without flow, shaker running, but disconnected from the orifice plate. Divergent orifice, $\Delta p_0 = 12$ cm H_2O, $f = 150$ Hz.

samples of the broadband sound component were then obtained by subtracting the periodic component from each original record.

In most of the pulsating flows investigated, the magnitude of the broadband component was much smaller than that of the periodic component. Possible contamination by background mechanical noise due to vibrations of the electro-dynamic shakers was investigated. To evaluate the relative importance of mechanical and aerodynamic noise, the sound pressure associated with the operation of the shakers was measured with the flow supply turned off, and with the orifice either connected or disconnected from the shaker plungers. The results are shown in Fig. 3. Also shown in the figure is the broadband component of the sound obtained using the ensemble-average method for a divergent orifice and a mean pressure of 12 cm H_2O. The broadband sound signal was much greater in magnitude when flow was present than when there was background noise (the signal-to-noise ratio was about 20 for the case shown), which confirmed that the random component of the sound was mostly related to air flow, and not to noise from the operation of the shakers.

B. Probability density function

For a random pressure signal of a finite time interval, T, the probability density function (PDF) was estimated using (Bendat and Piersol, 1986)

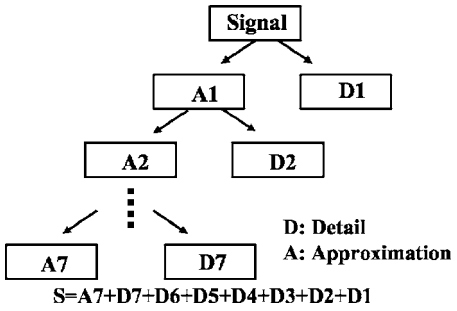


FIG. 4. Illustration of a seven-level discrete wavelet decomposition.

$$f(p) = \frac{T_p(\Delta p)}{T \cdot \Delta p}, \tag{1}$$

where T_p is the total time for which the pressure signal lies within a narrow pressure interval of width Δp centered on p. For comparisons between stationary and pulsating flows, Gaussian distribution regressions of the broadband sound PDFs were performed. The mean value and the standard deviation were estimated. The standard deviation indirectly measures the broadband sound energy. A normalized difference factor was defined to quantify the agreement between PDFs,

$$\varepsilon = \frac{|\sigma_{\text{unsteady}} - \sigma_{\text{steady}}|}{\sigma_{\text{steady}}} \times 100\,\%, \tag{2}$$

where σ_{unsteady} and σ_{steady} are the estimated standard deviations of the broadband sound for pulsating and comparable stationary flows, respectively.

C. Wavelet analysis

The turbulence-generated sound, periodically modulated by the orifice oscillation, is nonstationary in nature and has broad frequency content. Its characteristics in the frequency and time domains were investigated using a wavelet analysis. A wavelet transform produces a time-scale view of a signal, contrasting with the time-frequency view typical of short-time Fourier transforms. This suggests it might be an interesting tool for the analysis of turbulence and turbulence-produced sound, which involve various scales of fluid motions. Wavelets are irregular and transient functions. The mother wavelet is scaled so that it extends over various durations, therefore spanning a flexible window with varying resolution. Higher scales, corresponding to stretched wavelets of longer period and lower frequency, are suitable for the analysis of coarser signal features. Small scales correspond to short durations and they allow the analysis of fine signal features. The wavelet function is shifted over time to analyze local features of signals along the time axis.

The wavelet analysis in this study was performed using the Matlab Wavelet Toolbox. Discrete wavelet decomposition with scales and time positions based on powers of 2 was used. Signals were passed through two complementary filters and decomposed into a detail and an approximation in the time domain, as illustrated in Fig. 4. The two filters are determined by the specified wavelet and act like a low-pass and high-pass filter, respectively. The detail contains the small-

scale high-frequency content of the signal, while the approximation contains larger scale low-frequency information. The detail of level j, D_j, was calculated using (Mallat, 1998; Misiti *et al.*, 2005):

$$D_j(t) = \sum_k C(j,k)\psi_{j,k}(t), \tag{3}$$

where $C(j,k)$ is the discrete wavelet coefficient and $\psi_{j,k}$ is the scaled wavelet. These are defined as

$$C(j,k) = \int_R p(t)\psi_{j,k}(t)dt, \quad \psi_{j,k}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{4}$$

$$a = 2^j,\ b = k2^j,$$

where $p(t)$ is the signal of interest. Scale orders and time steps are denoted by indices j and k, respectively. The approximation of level J, A_J, contains the large-scale component of the original signal:

$$A_J = \sum_{j>J} D_j. \tag{5}$$

The approximation was then passed through the next level wavelet decomposition, using a wavelet of higher scale which is suitable for the analysis of the coarser features left in the approximation, and decomposed into a detail and a higher level approximation. The newly obtained detail and approximation contained again the high- and low-frequency components of the approximation of the previous level, respectively. This process was repeated until the details captured the entire frequency content of the signal under study (Fig. 4).

It should be noted that the two complementary filters were low-pass and high pass filters in terms of scale rather than frequency. Depending on the shape of the wavelet used, various frequency ranges could pass though the low-pass and high-pass filter. The frequency response of these filters can be evaluated by calculating the power spectra of the details and the approximations of white noise passing through the discrete wavelet decomposition. One example is shown in Fig. 5 for the Daubechies 1 (db1) wavelet (Daubechies, 1988). A white noise of normal distribution with a zero mean and unit standard deviation was used. As the scale increased, the frequency of spectral peak values associated with each detail decreased. The peak frequency values for the seven details, in the order of increasing scale, and the seventh approximation were determined from Fig. 5 to be 14 576, 6184, 3072, 1504, 764, 368, 188, and 0 Hz, respectively. For each detail or approximation, several frequency components were included in addition to the main component of energy around the peak frequency.

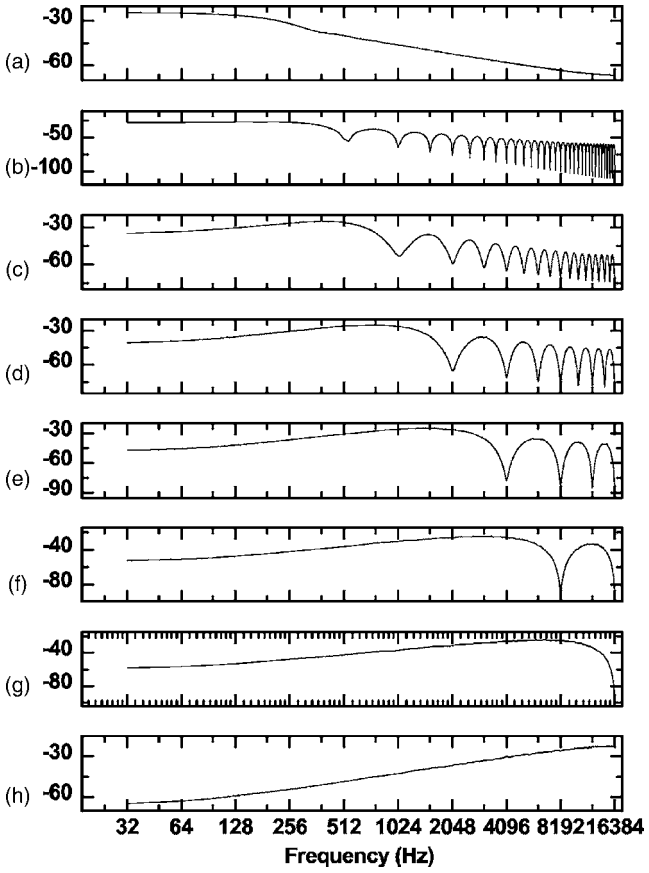


FIG. 5. Power spectra (dB) of the seven-level details and approximations for a white noise input. (a) Seventh approximation; (b) seventh detail; (c) sixth detail; (d) fifth detail; (e) fourth detail; (f) third detail; (g) second detail; and (h) first detail.

IV. RESULTS

A. Probability density functions

The probability density function of the broadband sound was calculated for selected phase values over one cycle in the case of pulsating flows. Additional measurements were then made for stationary flows and fixed orifice, for values of the orifice area and the transglottal pressure matching values recorded in the pulsating flow experiments for the selected phase values over each cycle. Comparisons were made between the sound pressure PDFs of the pulsating flows at specific times over one cycle and comparable stationary flows with the same orifice area and transglottal pressure.

Figure 6 shows the results for the model with a uniform orifice profile driven at 20 Hz. Comparisons were made at six specific instants during one cycle [Figs. 6(a)–6(f)]. These correspond respectively to the beginning of the opening stage (case a), the middle of the opening stage (case b), the end of the opening stage (case c), the beginning of the closing stage (case d), the middle of the closing stage (case e), and the end of the closing stage (case f). These instants are labeled in the corresponding orifice area time history, shown in Fig. 6(g). The normalized difference factors, ε, defined in Eq. (2) are also shown in the figure for each case. Broadband sound production in the pulsating flow case increased as the orifice opened, reached a maximum when the orifice area was nearly maximum, and decreased as the orifice closed. This is qualitatively consistent with the scaling laws for the sound power observed in Zhang *et al.* (2002a), which states that the sound power radiated from stationary jets increases approximately with U^6A^2 at low frequencies, where U is the jet centerline velocity and A is the orifice area.

The agreement between the PDFs in stationary and pulsating flows evolves over one cycle. At the beginning of the

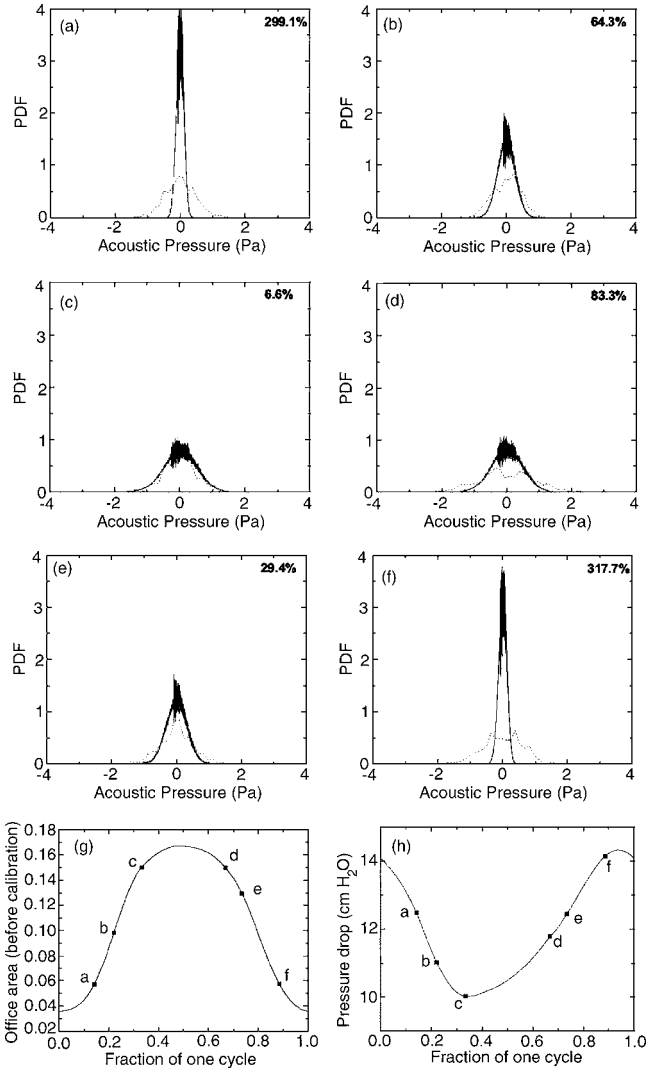


FIG. 6. Probability density function of acoustic pressure for straight orifice driven at 20 Hz. Comparison between the probability density functions of nonstationary and stationary jets was made at six different instants [(a)–(f)] during one orifice oscillation cycle as labeled in panel (g) (orifice area) and panel (h) (transglottal pressure drop).———stationary flow;– –pulsating flow. The number in the upper right corner of panels (a)–(f) quantifies the difference in PDF as defined by Eq. (2).

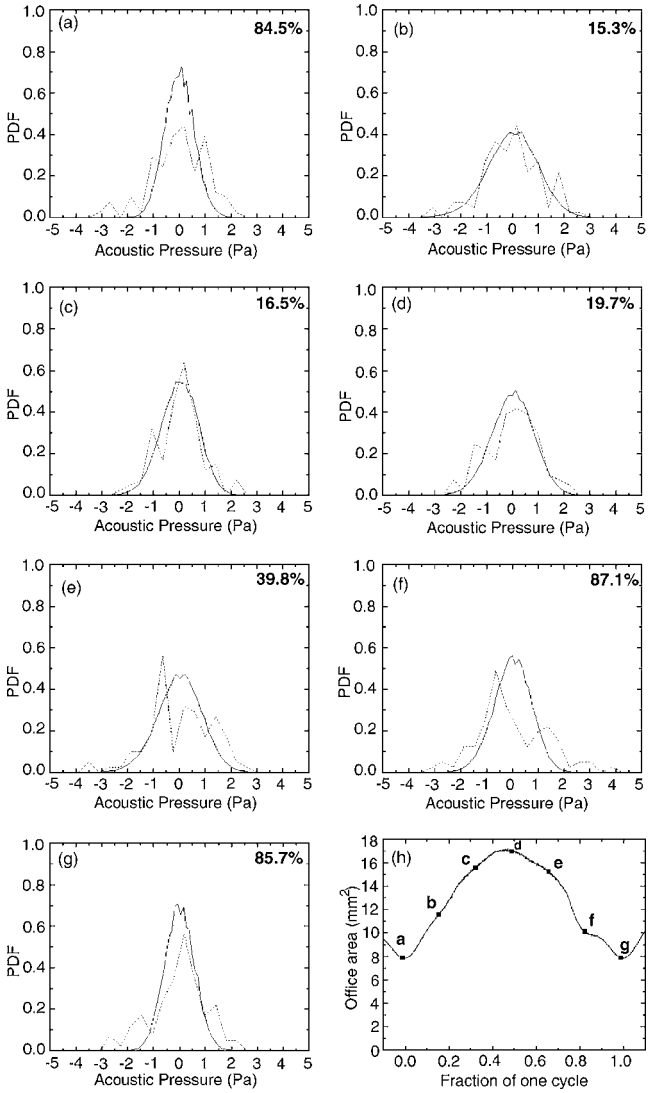


FIG. 7. Probability density function of acoustic pressure for divergent orifice driven at 20 Hz. Comparison between the probability density functions of nonstationary and stationary jets was made at seven different instants [(a)–(g)] during one orifice oscillation cycle as labeled in the orifice area function [panel (h)].———stationary flow;– –pulsating flow. The number in the upper right corner of panels (a)–(g) quantifies the difference in PDF as defined by Eq. (2).

duty cycle (time t_a), the agreement was poor. The PDF in pulsating flows spanned a larger range of acoustic pressure values (or a larger standard deviation) than that in stationary flows, indicating stronger broadband sound production during the opening phase in pulsating flows than in comparable stationary flows. This suggests that the unsteady effects played an important role in the broadband sound production at that time. The unsteadiness of flow was generally greater during the acceleration and deceleration stages than over the middle of the cycle when the flow was apparently into a quasi-stationary state. It is also possible that the effects of persisting vortical structures from the previous cycle may not have been completely dissipated before the beginning of a new cycle, resulting in more broadband sound production than that for a comparable stationary flow. The agreement between the PDFs for stationary and pulsating flows improved as the orifice area increased (at time t_b), after the flow was better established. The best agreement was obtained at time t_c, at which the orifice opening was nearly maximum and the turbulence energy was most probably nearly maximum. During the closing stage of the duty cycle, the agreement was less good (at times t_d and t_e). For time t_f, near the end of the cycle, the flow rate was small and the PDFs were in poor agreement. Note that the broadband sound amplitudes were almost always larger in pulsating flows than that in comparable stationary flows.

Figures 7–9 show similar results for a divergent orifice driven at 20, 100, and 200 Hz, respectively. Figure 10 shows the results for a convergent orifice driven at 100 Hz. Similar observations can be made for all these cases. The quasi-steady approximation was only valid during part of the oscillation cycle. The agreement between the PDF for pulsating and stationary flows was generally poor when the orifice area was nil or small. The agreement improved as the orifice area increased, and the best agreement was generally obtained

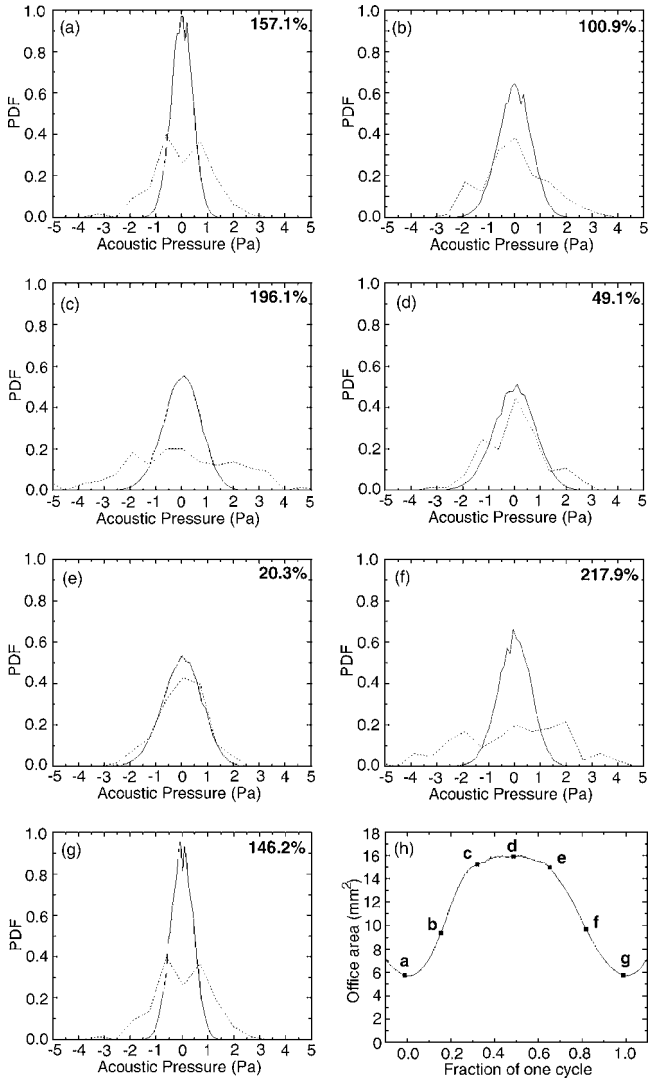


FIG. 8. Probability density function of acoustic pressure for divergent orifice driven at 100 Hz. Comparison between the probability density functions of nonstationary and stationary jets was made at seven different instants [(a)–(g)] during one orifice oscillation cycle as labeled in the orifice area function [panel (h)].———stationary flow;– –pulsating flow. The number in the upper right corner of panels (a)–(g) quantifies the difference in PDF as defined by Eq. (2).

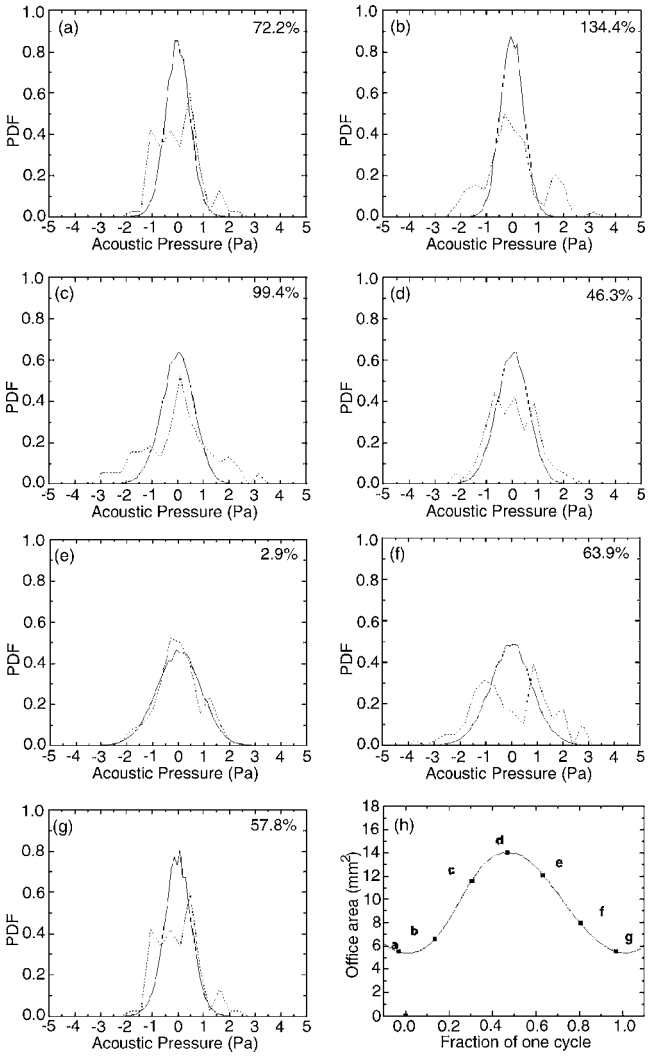


FIG. 9. Probability density function of acoustic pressure for divergent orifice driven at 200 Hz. Comparison between the probability density functions of nonstationary and stationary jets was made at seven different instants [(a)–(g)] during one orifice oscillation cycle as labeled in the orifice area function [panel (h)].———stationary flow;– –pulsating flow. The number in the upper right corner of panels (a)–(g) quantifies the difference in PDF as defined by Eq. (2).

when the orifice opening was nearly maximum. Although not as apparent as in Fig. 6, there was generally more broadband sound production in pulsating flows than that in comparable stationary flows during acceleration and deceleration, indicating the important role played by the flow unsteadiness.

The centerline jet velocity was almost constant over the duty cycle, except in the early phases of jet development and quenching (Mongeau *et al.*, 1997; Zhang, 2002). An increase in frequency increases the excitation Strouhal number and, therefore, in theory, the possible influence of unsteady effects. However, no significant and consistent influence of the driving frequency on the agreement between the PDFs of pulsating and comparable stationary flows (Figs. 7–9) was observed in the experiments. This may be because the driving frequency was not high enough to have a significant effect on the development of large-scale turbulence, which contributes the most to the total broadband sound. The driving frequency is also expected to have an impact on small-scale turbulent structures and their contribution to the broadband sound production, as discussed further in the next section.

The large fluctuations in the PDF for the pulsating jets were partially due to the limited number of averages. Another potential source of errors was a small drift in the orifice area history over time due to slight relaxation of the rubber orifice plates, which may have caused variability during the experiments. In particular, slight changes in duty cycle and collision times between the two vocal folds could have made the flow and the sound radiation vary slightly from cycle to cycle. This would lead to errors in the ensemble-averaging process. Evidence of this type of error is apparent in Fig. 8(f), in which a periodic component was present in the sound signals. This points out at the possibility that the periodic component may not have been completely removed in that particular case.

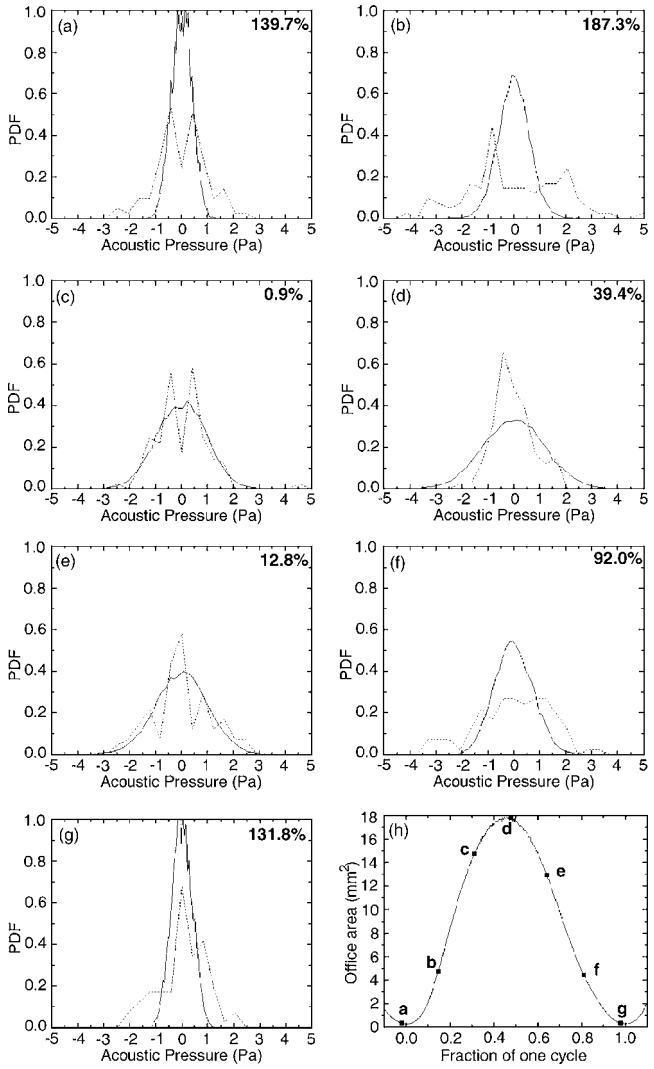


FIG. 10. Probability density function of acoustic pressure for convergent orifice driven at 100 Hz. Comparison between the probability density functions of nonstationary and stationary jets was made at seven different instants [(a)–(g)] during one orifice oscillation cycle as labeled in the orifice area function [panel (h)].———stationary flow;– –pulsating flow. The number in the upper right corner of panels (a)–(g) quantifies the difference in PDF as defined by Eq. (2).

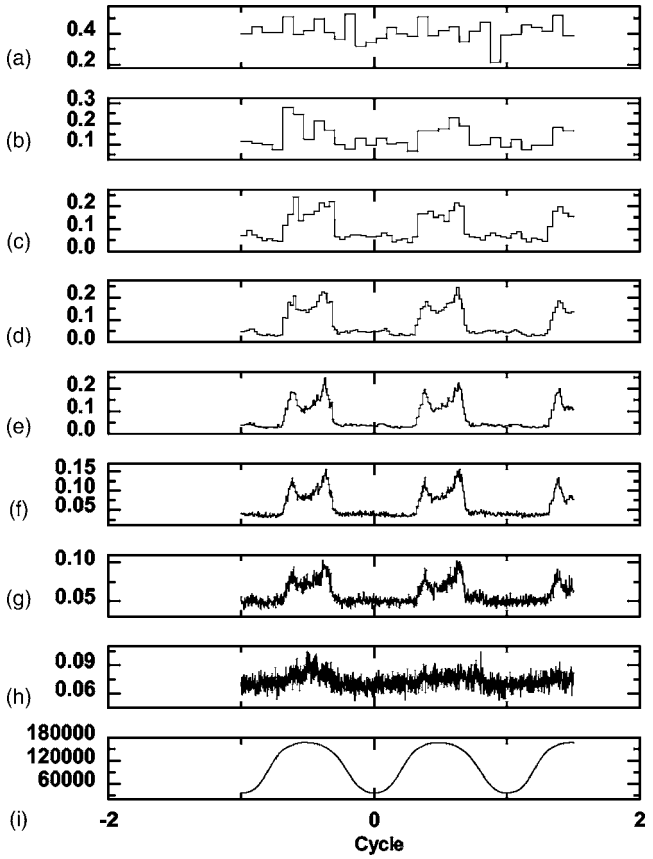


FIG. 11. Wavelet decomposition of the broadband sound generated by pulsating confined jets through straight orifice driven at 20 Hz. (a) Seventh approximation; (b) seventh detail; (c) sixth detail; (d) fifth detail; (e) fourth detail; (f) third detail; (g) second detail; (h) first detail, and (i) orifice area function in arbitrary unit.

B. Wavelet analysis

Broadband sound produced by turbulence includes contributions from turbulent structures of varying length and time scales, each corresponding to a characteristic frequency distinct from that of the periodic component. Sound production by turbulent structures of different scales during one oscillation cycle was investigated using wavelet analysis. The Daubechies 1 wavelet, which resembles a step function, was used in this study. A seven-level discrete wavelet transform was applied to each broadband sound record. The resulting details and approximations were then ensemble averaged on an energy basis to obtain mean values.

Figure 11 shows the wavelet coefficients for the broadband component of the sound obtained for a uniform orifice driven at 20 Hz, with $\Delta p = 12$ cm H_2O. The details of each level and the approximation at the seventh level are shown. Also shown is the orifice area [Fig. 11(i)]. No corrections were made to account for time delays due to either flow convection or sound propagation and reflection. The amplitude of each detail and approximation measured the sound energy at each scale and time value. Significant sound energy was present at very large scales or low frequencies (the seventh approximation). This is consistent with the high sound level at low frequencies in the stationary jets sound spectra and was attributed to local turbulent pressure fluctuations over the microphones (Zhang *et al.*, 2002a). The turbulent sound energy decreased monotonically with the nominal frequency. The wavelet coefficients at all levels featured a modulating envelope which had the same frequency as the orifice oscillation. During one cycle, sound energy was significant only over the part of the cycle at which the orifice area was nearly maximum. A threshold value existed in the orifice area below which sound radiation was negligible. In pulsating jet flows, the turbulence-generated sound is equal to the rate of change in the axial kinetic energy in the source region (Zhao *et al.*, 2001). Thus, the total axial kinetic energy change in the source region due to the flow pulsation acts as a low-frequency quadrupole source, which may give rise to the modulation at the driving frequency shown in Fig. 11.

For intermediate scales, the wavelet coefficients culminated twice over each cycle, shortly before and after the instant of maximum orifice area. The first peak appeared to be associated with the developing part of the jet. The second

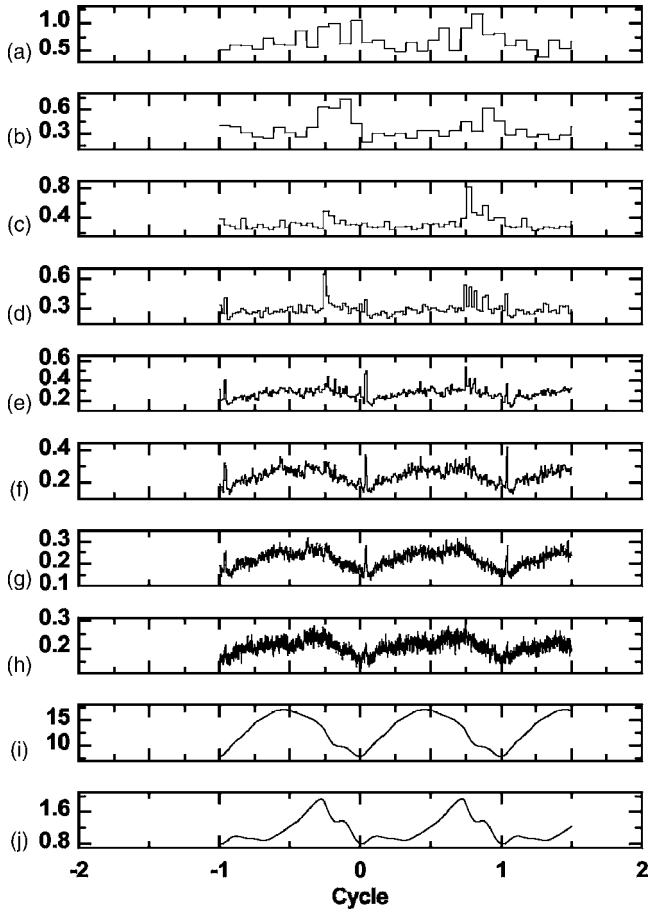


FIG. 12. Wavelet decomposition of the broadband sound generated by pulsating confined jets through divergent orifice driven at 20 Hz. (a) Seventh approximation; (b) seventh detail; (c) sixth detail; (d) fifth detail; (e) fourth detail; (f) third detail; (g) second detail; (h) first detail; (i) orifice area function in mm^2; and (j) U^6A^2.

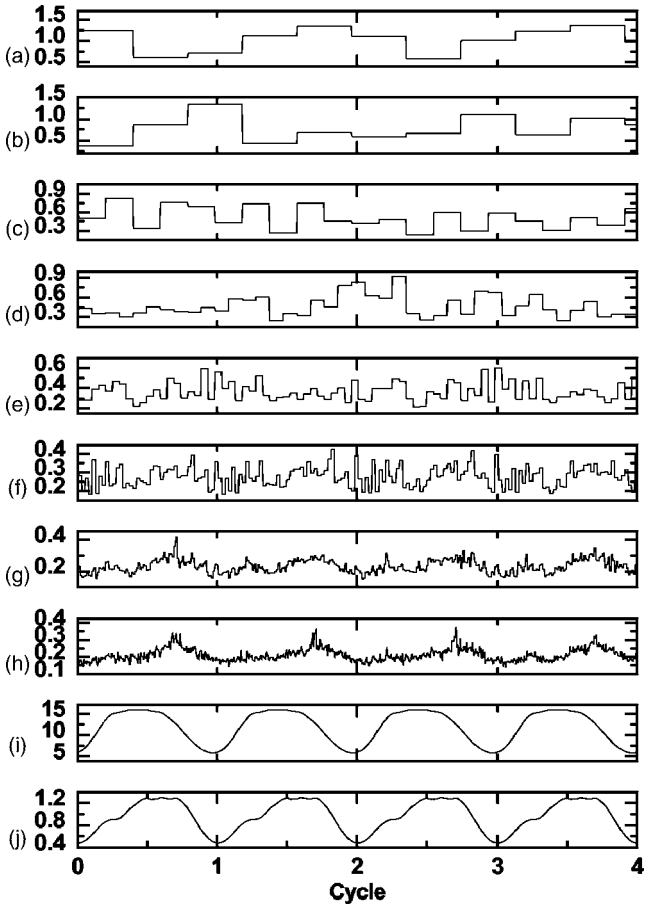


FIG. 13. Wavelet decomposition of the broadband sound generated by pulsating confined jets through divergent orifice driven at 100 Hz. (a) seventh approximation; (b) seventh detail; (c) sixth detail; (d) fifth detail; (e) fourth detail; (f) third detail; (g) second detail; (h) first detail; (i) orifice area function in mm^2; and (j) U^6A^2.

peak was associated with the quenching of the flow during orifice closing. Asymmetric behavior about the instant of the maximum orifice opening was observed, especially at small scales or high frequencies. The second peak was slightly delayed in time, and its amplitude was slightly larger than that of the first peak. For small-scale (high-frequency) components [e.g., the first detail in this case, Fig. 11(h)], the first peak was smaller in amplitude than that for the larger scales, indicating lower high-frequency sound production compared with the mid-frequency components during the opening stage (or the jet formation stage). This may indicate that the small-scale flow structures need more time to develop than the underlying large scale flow.

Figure 12 shows the wavelet decomposition for the case of a divergent orifice driven at 20 Hz. The orifice displacement was very small, and thus the orifice was open during the entire cycle. The details again show a modulation at the driving frequency. The two peaks are still discernable, although not so clearly. Also shown in the figure [Fig. 12(j)] is the sound power scaling law U^6A^2 developed in Zhang *et al.* (2002a) for low frequencies. This scaling law, describing the sound power produced in stationary flow, would yield a first approximation of the sound power produced over one cycle in pulsating flows. Neither the scaling law nor the orifice area history was in good agreement with the wavelet details, indicating again that the broadband sound generation process may not be perfectly quasi-steady. However, the U^6A^2 scaling law was generally in better agreement with the wavelet details than with the orifice area. Considering the non-quasi-steady behavior discussed in the previous section, the peaks in the U^6A^2 scaling law coincided approximately with the peaks in the details of the wavelet decomposition. This is particularly true for the high level of details [compare Fig. 12(b) with 12(j)]. The high level of detail corresponds to the dominant contributions from large-scale turbulence and appears to be a good indicator of sound power.

The experiments were repeated at higher frequencies, more typical of voice production. Figures 13 and 14 show the wavelet decomposition for a divergent orifice driven at 100 and 200 Hz, respectively. Figure 15 shows the wavelet decomposition for a convergent orifice driven at 100 Hz. The low-frequency modulation was present for all cases, except in the high level of detail for the case of the divergent orifice driven at 100 Hz. As discussed earlier, the periodic component may not be completely removed by the ensemble-averaging process for this case. The wavelet details showed two peaks over one cycle for intermediate scales. The first peak, within the opening stage, gradually disappeared as the

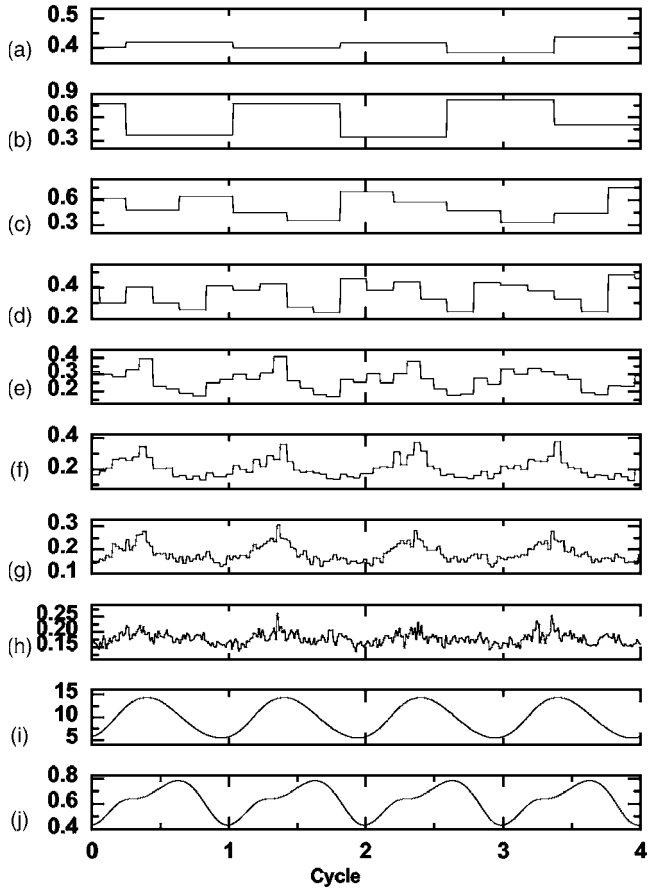


FIG. 14. Wavelet decomposition of the broadband sound generated by pulsating confined jets through divergent orifice driven at 200 Hz. (a) seventh approximation; (b) seventh detail; (c) sixth detail; (d) fifth detail; (e) fourth detail; (f) third detail; (g) second detail; (h) first detail; (i) orifice area function in mm^2; and (j) U^6A^2.

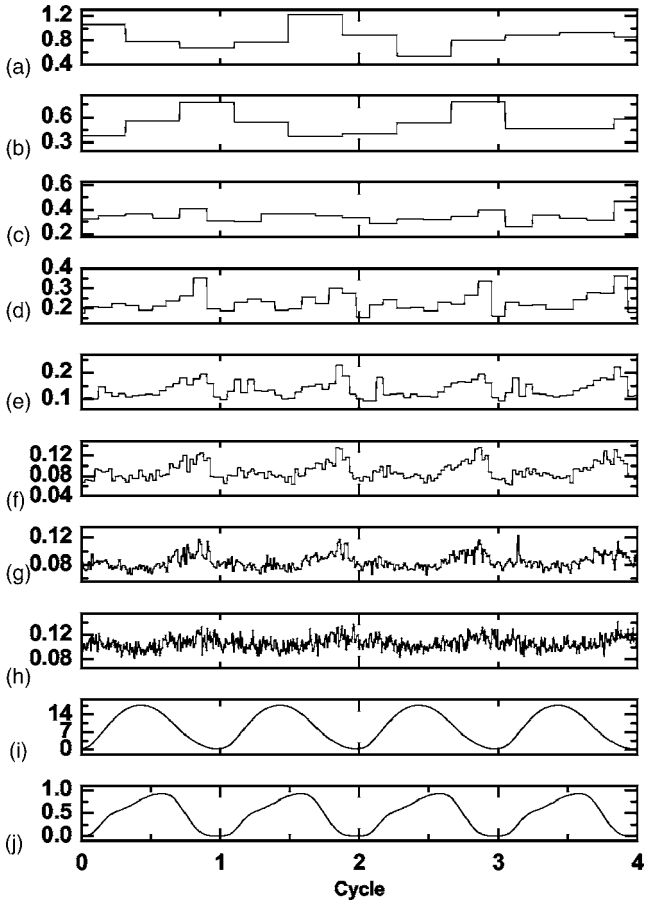


FIG. 15. Wavelet decomposition of the broadband sound generated by pulsating confined jets through convergent orifice driven at 100 Hz. (a) seventh approximation; (b) seventh detail; (c) sixth detail; (d) fifth detail; (e) fourth detail; (f) third detail; (g) second detail; (h) first detail; (i) orifice area function in mm^2; and (j) U^6A^2.

frequency increased. The general shape of the wavelet details matched the scaling law U^6A^2 better than the orifice area function.

For the same level of details, the first peak was less apparent as the driving frequency was increased. This indicates that at frequencies typical of voice production, the turbulent field, especially the fine-scale structures, may not have enough time to fully establish itself. Most of the turbulence sound at high frequencies was produced during the orifice closing stage. No jet tones were observed for jet flows through a divergent orifice, in contrast to that observed in the case of stationary jets (Zhang *et al.*, 2004). Those feedback mechanisms responsible for the jet tones may not have enough time to lock in for pulsating flows (Pelorson *et al.*, 1994).

Note that for the case of the divergent orifice and 200 Hz, the details were slightly ahead of the orifice area function in shape. This misalignment was probably due to delays associated with acoustic propagation and flow convection, which were not accounted for and may be important at high driving frequencies.

The analysis above has been repeated using other wavelet functions. Although details differ, qualitatively similar results were obtained. It is possible that some wavelets are more appropriate than others in this particular application. However, this is beyond the scope of this study and was not investigated further.

V. DISCUSSION

A major difference between the mechanical model and human phonation is that the motion of the vocal fold model in this study was imposed independently of the flow field and the acoustical environment, rather than induced by the glottal flow as in the case of human phonation. The deformation amplitude of the rubber vocal fold model was limited. The fluid-structure interactions within the human glottis were therefore not reproduced in the driven physical model. Such interaction may help to smooth the glottis motion, especially during orifice opening and closing, therefore reducing flow unsteadiness. The flow in real speech may not be as unsteady as it is in the mechanical model. This issue will be further investigated using more realistic vocal fold models.

Complete orifice closure was not always achieved in the experiments. Incomplete closure of the orifice implies that there is always a mean flow through the orifice. The possible effects of a nonzero mean flow on the broadband sound production and the validity of the quasi-steady approximation were not investigated in this study. The presence of a mean

flow might increase the radiated broadband sound energy during the initial opening and final closing stages, due to the persisting turbulent structures downstream in the vocal tract through the entire oscillating cycle, as compared to the case of total orifice closure. Small-scale turbulence would likely have more time to develop when a mean flow is present, leading to more sound production at high frequencies. However, this effect may be counteracted by the reduced flow unsteadiness due to the reduced orifice vibration amplitude, other experimental parameters remaining unchanged. This is particularly true in the initial opening and final closing stages, during which the flow unsteadiness boosts the broadband sound production as shown in this study. On the other hand, reduced flow unsteadiness implies that the quasi-steady models should better predict the broadband sound production than in the case of complete orifice closure. In the extreme case, when the oscillation amplitude is very small (e.g., in a falsettolike vibration), the flow is expected to be predicted well by using the quasi-steady approximation.

Although deviations from the quasi-steady approximation were observed in the broadband sound production, the scaling law for sound power (U^6A^2) developed in previous stationary flow studies qualitatively predicted the unsteady broadband sound production during one oscillation cycle. This scaling law for the sound power and the scaling law for the source spectral function (Zhang *et al.*, 2002a) may be used for noise modeling in high-quality synthesis of normal and pathological voices. To implement the broadband sound component, each glottal oscillation cycle may be subdivided into short-duration time segments. A typical velocity and diameter may then be assigned to each time segment, with their values determined from the harmonic source (either a prescribed pulse train or derived from speech signals using inverse filtering) during the corresponding time segment. Spectral filters can then be designed from the typical velocity, diameter, and the appropriate spectral similarity law developed in Zhang *et al.* (2002a). For each time segment, a random white noise of longer duration is passed through the corresponding spectral filter and a constituent broadband sound is generated. The final broadband sound is obtained by assembling these short segments of the constituent broadband sounds, properly scaled, in sequence. The unsteady effects (deviations from the quasi-steady approximation) can be accounted for by introducing a scaling factor for each time segment. The resulting broadband sound could then be superimposed onto the periodic sound component. Others models such as the Meyer-Eppler model could be also implemented. Comparisons between the qualities of the resulting voice output from these two models as well as original speech signals may be used to evaluate the effectiveness of the developed sound power and spectral similarity laws.

VI. CONCLUSIONS

Experiments were performed to study the characteristics of broadband sound generation by confined pulsating jets. The broadband sound component was isolated from measured sound signals by the method of ensemble averaging. The probability density function (PDF) of the broadband sound was calculated and compared to that of the sound radiated from a stationary jet flow with the same orifice area for the same flow boundary conditions. A good agreement was obtained between the PDF of the pulsating jets and the stationary jets at instants during one cycle when the orifice area was close to its maximum value. Discrepancies between the PDFs at the initial and ending stages of the orifice oscillating cycle were observed due to flow acceleration and deceleration. This indicates that the quasi-steady approximation may be valid for the broadband component when the turbulence is well established and the turbulence energy is comparatively large.

A wavelet analysis (seven-level discrete wavelet transform using the Daubechies 1 wavelet) of the broadband sound showed that random sound production was modulated at the driving frequency. Although deviations from quasi-steady approximation were observed, the scaling law for sound power developed in previous studies matched qualitatively the general shape of the sound production by the large-scale turbulence during one orifice cycle. Dual peaks in the wavelet coefficients were observed during one cycle for the orifice area near its maximum value. The first peak was most likely associated with jet formation, while the second was associated with flow deceleration during orifice closing. The first peak became less apparent at high frequency, suggesting that there was very little small-scale or high-frequency sound production during the opening phase. Increased driving frequency was found to have more impact on sound from small-scale turbulence than that from large-scale components. As the driving frequency increased, sound production during the opening stage was also reduced due to the shorter time available for turbulence to develop.

ACKNOWLEDGMENT

This study was supported by Research Grant No. RO1 DC03577 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health.

Bendat, J. S., and Piersol, A. G. (**1986**). *Random Data: Analysis and Measurement Procedures* (Wiley, New York).

Dabiri, J. O., and Gharib, M. (**2005**). "Starting flow through nozzles with temporally variable exit diameter," J. Fluid Mech. **538**, 111–136.

Daubechies, I. (**1988**). "Orthonormal bases of compactly supported wavelets," Commun. Pure Appl. Math. **41**, 909–996.

Davies, H. G., and Ffowcs Williams, J. E. (**1968**). "Aerodynamic sound generation in pipe," J. Fluid Mech. **32**(4), 765–778.

Howe, M. S., and McGowan, R. S. (**2005**). "Aeroacoustics of [s]," Proc. R. Soc. London, Ser. A **461**, 1005–1028.

Iguchi, M., Yamazaki, H., Yamada, E., and Morita, Z. (**1990**). "Velocity and turbulence intensity in a pulsating jet through a sudden expansion," Trans. Jpn. Soc. Mech. Eng., Sec. B **56**, 1659–1664.

Krane, M. H. (**2005**). "Aeroacoustic production of low-frequency unvoiced speech sounds," J. Acoust. Soc. Am. **118**, 410–427.

Mallat, S. G. (**1998**). *A Wavelet Tour of Signal Processing* (Academic, San Diego).

Meyer-Eppler, W. (1953). "Zum Erzeugungsmechanismus der Geräuschlaute (On the production mechanism of the obstruent sounds)," Z. Phonetik und allgemeine Sprachwissenschaft (Journal of Phonetics and Linguistics) **7**, 196–212.

Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. M. (**2005**). *Wavelet Toolbox User's Guide* (The MathWorks, Inc., Natick, MA).

Mongeau, L., Franchek, N., Coker, C. H., and Kubli, R. A. (**1997**). "Characteristics of a pulsating jet through a small modulated orifice, with application to voice production," J. Acoust. Soc. Am. **102**, 1121–1133.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Zhang, Z. (2002). "Experimental study of sound generation by confined jets with application to human phonation," Ph.D. thesis, Purdue University, West Lafayette, IN.

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002a**). "Broadband sound generation by confined turbulent jets," J. Acoust. Soc. Am. **112**(2), 677–689.

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002b**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**(4), 1652–1663.

Zhang, Z., Mongeau, L., Frankel, S. H., Thomson, S., and Park, J. (**2004**). "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**(3), 1720–1728.

Zhao, W., Frankel, S. H., and Mongeau, L. (**2001**). "Numerical simulations of sound from confined pulsating axisymmetric jets," AIAA J. **39**(10), 1868–1874.

Evidence for attractors in English intonation

Bettina Braun,[a] Greg Kochanski, Esther Grabe, and Burton S. Rosner
Phonetics Laboratory, University of Oxford, 41 Wellington Square, Oxford OX1 2JF, United Kingdom

(Received 11 May 2005; revised 15 March 2006; accepted 17 March 2006)

Although the pitch of the human voice is continuously variable, some linguists contend that intonation in speech is restricted to a small, limited set of patterns. This claim is tested by asking subjects to mimic a block of 100 randomly generated intonation contours and then to imitate themselves in several successive sessions. The produced f_0 contours gradually converge towards a limited set of distinct, previously recognized basic English intonation patterns. These patterns are "attractors" in the space of possible intonation English contours. The convergence does not occur immediately. Seven of the ten participants show continued convergence toward their attractors after the first iteration. Subjects retain and use information beyond phonological contrasts, suggesting that intonational phonology is not a complete description of their mental representation of intonation.
 [DOI: 10.1121/1.2195267]

PACS number(s): 43.70.Fq, 43.71.Bp, 43.66.Ba, 43.66.Fe [AL] Pages: 4006–4015

I. INTRODUCTION

The pitch of the human voice is continuously variable. Nevertheless, phonologists often assert that any language uses only a small set of different patterns to control intonation (variation in pitch, whose primary acoustic correlate is fundamental frequency or f_0). Intonation in English, for example, is said to behave this way (Cruttenden, 1997; Kingdon, 1958; O'Connor and Arnold, 1961). Similar claims have been made for numerous other languages. See Hirst and Cristo (1998) for discussions of intonation patterns in European and non-European languages.

Listeners supposedly interpret and make linguistic sense of continuous pitch changes in speech via such basic patterns (Brazil, 1985; Jun, 2005; Pierrehumbert and Hirschberg, 1990). In intonation languages, such as Dutch, German, or English, pitch variations seem to help in signaling prosodic phrasing, different information structure packaging (Steedman, 2000; Vallduví and Engdahl, 1996), or attitudinal and emotional information (Scherer, 1985). The belief in a small set of basic intonation patterns is based primarily on a linguist's or a subject's conscious classification of contours (Gussenhoven and Rietveld, 1997; Kohler, 1991; Ladd and Morton, 1997). Experimental evidence on the validity of linguistic descriptions of intonation is very limited.

One line of attempts to obtain such evidence has used the "categorical perception" paradigm (Liberman *et al.*, 1957). A set of speech stimuli is generated whose f_0 contours are spaced along a continuum between two supposedly basic prototypes. In an identification task, participants classify each stimulus as belonging to one of the prototypes. In a discrimination task, the observers indicate whether they hear a difference between paired stimuli that are near neighbors on the continuum. The hallmark of "categorical perception" is that discrimination can be predicted from identification. Qualitatively, all examples within a category should be perceived as similar, making it difficult to discriminate pairwise between them. In contrast, discrimination should be easy for two items that straddle the boundary between categories, producing peak performance.

A maximum in the discrimination function was reported for early and medial peaks in German intonation (Kohler, 1991) and for the perception of high and low boundary tones in Dutch (Remijsen and van Heuven, 1999). However, discrimination within categories was better than the identification data predicted. In another experiment, Ladd and Morton (1997) tested the perception of peak height. Their listeners could classify the stimuli as normal or unusual, but discrimination was not even maximal across the classification boundary. In studies on lexical tone in Cantonese, the predictability of discrimination from identification varied with the type of contrast under study (Francis *et al.*, 2003). On the whole, these data show no clear examples of categorical perception.

Obviously, identification relies on conscious classification of the speech stimuli. Furthermore, it shows nothing about whether the ends of the continuum represent basic psychological structures or are simply transient categories imposed by the experiment. Color naming provides a case in point. In a paper by Doll and Thomas (1967), subjects were trained to label two different wavelengths and were then tested on intermediate wavelengths, to generate an identification function. Training on different pairs of wavelengths then resulted in different identification functions. Similar effects occur in speech perception (cf. Eisner and McQueen, 2005 and references therein). Such easily shifted category boundaries apparently cannot be deeply embedded in our perceptions. Indeed, boundaries can move so rapidly (Ladefoged, 1989; Ladefoged and Broadbent, 1957) that they might not even be stable over the duration of a categorical perception experiment. For these and other reasons, the concept of categorical perception and its attendant experimental paradigm have become increasingly disfavored [see Schouten *et al.* (2003) and Plomp (2002, pp. 137–145)].

In a different approach, Pierrehumbert and Steele (1989) constructed a set of contours that varied in the position of a

[a]Present address: Max Planck Institute for Psycholinguistics, Postbus 310 NL-6500 AH Nijmegen.

peak in f_0. They asked subjects to repeat each carrier utterance and measured the positions of the resulting peaks. Although the paper lacks a statistical analysis, the authors interpret the data as showing that the subjects produced bimodal distributions of peak positions. The paper introduced a valuable method, although it was limited in certain respects. Beyond the lack of statistics, one of the five subjects was an author and was therefore not naïve.

These previous experimental efforts have left undefined the number and properties of basic intonational patterns. Even their psychological reality can be doubted. We therefore set out to obtain better behavioral evidence on whether basic intonational patterns affect the perception and production of speech. To do this, we employ iterative mimicry. Mimicry is a simple behavioral response to language. It appears early in children's language development, long before grammar and comprehension are fully established (Loeb and Allen, 1993; Meltzoff and Decety, 2003; Snow, 1998). Results from mimicry therefore should provide a better picture of a subject's language processing than do conclusions based on introspection or explicit and conscious classification.

Bartlett (1932) long ago exploited a version of iterative mimicry to study drawing, among other things. His Method of Serial Reproduction required a subject to produce a drawing after short exposure to an original. The reproduction was presented to another subject who attempted a new reproduction. This between-subjects procedure was iterated some dozen or so more times. Bartlett found that the successive reproductions were gradually simplified or even transformed into something very different from the original drawing.

Our procedure of iterative mimicry builds on this method but remains strictly within each subject. It provides a substantially new approach to studying the way in which intonation is processed. Our technique is related to those used by Pierrehumbert and Steele (1989) and Repp and Williams (1987), but it extends their efforts in several ways. In our procedure, a subject first mimics an initial set of English utterances with widely varied f_0 contours. In a second session, the subject then mimics his/her own first productions. Continuing in two more sessions, he/she mimics his/her own productions from the immediately preceding session. If basic intonational contours are part of the subject's mental structures for English, the contours in the subject's productions should progressively reduce over sessions towards a few, well-distinguished forms.

Such well-distinguished contours can be idealized mathematically as "attractors," a concept from the mathematics of iterated functions. A function $f(x)$ is applied to a starting value x_0 and then successively to its own result, yielding a sequence of values x_0, $x_1=f(x_0)$, $x_2=f(x_1)$, …. A simple example of an attractor comes from the function $f(x)=x^2$. Zero is called a fixed point of this function, since $f(0)=0$. [Generally, w is a fixed point of a function g if $g(w)=w$.]

Consider values of the argument of $f(x)$ in the region around zero, where $|x_0|<1$. Applying $f(x)$ to any value in that region yields a result that has a smaller absolute value than the input value. Succeeding applications of the function produce still smaller values that approach zero as the sequence continues. For the function x^2, then, zero is a *stable* fixed point. Nearby values will converge over iterations toward any stable fixed point. This convergence is a key feature of an attractor. The points from which the sequence converges form an attractor's "basin of attraction."

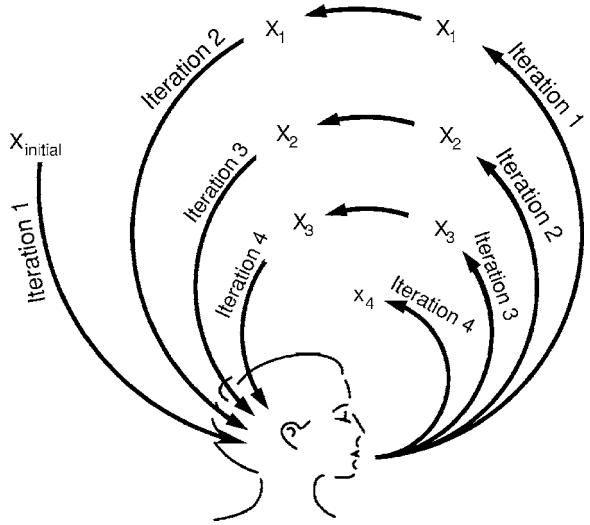


FIG. 1. Scheme of the experiment. The numbers refer to the four experimental sessions, called iterations. In iteration 1, the subject mimics the initial synthesized stimuli X_{initial}, producing responses x_1. In each later iteration, the subject mimics his or her own productions from the preceding session. Each response serves as a stimulus in the next iteration.

For iterative mimicry, x represents a description of an entire intonation contour as a vector of f_0 values for each moment in time. Then $f(x)$ represents the transformation from the sound entering the subject's ears to the sound produced by the subject. If each x in a subset of starting contours converges over successive imitations towards a fixed shape, this would provide evidence for the existence of an attractor in intonation. That subset of contours would lie in a basin of attraction. Different subsets of starting contours that converged towards different fixed shapes would lie in different basins of attractions. The attractors themselves would represent underlying mental structures.

Using iterative mimicry to search for underlying attractors has a crucial advantage over the other experimental methods described thus far: it does not require any conscious classification of intonation by either the subject or the experimenter. This experimental procedure comes much closer than others to the actual use of language in conversation. We respond to language, but rarely do we consciously and explicitly analyze the intonation we hear. Hence, our method avoids the possibility that conscious reports on intonation may not correspond to responses that speakers would make in a conversation.

II. PROCEDURES

A. General design

We designed the experiment in analogy with the iterated-function definition of attractors given above. Figure 1 shows this design. Each subject serves in four experimental sessions. In the first session, "iteration 1," the subject mimics an initial set of utterances X_{initial}. The f_0 contours of these utterances are systematically varied by resynthesis. Male and

female subjects hear initial stimuli that are based on a male and a female voice, respectively. We record the subject's response to each X_{initial}. We call this first set of responses x_1. These responses become the stimuli X_1 for the next session.

In that session, the subject mimics each utterance in X_1, producing a second set of responses x_2. These responses in turn become the stimuli X_2 for the next session. The responses x_3 from that session become the stimuli X_3 for the final session, which is "iteration 4." In that iteration, the subject mimics each utterance in X_3, yielding the final set of responses x_4. Except for the initial stimuli X_{initial}, a subject hears only his or her own utterances from the preceding iteration.

During each iteration, the subject needs to remember the intonation pattern for only a short time between stimulus and mimicry. A stimulus is presented to the subject who begins to respond immediately when ready. The mean interval within a session between the end of a stimulus and the beginning of the response to it was 130 ms with a standard deviation of 55 ms. Subjects therefore typically started moving their articulators before a stimulus ended. The median spacing between iterations (i.e., x_k to X_k) was 5 days, with a minimum spacing of 1 day. Additionally, we scrambled the order of presentation of stimuli between iterations so that subjects would not be able to track an utterance from one iteration to the next.

B. Participants

Participants were linguistically naïve speakers of Standard Southern British English, five males and five females. All had normal hearing and were between 19 and 30 years old. They received written instruction to imitate the speech and the melody of each sentence as closely as possible. They were informed that the initial stimuli were synthesized but that their task was not to imitate the voice *quality* of the stimuli. To ease any discomfort at hearing their own voice in the self-mimicking sessions, subjects were told that they would hear processed versions of their own speech. (See Sec. II E for details of the processing.)

C. Materials

The target sentences were chosen to be mostly sonorant subject-verb-object constructions with a total of seven syllables each. They were constructed so that in a neutral reading, accents occur on the first and the sixth syllables. Nine candidate sentences were generated.

D. Initial recording and selection

We interspersed these nine candidates with fillers of various lengths and syntactic structures, in order to obtain starting utterances for resynthesis. One male and one female speaker of Southern British English read the list, both with a rising (e.g., questioning) and falling (i.e., statement) final intonation. The recordings were made in a sound-treated room and digitized at 16 kHz with 16 bits/sample. The productions from the male and the female speaker were processed separately to provide initial stimuli for the male and female subjects, respectively. Having two versions of the recordings minimized the eventual size of the f_0 shifts that we would later need to impose; resynthesis quality generally declines as f_0 shifts become larger.

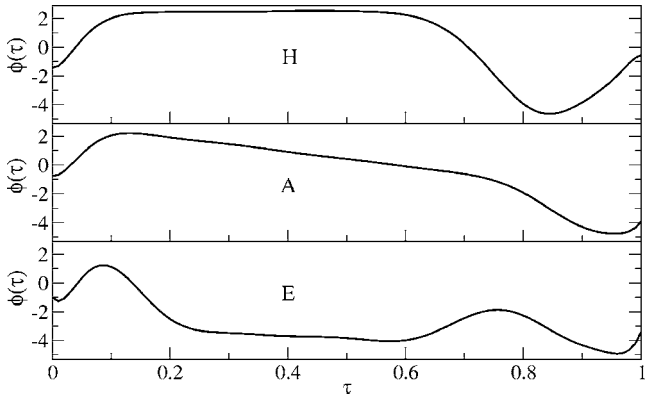


FIG. 2. The three basis contours from which initial stimuli were constructed (Grabe *et al.*, 2003). The ordinate is in semitones, relative to mean f_0; the abscissa is normalized time.

The nine candidate sentences were PSOLA-resynthesized with the three basis intonation contours A, E, and H of Grabe *et al.* (2003), see Fig. 2. (Resynthesis of A and E was based on the falling recordings, H on the rising recordings.) The basis contours were chosen to be maximally distinct among normal English contours. The authors separately assessed the quality and naturalness of the 54 resynthesized versions. Six sentences survived this initial selection.

Next, to avoid any possible bias toward either rising or falling patterns (Ryalls *et al.*, 1994), we constructed versions of these six sentences that were intermediate between the rising and falling recordings in terms of duration and LPC coefficients (see the Appendix). One more sentence was eliminated because the resulting audio had an unnatural segmental structure. This finally left the following five sentences for the main experiment, each with a male and female version:

(1) Anna will marry Marlon.
(2) Alan rode on a llama.
(3) Eleanor blamed our neighbour.
(4) Melanie won a million.
(5) Minnie will ring a lawyer.

1. Intonation contours for initial stimuli: $X_{(\text{initial},k)}$

One hundred fifteen initial f_0 contours were generated by taking linear combinations of the basis contours shown in Fig. 2. A combination was defined by $c_A \cdot a(\tau) + c_E \cdot e(\tau) + c_H \cdot h(\tau)$, where $a(\tau)$, $e(\tau)$, and $h(\tau)$ are the basis contours expressed as functions of normalized time τ that ranges from 0 to 1. We used combinations of c_A, c_E, and c_H such that $c_A + c_E + c_H = 1$, $c_A > -0.3$, $c_E > -0.3$, and $c_H > -0.3$.

The 115 combinations of c_A, c_E, and c_H were selected to avoid clusters. The selection algorithm operated iteratively. At each step, it chose 100 candidate samples of (c_A, c_E, c_H) from a uniform distribution and accepted the candidate that was farthest from all previously accepted samples. (Farthest is defined via the Euclidean distance between the two

(c_A, c_E, c_H) vectors: $[(C_A - c'_A)^2 + (c_E - c'_E)^2 + (c_H - c'_H)^2]^{1/2}$.) This resulted in a set of points that were more uniformly distributed than would result from independent random sampling. To avoid priming the subjects with the basis contours, we excluded regions where (c_A, c_E, c_H) was within 0.1 of (1,0,0), (0,1,0), or (0,0,1). This process finally produced initial intonation contours that were generally intermediate between the basis contours (when $c_A > 0$, $c_E > 0$, and $c_H > 0$) but also included some mildly exaggerated versions of the basis contours when $c_A < 0$, $c_E < 0$, or $c_H < 0$.

2. Synthesizing the initial stimuli

Each initial intonation contour was superimposed on a segmentally intermediate target sentence from the male and the female speaker, using PSOLA resynthesis. Although each subject heard the resynthesized male or female initial utterances with the same initial contours, f_0 was scaled up or down from those utterances to match the average f_0 of the participant's own speech. For the ten subjects, this ultimately produced 1150 synthetic utterances.

E. Data collection

Each experimental session used up to five successive blocks of stimuli:

- Block A: A practice block of 15 stimuli (generated per Sec. II D 1). These synthetic stimuli were the same on all iterations. Each experimental sentence was synthesized with three distinct contours.
- Block B: Re-recordings of unusable productions detected after the end of the previous iteration. We re-recorded when the subject spoke the wrong words or substantially hesitated in the midst of a sentence, when bursts of noise interfered (e.g., the subject touched the microphone), or when occasional technical problems arose.
- Block C: A block to test reproducibility. In the first iteration, this was identical to block A. In succeeding iterations, the stimuli in this block were always the productions of block C, iteration 1. The productions were adjusted to a consistent amplitude, and initial or final breath and lip-smack noises were removed before they were used as stimuli.
- Block D: This was the main experimental block. The first iteration utilized 100 synthesized stimuli. Each of the five finally selected sentences was resynthesized with 20 maximally separated contours to cover the space of possible contours. Each sentence carried a different block of contours. In all suceeding iterations, the stimuli were the productions of block D from the previous iteration. The productions were adjusted to a consistent amplitude, and initial or final breath and lip-smack noises were removed before they were used as stimuli. We randomized the order of the stimuli between each iteration so that the subjects could not keep track of the history of each stimulus.
- Block E: This block re-recorded blatant mistakes that the experimenter noticed while responses to block D were being recorded.

If a subject was dissatisfied with a production, one repeat was immediately available. Eleven percent of the utterances were re-recorded at once for this reason. Between blocks B and E, an additional 5% of the productions were re-recorded.

With 115 stimuli per session, 4 sessions, and 10 participants, the corpus contains 5200 responses in total. Of these, the 4000 utterances from blocks D (with some replacements from by B and E) are the basis of the results presented below.

F. Signal processing for mimicry data

Signal processing of the mimicry productions involved three main steps: inspection and modification of f_0 tracks, weighting, and normalization. The processing generally followed Kochanski *et al.* (2005). Before analysis, the f_0 tracks obtained from `get_f0` (Talkin, 1995) were screened for gross errors. Tracks that had frequency shifts of one octave, (±15 Hz) between successive points were inspected, as were tracks with points that fell more than 7 semitones below the participant's average f_0. Of the 4600 utterances from blocks C and D, 125 tracks required inspection. Where necessary, f_0 was modified. In 87 utterances some region was raised by an octave. It was lowered by an octave in one. One hundred seventeen productions had the marking of some region changed from voiced to unvoiced or irregularly voiced. The marked regions contributed to an indicator of irregular voicing $I(t)$, used in Eq. (1) below. The mean length of all modified or marked regions was 35 ms, totaling less than 1% of the data.

1. Weighting the mimicry data

For most sonorant sounds, especially in stressed syllables, the perceived pitch of speech correlates well with the output of algorithms that estimate fundamental frequency. Not all our data, however, meet that criterion. In the interest of using plausible sentences, we chose some that had incomplete sonority (e.g., after /d/ in "rode," sentence 2, above). Furthermore, some voiced sounds were not strongly periodic and may not have had a clear pitch.

Consequently, not all the data are of equal value for specifying the pitch that the subject heard and attempted to reproduce. A weight function is unavoidable, because pauses have no pitch and must therefore be excluded from further analysis.

We base the weight, $W(t)$, on two local acoustic measures that seem important in describing prosody: loudness, $L(t)$, and the degree of periodicity of the waveform, $A(t)$. The net result of $W(t)$ is to focus the analyses on the peak of the syllable, paying less attention to the margins. The choice of weight is partially motivated by the pitch tracking algorithm and statistical considerations. Another factor is that low-amplitude parts of speech are often buried in normal environmental noise and thus have little or no perceptual importance. [Substantial numbers of people speak in environments where the mean signal-to-noise ratio for speech is only 9 dB (Kochanski *et al.*, 2005).]

The weight of a datum is

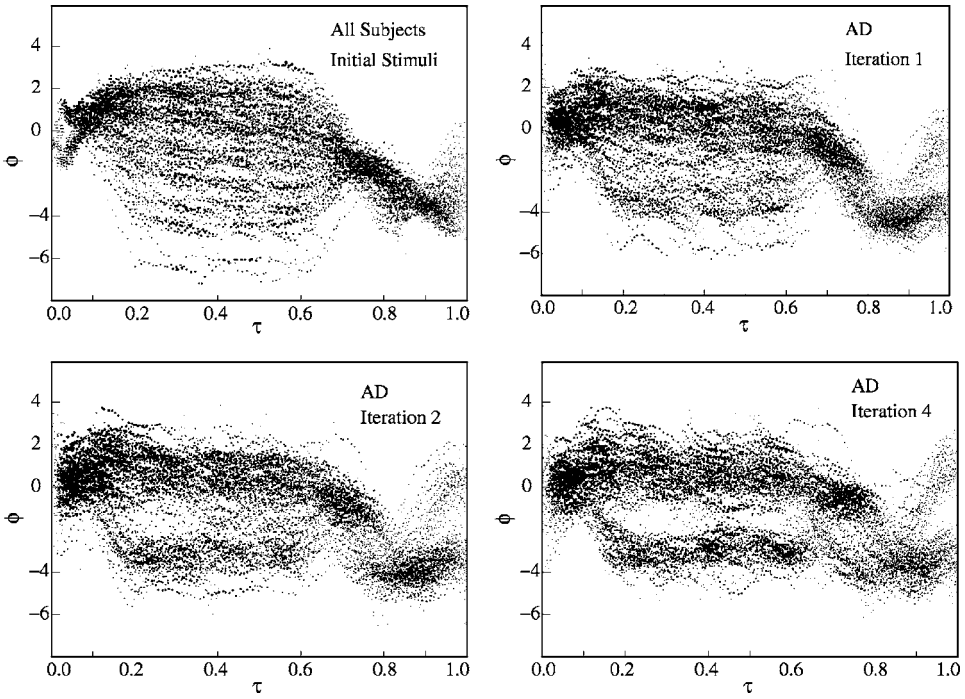


FIG. 3. Normalized f_0 contours, $\phi(\tau)$, for a typical subject, AD. In each panel, f_0 measurements are superposed from the 100 main experimental utterances. The time axis τ is normalized to the length of the utterance and the frequency axis is semitones relative to the speaker's mean f_0. The upper left panel shows f_0 contours for the initial stimuli. The next three panels contain response tracks for iterations 1 (upper right), 2 (lower left), and 4 (lower right).

$$W(t) = L^2(t) \cdot \max(1 - A^2(t), 0)^2 \cdot V(t) \cdot I^2(t) \cdot (1 + 2\tau), \tag{1}$$

which is $(1+2t/T)$ times $W_{f_0}(t)$ from Kochanski *et al.* (2005). In Eq. (1),

- $L(t)$ is the loudness (following Stevens, 1971),
- $A(t)$ is a measure of periodicity (ranging from 0 to approximately 1),
- $V(t)$ is the binary voicing estimate from the pitch detector, and
- $I(t)$ is a manual indicator of irregular voicing (its value is most often 1, but the value occasionally ranges down to zero for regions with no clear pitch), and
- τ is the normalized time, defined below in Sec. II F 2, which ranges from 0 to 1.

The last term in Eq. (1), $(1+2\tau)$, partially compensates for the typical decrease in $L(t)$ as an utterance progresses. Without some compensation, $W(t)$ would generally decline along the sentence. However, in two-accent sentences such as we use, the second (nuclear) accent is believed to be perceptually and semantically at least as important as the first (Cruttenden, 1997). The last term in Eq. (1) was chosen to provide approximate equality in $W(t)$ for the two accents. We have found that our results would not be substantially different if the coefficient in the last term were 1 or 4. Also, within the context of Kochanski *et al.* (2005), changing the weight function by raising it to the power 2 or 0.5 had minor effects. Consequently, we do not expect that the results presented here are critically dependent on the detailed definition of $W(t)$.

2. Normalization

For each f_0 track, we normalized the time axis to range from 0 to 1 by computing $\tau_{i,j}=t_{i,j}/T_i$, where i indicates an utterance, j selects a datum in the utterance, T_i is the duration of the ith utterance, and t and τ refer to real and normalized time, respectively. We normalized f_0 by dividing it by $\bar{f}$, the 10%-trimmed weighted average of $f_0(t)$ over all that subject's sentences from all iterations, and converting the quotient to semitones. This gave a normalized fundamental frequency $\phi(\tau)=12\log_2[f_0(t)/f]$.

III. RESULTS AND DISCUSSION

To check that the subjects' productions were representative of English, a phonologist (author BB) applied standard intonation labels (Beckman and Ayers, 1997) to a subset of the data. The labelling and classification of the contours generally proved straightforward. Ninety-six percent of the first imitations resulted in a previously recognized English contour. Furthermore, the contours tended to be stable. A contour received the same annotation in all four iterations 65% of the time; changes between iterations usually yielded another recognized English contour. Overall, only 15% of the contours could not be unambiguously classified.

Figure 3 shows the f_0 contours of the 100 initial stimuli and of iterations 1, 2, and 4 for subject AD. The areas of the dots are proportional to the loudness of speech multiplied by a measure of of periodicity [see Sec. II F 1, Kochanski *et al.* (2005)].

Near the center of the utterances (e.g., $0.2<\tau<0.6$, where τ is the normalized time), the spread is primarily due to the difference between basis contours E and the other two (A and H). The spread at the tail ($\tau>0.9$) contrasts H vs. (A and E). In the very first iteration, the distribution of f_0 contours for AD already develops some structure beyond that for the initial stimuli. Over the succeeding iterations, the broad distribution of f_0 near the centers of the utterances gradually splits into two branches.

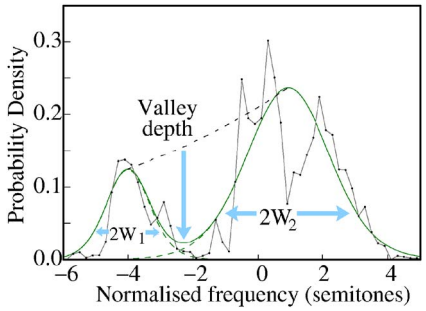


FIG. 4. Fit of a two **t**-density mixture to the central section of iteration 4, subject BP. The abscissa is f_0 in semitones relative to the speaker's average; the ordinate gives probability. The line with dots represents the histogram of the data to which the model was fitted. The smooth black curve is the best-fitting model; the dashed curves show its two components. The vertical arrow indicates where the valley depth is computed, to compare the density minimum to a density interpolated between the peaks of the two components. The width of each component is also indicated.

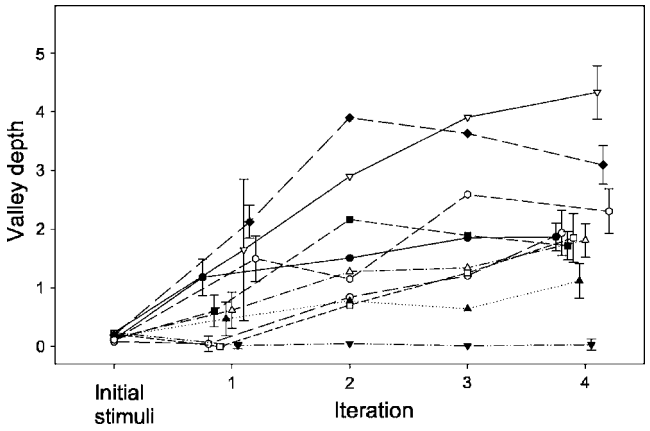


FIG. 5. Measure of bimodality in the distribution of central f_0 for initial stimuli and for mimicry responses as a function of iteration. Each curve gives results for one speaker. Ordinate is valley depth (see Fig. 4). Error bars are 1-σ from Monte Carlo simulations; other error bars (not shown) are of comparable size. (Subjects PC: ▼; EM: ▲: AD: ●; BP: ○; CB: ■; CM: □.)

Most of our subjects manifest clear splitting such as is seen for subject AD. (We show data from subject AD because they are at the median for two measures of this splitting derived below from Fig. 5. AD's final valley depth ranks sixth out of 10, and the increase in valley depth between iterations 1 and 4 ranks sixth out of 10.)

While participants *can* reproduce the intermediate contours [e.g., the "iteration 1" panel in Fig. 3, for $-2<\phi(\tau)<0$, $0.2<\tau<0.6$], those contours are unstable. They move over successive iterations into either the upper or lower stable branch (e.g., "iteration 4" panels of Fig. 3).

A. One branch or two? Modeling the distribution of central f_0

The f_0 tracks in the subjects' productions apparently collapsed into two branches over mimicry sessions. This collapse is consistent with the proposition that a subject's mental structures include certain intonation patterns as attractors. According to the mathematical model, the branches should become arbitrarily narrow, but human variability places a lower limit on the width of the branches. Statistical testing of the splitting of the f_0 tracks in our data therefore becomes necessary. To do this, we take a slice of data for each combination of subject and iteration between $\tau=0.375$ and $\tau=0.5$. Inspection showed that the distribution of such normalized f_0 values has longer tails than would a unimodal model constructed from a single Gaussian density or a bimodal model constructed from two Gaussians. We therefore represent the histogram of $\log(f_0)$ as a mixture of two of Student's **t** probability densities, one for each branch. This allows models that are somewhat heavier-tailed than Gaussian mixtures. Figure 4 shows the histogram of $\log(f_0)$ for iteration 4 from subject BP (lines with dots). It also displays the best-fitting mixture of **t** densities, along with the width of each density. The valley-depth measure is discussed below.

Through a statistical analysis, we compare how well one-and two-component models of the distribution represent such data from each subject and each iteration. The analysis uses a Markov chain Monte Carlo approach based on Bayes' Theorem. It produces samples from the joint distribution of the parameters that define the mixture and calculates the likelihood ratio (relative a posteriori probabilities) of the one- and two-density models for each subject.

1. Models for the probability density of f_0

The one-density model for the probability density of f_0 is

$$H_1(\phi;\theta_1)=\mathbf{t}\left(\frac{\phi-\mu_1}{\sigma_1},\eta_1\right), \tag{2}$$

where θ_1 is shorthand for the three parameters (μ_1,σ_1,η_1), and **t** is Student's **t** function. The two-density model is

$$H_2(\phi;\theta_2)=r\cdot\mathbf{t}\left(\frac{\phi-\mu_{21}}{\sigma_{21}},\eta_2\right)+(1-r)\cdot\mathbf{t}\left(\frac{\phi-\mu_{22}}{\sigma_{22}},\eta_2\right) \tag{3}$$

where θ_2 is shorthand for the five parameters $(\mu_{21},\mu_{22},\sigma_{21},\sigma_{22},\eta_2,r)$. In these equations, μ_k sets the center of a component, σ_k specifies its width, and η_k is the degrees-of-freedom parameter of the **t** density. We are not fitting this density to data. Instead, we are using it as a convenient way to represent a symmetric probability density that is heavy-tailed compared to a Gaussian. Therefore, η_k is a free parameter of the model: it controls how heavy the tails are. For the two-density distribution, r sets the relative probability masses of the two components.

We constrain the parameters so that $\sigma_k>0$, $0<r<1$, and $2<\eta_k<20$. ($\eta\geqslant 20$ makes the **t** density indistinguishable from a Gaussian.) We can safely constrain $\mu_{22}>\mu_{21}$ so that μ_{21} always represents the lower branch and μ_{22} represents the upper branch of f_0, as in the right-hand panels of Fig. 3.

Under the assumption of the one-density model θ_1, the probability of observing the data, $\Phi=\{\phi_{i,j}\}$ for j in the set of utterances and $0.375<\tau_{i,j}<0.5$, is

$$\log[P_1(\Phi|\theta_1)] = \sum_{\Phi} \log[H_1(\phi_{i,j};\theta_1)] \cdot \frac{W_{i,j}}{\bar{W}_j} \cdot \frac{\Delta t}{T_{\text{corr}}}. \quad (4)$$

For the two-density model θ_2, it is

$$\log[P_2(\Phi|\theta_2)] = \sum_{\Phi} \log[H_2(\phi_{i,j};\theta_2)] \cdot \frac{W_{i,j}}{\bar{W}_j} \cdot \cdot \frac{\Delta t}{T_{\text{corr}}}. \quad (5)$$

In these equations, $\bar{W}_j$ is the average weight across an utterance, and $\Delta t/T_{\text{corr}}$ is the ratio of the interval over which f_0 is measured, compared to the assumed correlation time of the f_0 measurements. We conservatively assumed T_{corr} = 100 ms, so $\Delta t/T_{\text{corr}}$=0.1. Effectively, this means that we use only 10% of our data in computing the significance of the two-component model. The actual correlation length of the f_0 measurements obtained from `get_f0` is hard to estimate precisely, due to the complexity of the algorithm. In strongly voiced regions, however, `get_f0` often reflects changes in fundamental frequency that occur over 10–20-ms time scales. A suitable value for T_{corr} therefore may well be substantially smaller than our 100-ms assumption. Reducing T_{corr} would make the two-component model more significant relative to the one-component model.

2. Statistical evaluation of the models

We used Bayes' theorem and a Markov chain Monte Carlo algorithm (Geyer, 1992; Metropolis *et al.*, 1953) to generate samples from the distributions of $P_1(\theta_1|\Phi)$ and $P_2(\theta_2|\Phi)$. These are the posterior joint probability densities of the parameters, given the observed data. We assumed flat priors over the parameter ranges specified above.

To implement statistical testing, we computed the likelihood ratio of two classes of hypotheses, where each class consists of a set of related models:

(1) the class of single-component distributions that are plausible fits to the data [specifically Eq. (2) where θ_1 selects the model within the class] and
(2) the class of two-component distributions that are plausible fits to the observed data [specifically Eq. (3) where θ_2 selects the model].

This is a straightforward application of Bayesian model averaging (Hoeting *et al.*, 1999). Assuming flat prior probabilities and no bias toward either model, the likelihood ratio is

$$R = \frac{\langle P_1(\theta_1|\Phi)\rangle}{\langle P_2(\theta_2|\Phi)\rangle} = \frac{\langle P_1(\Phi|\theta_1)\rangle}{\langle P_2(\Phi|\theta_2)\rangle}, \quad (6)$$

where the angle brackets, $\langle\ \rangle$, denote an expectation over the corresponding probability density, e.g.,

$$\langle P_1(\Phi|\theta_1)\rangle = \int P_1(\Phi|\theta_1) P_1(\theta_1|\Phi) d\theta_1. \quad (7)$$

The expectation value can therefore be implemented as an average over samples generated by the Markov chain Monte Carlo process, because the process picks samples from each region of volume $d\theta_1$ with probability $P_1(\theta_1|\Phi)\cdot d\theta_1$. Therefore,

$$\langle P_1(\Phi|\theta_1)\rangle = \sum\nolimits_{\theta_1} P_1(\Phi|\theta_1) \Big/ \sum\nolimits_{\theta_1} 1. \quad (8)$$

The expectation value for $\langle P_2(\Phi|\theta_1)\rangle$ is computed similarly.

If the likelihood ratio $R<1$, then the two-component model is a better representation of the data. By the rules of hypothesis testing, however, we do not reject the single-component model unless R falls below a confidence limit of 0.001. Our tests show that the two-density model for iterations 2–4 is statistically significant (with R actually less than 10^{-4}) for all but subject PC. On iteration 4 for PC, R <0.01 and the means of the components differ by 2.2 semitones. The components, however, have widths of 1.7 and 0.9 semitones and therefore overlap seriously. Although PC's distribution of f_0 values may require two components to represent it well, the components do not separate into clearly distinct branches.

In summary, in almost all cases, two **t** densities give a significantly better fit to the observed distributions of f_0 than does one. The best-fit **t** densities typically have noticeably but not dramatically longer tails than would Gaussians. We find η=13 with an intersubject standard deviation of 5 in Eqs. (2) and (3).

B. Valley depth

Having two components is a necessary but not sufficient condition for establishing bimodality, as the **t** densities of a two-component fit could overlap strongly. Their means could even virtually coincide. To provide further evidence for bimodality, we compute a valley depth measure (Fig. 4) from a fit of two mixed **t** densities to a histogram of central f_0. The computation finds the minimum of the curve of $H_2(\phi;\theta_2)$ between μ_{21} and μ_{22} and compares it to the value interpolated between the peaks. The minimum is obtained by simple iterative searching; the interpolation is linear on $\log(H_2)$ between $H_2(\mu_{21};\theta_2)$ and $H_2(\mu_{22};\theta_2)$. The valley depth is the negative log of the density at the minimum, divided by the interpolated density at the same point. Valley depth is zero or negative for a unimodal density. Values greater than one typically imply two well-separated components and hence two clusters.

The valley depth measure was obtained for each Monte Carlo sample of θ_2 generated from $P_2(\theta_2|\Phi)$. Figure 5 shows the set of valley depths for each speaker plotted against iteration number. The lines show the mean; standard deviations are shown for the first and last iterations. The error bars are conservative. As stated previously, our computations assumed that f_0 measurements are correlated over a large 100-ms stretch because of the dynamic programming algorithm in `get_f0`. The correlation length in such strongly voiced regions, however, is typically no larger than 20 ms. The error bars in the figure would shrink in proportion to the square-root of the correlation length. They are perhaps about half as large as plotted. The figure also shows fits to the f_0 contours for the initial stimuli.

The mean valley depth starts at 0.16 for the stimuli. Averaged over all subjects, it increases from 0.84 for itera-

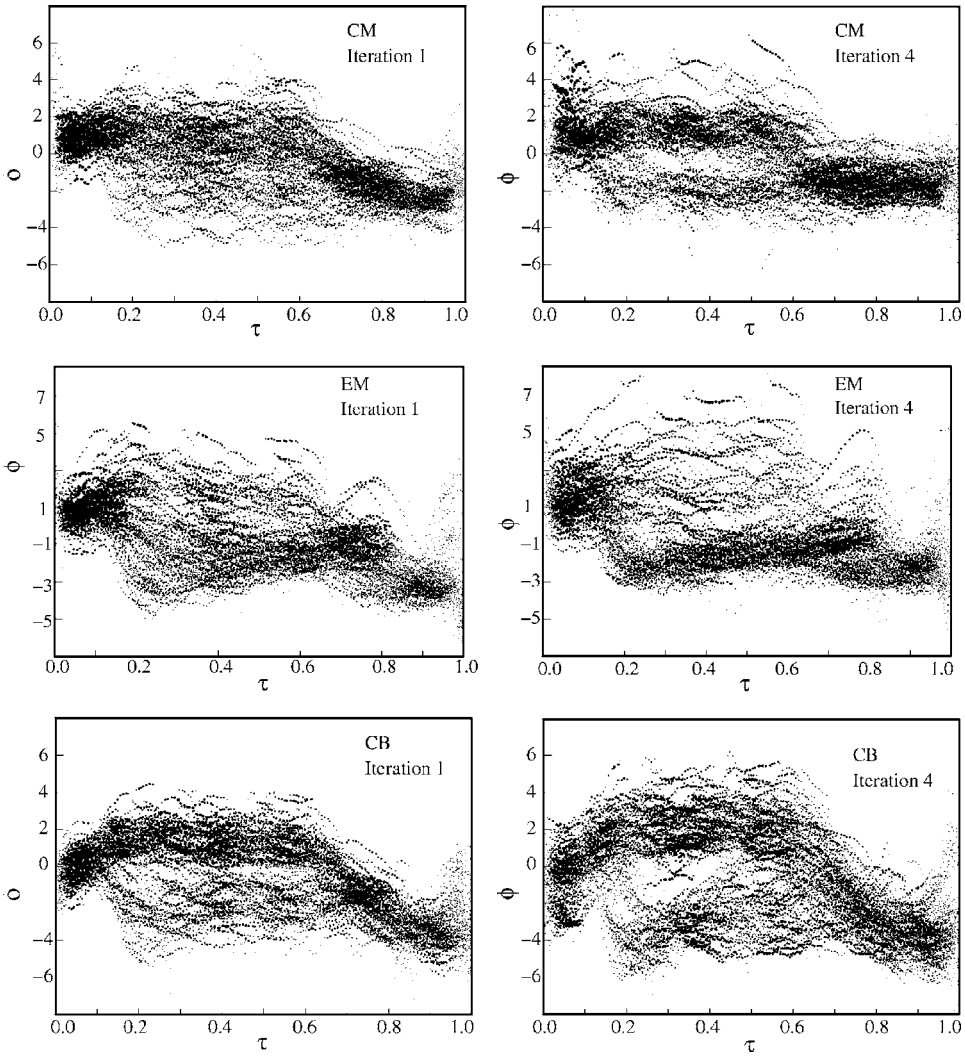


FIG. 6. Variation between subjects. This figure shows normalized f_0 contours, $\phi(\tau)$, for subjects CM, EM, and CB (top to bottom), plotted as per Fig. 3. Data from iterations 1 (left) and 4 (right); the initial stimuli are identical, as displayed in the upper-left panel of Fig. 3.

tion 1 to 1.96 for iteration 4. This is a highly significant increase ($z=8.2$, $P<10^{-6}$). Individually, nine subjects show some increase, and seven of them show a significant increase ($P<0.05$).

The exceptional subject PC (point-down filled triangles) gives a valley depth close to zero in most of the Monte Carlo samples. There is no valley between the two components fitted to PC's data, in line with the relatively high values of R found above for that subject.

Most subjects take more than one iteration to reach a stable state. Their f_0 contours did not collapse immediately into two branches. Therefore, between perception and production on a given trial, most subjects remember some details of the stimulus beyond the shape of a particular "stable attractor" contour. Intonational phonology with its choices among a small set of discrete contours cannot be a complete description of the subject's mental representations of the contours they hear.

C. Variation in and around the attractors

Subjects do not finally settle exactly on the attractor contours. The standard deviations of the two branches shrink (see Figs. 3 and 4) to about 0.7 semitones for the lower branch and 1.1 semitones for the upper. Frequency shifts smaller than this may often not be linguistically useful because they cannot be reliably perceived, remembered, or reproduced. These standard deviations are similar to rms differences between repetitions of the same utterance (Holm and Bailly, 2002, Sec. III A).

Figure 6 shows the variation in attractors across subjects. Generally, the upper branch either rises early or starts high, stays high, and then drops. The lower branch either starts high or has an early peak near $\tau=0.1$ and typically has a smaller peak around $\tau=0.8$. Final rises are common but often they sound flat or even falling, perhaps due to the final decline in loudness starting near $\tau=0.92$.

Subject CM (top of Fig. 6) has no peaks in the lower branch (unlike AD in Fig. 3). No bimodality is visible in iteration 1, but branches develop in iteration 2 or 3 (not shown) and are clear by iteration 4. Subject EM (middle) has a poorly defined upper branch, with a valley depth that is next-to-lowest in Fig. 5. Even for this subject, however, a lower branch forms by iteration 1 and becomes narrower between iterations 1 and 4. Subject CB (bottom panel) shows two branches even in iteration 1 which spread apart over the iterations. CB has a well-defined peak early in the bottom branch, near $\tau=0.15$, and a broad, ill-defined later peak.

The initial stimuli include both rising-tail and falling-tail

f_0 tracks, derived from the three basis contours. Nonetheless, we do not clearly observe independent traces of the third basis contour in the productions, nor do we see strong distinctions in the tails of the productions. Only one subject (AD, shown in Fig. 3) has three clear groups of contours. While several subjects show late bimodality around $\tau=0.95$, correlations between the tails and the centers of utterances complicate any interpretation.

In summary, the f_0 contours are typically bimodal. They become more strongly bimodal with every successive iteration. We see attractors for two of the three basis contours that we expected; overall our data do not show an association of contour H (rising tail) with an attractor.

D. Qualitative analysis

In the phonologist's (author BB) labeling three contours occurred particularly often: two peaks (35%), a hat pattern (40%), and one with a final rise (17%). Recall that 96% of the labels applied to the first imitations resulted in a previously recognized English contour and that the labels showed stability over iterations. The labeling therefore indicates a collapse in the very first iteration towards some stable attractor, whereas the acoustic analysis points to a gradual movement towards it. Linguistic categories could be assigned to most of the contours, even if—acoustically—a contour is not yet close to a stable attractor but is within its basin of attraction. This could be interpreted as a "magnet effect" for English intonation, analogous to Kuhl (1991) and Guenther and Gjaja (1996).

One interpretation of this result is that the attraction happens at the perceptual level rather than at the level of motor control. If so, then the phonologist perceived the speech by way of a mapping that is presumably similar to the subject's. The phonologist's perception of a contour from iteration 1 would then have undergone two mappings: one imposed by the subject on an initial stimulus and one imposed by the phonologist on the subject's response. The phonologist's perception would then effectively be one iteration further along than the acoustic analysis. Her perceived contours therefore would be closer to the (discrete) attractors.

Our findings are compatible with the results of Ladd and Morton (1997). They reported that subjects discriminate between numerous contours that (for instance) signal presence of emphasis but that subjects also readily assign contours to either an "emphasis" or "no emphasis" category. The corresponding feature of our results is the preservation of some of the difference information, as evidenced by the gradual collapse to the attractor.

One speculative way to explain the ease of conscious classification, the slow collapse of the acoustic properties toward attractors, and the discrimination of contours within categories is to assume that a heard f_0 contour is internally represented as an attractor plus a partially remembered difference between that attractor and the presentation. The phonologist may then suppress the difference information when classifying contours, retaining only the category information. We recognize, however, that there are other possible explanations of the three effects enumerated above.

IV. CONCLUSIONS

Although subjects can hear and imitate randomly generated contours that are not normally found in English, over several iterations their productions converge onto a limited set of distinct contours. These function mathematically as attractors and correspond to some (but not necessarily all) common English intonation contours. The results of iterative mimicry provide objective support for the existence of basic intonation patterns that act as attractors. Plausibly, the attractors are either a description of clusters of episodic traces/exemplars (Goldinger, 1998), or are a soft targets along the lines of (e.g., Kochanski *et al.*, 2003) that represent the position of the cluster center but allow some variability around it. Our results are not consistent with the hypothesis that subjects have available only a discrete phonological description of the intonation. Subjects actually perceive, remember, and use acoustic detail above and beyond what international phonology normally represents.

Our attractor contours have a parallel in the production and perception of vowels. In speech production, vowels are highly variable. Nevertheless, infants extract the vowels of their native language, and the adult vowel space contains regions of substantial size where a given vowel is heard reliably. The Perceptual Magnet Theory of Phonetic Learning (Guenther and Gjaja, 1996; Kuhl, 1991) shows how these local vowel regions emerge: a set of magnets acts as a prototype and warps the continuously variable vowel space around each of them.

While there are similarities with vowel magnets, intonation contours are dramatically extended in time. Vowels can be identified from brief bursts of sound (20–60 ms), while the attractors seen here are global properties of an entire utterance (about 1 s). Consequently, the mechanism that recognizes and processes these extended attractors may be different from the mechanism that handles vowels.

Stable attractors in the mapping between produced intonation and heard intonation have implications for language learning and development. To the extent that mimicry is involved in the normal use of language, the distribution of intonation contours that people hear should have a high density of examples near each attractor. Notably, children should be exposed to this distribution of contours. If the attractor contours in the children's mappings are based upon especially commonly heard examples, this could provide the mechanism for the stable transmission of the international aspects of language from one generation to the next in analogy to Maye *et al.* (2001) and Saffran *et al.* (1996).

ACKNOWLEDGMENTS

This work has been supported by the Oxford University Research Development Fund. We thank John Coleman for assistance with the stimulus generation and for tolerance and advice, and we thank Cindy Pribble Kochanski for helping to test the experimental software and design.

APPENDIX: SYNTHESIS PROCEDURES

A LPC analysis was made of each of the utterances from the original male and female speakers, using the Entropics

program `refcof`). Each analysis yielded 17 reflection coefficients; three voice source parameters were also obtained from `get_f0`.

For each of the candidate utterances, the parameters for the versions with falling and rising intonation were matched by dynamic time warping. This algorithm finds a monotonic mapping between the time axis of one utterance and the time axis of another that minimizes the mean magnitude of the vector difference between two sets of parameters. This yielded a time series of the 20 speech parameters that was half-way between that of the two starting utterances, both in terms of segment durations and of segmental properties like formant frequencies. Each of the candidate sentences was then synthesized from the intermediate parameter.

Bartlett, F. C. (**1932**). *Remembering* (Cambridge U. P., Cambridge).

Beckman, M., and Ayers, G. (1997). "Guidelines for ToBI labelling," Tech. Report, Linguistics Department, Ohio State University, http://ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf.

Brazil, D. (**1985**). *The Communicative Value of Intonation in English*, Discourse Analysis Monograph, No. 2 (Bleak House Books and English Language Research, Birmingham, UK).

Cruttenden, A. (**1997**). *Intonation*, 2nd ed. (Cambridge U. P., Cambridge).

Doll, J. J., and Thomas, D. R. (**1967**). "Effects of discrimination training on stimulus generalization for human subjects," J. Exp. Psychol. **75**, 508–512.

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

Francis, A. L., Ciocca, V., and Ng, B. K. C. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**, 1029–1044.

Geyer, C. J. (**1992**). "Practical Markov Chain Monte Carlo," Stat. Sci., 473–483.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Grabe, E., Rosner, B. S., García-Albea, J. E., and Zhou, X. (**2003**). "Perception of English intonation by English, Spanish, and Chinese listeners," Lang Speech **46**(4), 375–401.

Guenther, F. H., and Gjaja, M. N. (**1996**). "The perceptual magnet effect as an emergent property in neural map formation," J. Acoust. Soc. Am. **100**, 1111–1121.

Gussenhoven, C., and Rietveld, A. M. C. (**1997**). "Empirical evidence for the contrast between H^* and L^* in Dutch rising intonation contours," in *Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop*, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis.

Hirst, D., and Cristo, A. D., (Eds.) (**1998**). *Intonation Systems: A Survey of Twenty Languages* (Cambridge U. P., Cambridge, UK).

Hoeting, J. A., Madigan, D., Raferty, A., and Volinsky, C. T. (**1999**). "Bayesian Model averaging: A tutorial," Stat. Sci. **14**(4), 382–417; http://www.stat.colostate.edu/~jah/papers/statsci.pdf.

Holm, B., and Bailly, G. (**2002**). "Learning the hidden structure of intonation: Implementing various functions of prosody," in *Proceedings of Speech Prosody 2002*, http://www.isca-speech.org/archive/sp2002/sp02_399.pdf.

Jun, S.-A., (Ed.) (**2005**). *Prosodic Typology. The Phonology of Intonation and Phrasing* (Oxford U. P., Oxford, UK).

Kingdon, R. (**1958**). *The Groundwork of English Intonation* (Longman, London).

Kochanski, G., Shih, C., and Jing, H. (**2003**). "Quantitative measurement of prosodic strength in Mandarin," Speech Commun. **41**(4), 625–645; http://dx.doi.org/10.1016/S0167-6393(03)00100-6.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). "Loudness predicts prominence: Fundamental frequency lends little," J. Acoust. Soc. Am. **118**(2), 1038–1054; URL http://dx.doi.org/10.1121/1.1923349.

Kohler, K. J. (**1991**). "Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics," Arb. Inst. Phon. Univ. Kiel (AIPUK) **25**, 115–185.

Kuhl, P. (**1991**). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories," Percept. Psychophys. **50**, 93–107.

Ladd, D. R., and Morton, R. (**1997**). "The perception of intonational emphasis: continuous or categorical?," J. Phonetics **25**, 313–342.

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.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **61**, 379–388.

Loeb, D. F., and Allen, G. D. (**1993**). "Preschoolers' imitation of intonation countours," J. Speech Hear. Res. **36**(1), 4–13.

Maye, J., Werker, J. F., and Gerken, L. (**2001**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**, B101–B111.

Meltzoff, A. N., and Decety, J. (**2003**). "What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience," Philos. Trans. R. Soc. London **248**, 491–500.

Metropolis, N., Rosenbluth, A. E., Rosenbluth, M. N., Teller, A. H., and Teller, E. (**1953**). "Equations of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092.

O'Connor, J. D., and Arnold, G. F. (**1961**). *Intonation of Colloquial English* (Longman, London).

Pierrehumbert, J. B., and Hirschberg, J. (**1990**). "The meaning of intonation in the interpretation of discourse," in *Intentions in Communication*, edited by P. Cohen, J. Morgan, and M. Pollack (MIT, Cambridge, MA), Chap. 14, pp. 271–311.

Pierrehumbert, J. B., and Steele, S. A. (**1989**). "Categories of tonal alignment in English," Phonetica **46**, 181–196.

Plomp, R. (**2002**). *The Intelligent Ear: On the Nature of Sound Perception* (Erlbaum, Mahwah, NJ).

Remijsen, B., and van Heuven, V. J. (**1999**). "Gradient and categorical pitch dimensions in Dutch: Diagnostic test," in *Proc. of the 14th International Congress of the Phonetic Sciences (ICPhS)*, San Francisco, CA, pp. 1865–1868.

Repp, B. H., and Williams, D. R. (**1987**). "Categorical tendencies in self-imitating self-produced vowels," Speech Commun. **6**, 1–14.

Ryalls, J., Dorze, C. L., Lever, N., Quellet, L., and Larfeuil, C. (**1994**). "The effects of age and sex on speech intonation and duration for matched statements and questions in French," J. Acoust. Soc. Am. **95**(4), 2274–2276.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month-old infants," Science **274**, 1926–1928.

Scherer, K. R. (**1985**). "Vocal affect signalling: A comparative approach," in *Advances in the Study of Behavior*, edited by J. Rosenblatt, C. Beer, M.-C. Busnel, and P. Slater (Academic, New York), Vol. **15**, pp. 189–244.

Schouten, B., Gerrits, E., and von Hessen, A. (**2003**). "The end of categorical perception as we know it," Speech Commun. **41**, 71–80; http://dx.doi.org/10.1016/S0167-6393(02)00094-8.

Snow, D. (**1998**). "Children's imitation of intonation contours: Are rising tones more difficult than falling tones?," J. Speech Lang. Hear. Res. **41**(3), 576–587.

Steedman, M. (**2000**). "Information structure and the syntax-phonology interface," Linguist. Inq. **31**(4), 649–689.

Stevens, S. S. (**1971**). "Perceived level of noise by mark VII and decibels," J. Acoust. Soc. Am. **51**(2), (part 2) 575–602.

Talkin, D. (**1995**). "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, edited by W. B. Klein and K. K. Palival (Elsevier, Amsterdam).

Vallduvi, E., and Engdahl, E. (**1996**). "The linguistic realisation of information packaging," Linguistics **34**, 459–519.

Speech categorization in context: Joint effects of nonspeech and speech precursors

Lori L. Holt[a)]
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 20 March 2005; revised 10 March 2006; accepted 14 March 2006)

The extent to which context influences speech categorization can inform theories of pre-lexical speech perception. Across three conditions, listeners categorized speech targets preceded by speech context syllables. These syllables were presented as the sole context or paired with nonspeech tone contexts previously shown to affect speech categorization. Listeners' context-dependent categorization across these conditions provides evidence that speech and nonspeech context stimuli jointly influence speech processing. Specifically, when the spectral characteristics of speech and nonspeech context stimuli are mismatched such that they are expected to produce opposing effects on speech categorization the influence of nonspeech contexts may undermine, or even reverse, the expected effect of adjacent speech context. Likewise, when spectrally matched, the cross-class contexts may collaborate to increase effects of context. Similar effects are observed even when natural speech syllables, matched in source to the speech categorization targets, serve as the speech contexts. Results are well-predicted by spectral characteristics of the context stimuli. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195119]

PACS number(s): 43.71.An, 43.71.−k, 43.66.Ba, 43.66.−x [ALF] Pages: 4016–4026

I. INTRODUCTION

Context plays a critical role in speech categorization. Acoustically identical speech stimuli may be perceived as members of different phonetic categories as a function of the surrounding acoustic context. Mann (1980), for example, has shown that listeners' categorization of a series of speech stimuli ranging perceptually from /ga/ to /da/ is shifted toward more "ga" responses when these target syllables are preceded by /al/. The same stimuli are more often categorized as "da" when /ar/ precedes them. Such context-dependent phonetic categorization is a consistent finding in speech perception (e.g., Lindblom and Studdert-Kennedy, 1967; Mann and Repp, 1981; see Repp, 1982 for review).

Consideration of how to account for context-dependent speech perception highlights larger theoretical issues of how best to characterize the basic representational currency and processing characteristics of speech perception. Relevant to this interest, an avian species (Japanese quail, *Coturnix coturnix japonica*) has been shown to exhibit context-dependent responses to speech (Lotto *et al.*, 1997). Birds operantly trained to peck a lighted key in response to a /ga/ stimulus peck more robustly in later tests when test syllables are preceded by /al/. Correspondingly, birds trained to peck to /da/ peck most vigorously to test stimuli when the are preceded by /ar/. Thus, birds exhibit shifts in pecking behavior contingent on preceding context analogous to context-dependent human speech categorization. The birds had no previous experience with speech, so their behavior cannot be explained on the basis of learned covariation of acoustic attributes across contexts or on the basis of existing phonetic categories. It is also unlikely that quail have access to specialized speech processes or knowledge of the human vocal tract. The parallels between quail and human behavior suggest a possible role for general auditory processing, not specific to speech or dependent upon extensive experience with the speech signal, in context-dependent speech perception.

In accord with the hypothesis that general, rather than speech-specific, processes play a role in context-dependent speech perception there is evidence that nonspeech acoustic contexts affect speech categorization by human listeners. Following the findings of Mann (1980), Lotto and Kluender (1998) synthesized two sine-wave tones, one with a higher frequency corresponding to the third formant (F3) offset frequency of /al/ and the other with a lower frequency corresponding to the /ar/ F3 offset frequency. When these nonspeech stimuli preceded a /ga/ to /da/ target stimulus series like that studied by Mann (1980), speech categorization was influenced by the precursor tones. Listeners more often categorized the syllables as "ga" when they were preceded by the higher-frequency sine-wave tone modeling /al/. The same stimuli were more often categorized as "da" when the tone modeling /ar/ preceded them. Thus, nonspeech stimuli mimicking very limited spectral characteristics of speech contexts also influence speech categorization.

Nonspeech-elicited context effects on speech categorization appear to be a general phenomenon. Holt (1999; Holt and Lotto, 2002) reports that sine-wave tones or single formants situated at the second formant (F2) frequency of /i/ versus /u/ shift categorization of syllables ranging perceptually from /ba/ to /da/ in the same manner as the vowels they model. Likewise, flanking nonspeech frequency-modulated glides that follow the F2 formant trajectories of /bVb/ and /dVd/ syllables influence categorization of the intermediate vowel (Holt *et al.*, 2000). A number of other studies demonstrate interactions of nonspeech context and speech percep-

a)Electronic mail: lholt@andrew.cmu.edu

 0001-4966/2006/119(6)/4016/11/$22.50

tion (Fowler *et al.*, 2000; Kluender *et al.*, 2003; Watkins and Makin, 1994, 1996a, 1996b) and the effects appear to be reciprocal. Stephens and Holt (2003) report that preceding /al/ and /ar/ syllables modulate perception of following nonspeech stimuli. Follow-up studies have demonstrated that listeners are unable to relate the sine-wave tone precursors typical of these studies to the phonetic categories the tones model (Lotto, 2004); context-dependent speech categorization is elicited even with nonspeech precursors that are truly perceived as nonspeech events.

There is evidence that even temporally nonadjacent nonspeech precursors can influence speech categorization. Holt (2005) created "acoustic histories" composed of 21 sine-wave tones sampling a distribution defined in the acoustic frequency dimension. The acoustic histories terminated in a neutral-frequency tone that was shown to have no effect on speech categorization. In this way, the context immediately adjacent to the speech target in time was constant across conditions. The mean frequency of the acoustic histories differentiated conditions, with distribution means approximating the tone frequencies of Lotto and Kluender (1998). Despite their temporal nonadjacency with speech targets, the nonspeech acoustic histories had a significant effect on categorization of members of a following /ga/ to /da/ speech series. In line with previous findings, the higher-frequency acoustic histories resulted in more "ga" responses whereas the lower-frequency acoustic histories led to more "da" responses. These effects were observed even when as much as 1.3 s of silence or 13 repetitions of the neutral tone separated the acoustic histories and the speech targets in time.

In each of the cases for which effects of nonspeech contexts on speech categorization have been observed, the nonspeech contexts model limited spectral characteristics of the speech contexts. As simple pure tones or glides, they do not possess structured information about articulatory gestures. Moreover, even the somewhat richer nature of the acoustic history tone contexts of Holt (2005) are far removed from the stimuli that may be perceived as speech in sine-wave speech studies (e.g., Remez *et al.* 1994). The commonality shared between the tones composing the acoustic histories and sine-wave speech is limited to the fact that both make use of sinusoids. The tonal sine-wave speech stimuli are composed of three or four concurrent time-varying sinusoids, each mimicking the center frequency and amplitude of a natural vocal resonance measured from a real utterance. Thus, the sine-wave replicas that may give rise to speech percepts possess an overall acoustic structure that much more closely mirrors the speech spectrum it models. By contrast, the single sine-waves of, for example, Lotto and Kluender (1998) or the sequences of sine waves of Holt (2005) are far more removed from the precise time-varying characteristics of speech. The tones composing the acoustic histories of Holt (2005) are single sinusoids of equal amplitude, separated in time (not continuous), and randomized on a trial-by-trial basis. The nonspeech contexts provide neither acoustic structure consistent with articulation nor acoustic information sufficient to support phonetic labeling (see Lotto, 2004). What they do share with the speech contexts they model is a very limited resemblance to the *spectral* information that differentiates, for example, the /al/ from /ar/ contexts that have been shown to influence speech categorization (Mann, 1980).

The directionality of the context-dependence is likewise predictable from this spectral information. Across the observations of context-dependent speech categorization for speech and nonspeech contexts, the pattern of context-dependent categorization is spectrally contrastive (Holt, 2005; Lotto *et al.*, 1997; Lotto and Kluender, 1998); precursors with acoustic energy in higher frequency regions (whether speech or nonspeech, e.g., /al/ or nonspeech sounds modeling the spectrum of /al/) shift categorization toward the speech category characterized by lower-frequency acoustic energy (i.e., /ga/) whereas lower-frequency precursors (/ar/ or nonspeech sounds modeling /ar/) shift categorization toward the higher-frequency alternative (i.e., /da/). The auditory perceptual system appears to be operating in a manner that serves to emphasize spectral change in the acoustic signal. Contrastive mechanisms are a fundamental characteristic of perceptual processing across modalities. General mechanisms of auditory processing that produce spectral contrast may give rise to the results observed for speech and nonspeech contexts in human listeners with varying levels and types of language expertise (Mann, 1986; Fowler *et al.*, 1990) and in quail subjects (Lotto *et al.*, 1997). Neural adaptation and inhibition are simple examples of neural mechanisms that exaggerate contrast in the auditory system (Smith, 1979; Sutter *et al.*, 1999), but others exist at higher levels of auditory processing (see e.g., Delgutte, 1996; Ulanovsky *et al.*, 2003; 2004) that produce contrast without a loss in sensitivity (Holt and Lotto, 2002). The observation of nonspeech context effects on speech categorization when context and target are presented to opposite ears (Holt and Lotto, 2002; Lotto *et al.*, 2003) and findings demonstrating effects of nonadjacent nonspeech context on speech categorization (Holt, 2005) indicate that the mechanisms are not solely sensory.[1] Moreover, there is evidence that mechanisms producing spectral contrast may operate over multiple time scales (Holt, 2005; Ulanovsky *et al.*, 2003, 2004).

By this general perceptual account, speech- and nonspeech-elicited context effects emerge from common processes that are part of general auditory processing. These mechanisms are broadly described as spectrally contrastive in that they emphasize spectral change in the acoustic signal, independent of its classification as speech or nonspeech or whether the signal carries information about speech articulation. So far, observed effects have been limited to the influence of speech *or* nonspeech contexts on speech categorization (or, conversely, the effects of speech contexts on nonspeech perception, Stephens and Holt, 2003). However, an account that relies upon spectral contrast makes strong directional predictions about context-dependent speech categorization in circumstances in which both speech and nonspeech contexts are present. Specifically, this account predicts that when both speech and nonspeech are present as context, their effects on speech categorization will be dictated by their spectral characteristics such that they may either cooperate or conflict in their direction of influence on speech categorization as a function of how they are paired. If the speech and nonspeech contexts are matched in the distri-

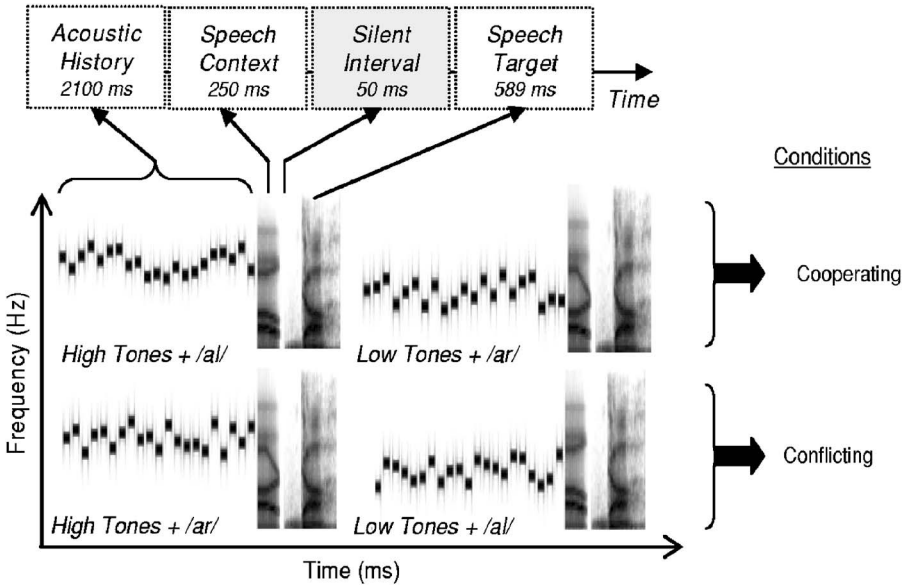


FIG. 1. At the top, an illustration displays the elements of each stimulus. Representative spectrograms (on time × frequency axes) below show example stimuli from Experiment 1 conditions. Stimuli from the Cooperating condition (top row) are composed of spectrally matched speech and nonspeech contexts that have been shown previously to shift speech categorization in the same direction. Examples of Conflicting condition stimuli for which spectrally mismatched nonspeech and speech precursors have opposing effects on speech categorization are shown in the bottom row.

bution of spectral energy that they possess such that they are expected to shift speech categorization in the same direction, then nonspeech may collaborate with speech to produce greater effects of context than observed for speech contexts alone. Conversely, when nonspeech and speech contexts possess spectra that push speech categorization in opposing directions, nonspeech contexts should be expected to lessen the influence of speech contexts on speech categorization. As a means of empirically examining the hypotheses arising from this account, the present experiments examine speech categorization when *both* speech and nonspeech signals serve as acoustic context, specifically investigating the degree to which they may jointly influence speech categorization.

II. EXPERIMENT 1

The aim of this study thus is to assess the relative influence of speech and jointly presented nonspeech contexts on speech categorization. Experiment 1 examines speech categorization of a /ga/ to /da/ syllable series across three contexts: (1) preceding /al/ and /ar/ syllables; (2) the same speech syllables paired with spectrally matched nonspeech acoustic histories (as described by Holt, 2005) that shift speech categorization in the same direction (e.g., High Mean acoustic histories paired with /al/); (3) the same speech syllables paired with spectrally mismatched nonspeech acoustic histories that shift speech categorization in opposing directions (e.g., Low Mean acoustic histories paired with /al/). Whereas the speech contexts remain consistent across conditions, the nonspeech contexts vary. Thus, if speech and nonspeech contexts fail to jointly influence speech categorization there will be no significant differences in speech categorization across conditions and, as in previous studies, speech targets preceded by /al/ will be more often categorized as "ga" than the same targets preceded by /ar/. If, however, the two sources of acoustic context mutually influence speech categorization as predicted by a general perceptual/cognitive account of context effects in speech perception then the observed context effects will vary across conditions and the relative influence of each context source on speech categorization can be assessed.

A. Methods

1. Participants

Ten adult monolingual English listeners recruited from the Carnegie Mellon University community participated in return for a small payment or course credit. All participants reported normal hearing.

2. Stimuli

Stimulus design is schematized in Fig. 1. For each stimulus an acoustic history composed of 21 sine-wave tones preceded a speech syllable context stimulus, a 50-ms silent interval, and a speech target drawn from a stimulus series varying perceptually from /ga/ to /da/.

a. Speech. Speech target stimuli were identical to those described previously (Holt, 2005; Wade and Holt, 2005). Natural tokens of /ga/ and /da/ spoken in isolation were digitally recorded from an adult male monolingual English speaker (CSL, Kay Elemetrics; 20-kHz sample rate, 16-bit resolution). From a number of natural productions, one /ga/ and one /da/ token were selected that were nearly identical in spectral and temporal properties except for the onset frequencies of F2 and F3. LPC analysis was performed on each of the tokens and a nine-step sequence of filters was created (Analysis-Synthesis Laboratory, Kay Elemetrics) such that the onset frequencies of F2 and F3 varied approximately linearly between /g/ and /d/ endpoints. These filters were excited by the LPC residual of the original /ga/ production to create an acoustic series spanning the natural /ga/ and /da/ end points in approximately equal steps. Each stimulus was 589 ms in duration. The series was judged by the experimenter to comprise a gradual shift between natural-sounding /ga/ and /da/ tokens and this impression was confirmed by regular shifts in phonetic categorization across the series by participants in the Holt (2005) and Wade and Holt (2005)

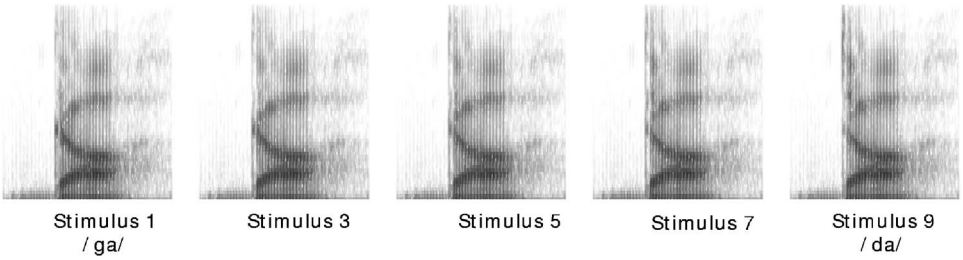


FIG. 2. Spectrograms of the odd-numbered stimuli along the nine-step /ga/ to /da/ series that served as categorization targets in Experiments 1 and 2.

studies. These speech series members served as categorization targets for each experimental condition. Spectrograms of odd-number series stimuli are shown in Fig. 2.

In addition, there were two speech context stimuli. These 250-ms syllables corresponded perceptually to /al/ and /ar/ and were composed of a 100-ms steady-state vowel followed by a 150-ms linear formant transition. Stimuli were synthesized using the cascade branch of the Klatt (1980) synthesizer. These stimuli were identical to those shown in earlier reports to produce spectrally contrastive context effects on perception of speech (Lotto and Kluender, 1998) and nonspeech (Stephens and Holt, 2003). Lotto and Kluender (1998) provide full details of stimulus synthesis.

b. Nonspeech. Acoustic histories were created as described by Holt (2005). Each acoustic history was composed of 21 70-ms sine-wave tones (30-ms silent intervals) with unique frequencies. Distributions' mean frequencies (1800 and 2800 Hz) were chosen based on the findings of Lotto and Kluender (1998), who demonstrated that single 1824 versus 2720 Hz tones produce a spectrally contrastive context effect on speech categorization targets varying perceptually from /ga/ to /da/. "Low Mean" acoustic histories were composed of 1300–2300 Hz tones ($M=1800$ Hz, 50-Hz steps). "High Mean" acoustic histories possessed tones sampling 2300–3300 Hz ($M=2800$ Hz, 50-Hz steps).

To minimize effects elicited by any particular tone ordering, acoustic histories were created by randomizing the order of the 21 tones on a trial-by-trial basis. Each trial was unique; acoustic histories within a condition were distinctive in surface acoustic characteristics, but were statistically consistent with other stimuli drawn from the distribution defining the nonspeech context. Thus, any influence of acoustic histories on speech categorization is indicative of listeners' sensitivity to the long-term spectral distribution of the acoustic history and not merely to the simple acoustic characteristics of any particular segment (for further discussion see Holt, 2005).

Tones comprising the acoustic histories were synthesized with 16-bit resolution and sampled at 10 kHz using MATLAB (Mathworks, Inc.). Linear onset/offset amplitude ramps of 5 ms were applied to all tones. Target speech stimuli were digitally down-sampled from their recording rate of 20–10 kHz and both tones and speech tokens were digitally matched to the rms energy of the /da/ end point of the target speech series.

As discussed in Sec. I, very broad interpretation of the kind of acoustic energy that may carry articulatory information may cause concern that the High and Low mean acoustic histories could serve as information about articulatory events and perhaps lead listeners to identify the nonspeech acoustic histories phonetically. To allay this concern, 10 monolingual English participants who reported normal hearing were tested in a pilot stimulus test. These participants did not serve as listeners in any of the reported experiments and had not participated in experiments of this sort before. These listeners identified the High and Low mean acoustic histories as "al" or "ar" in the context of the following speech syllable pairs described above. If the limited spectral information that the acoustic histories model from the /al/ and /ar/ contexts serves as information about articulatory events, we should expect High mean acoustic histories to elicit more "al" responses and Low mean acoustic histories to elicit more "ar" responses. This was not the case. Listeners' phonetic labeling of the High versus Low mean acoustic histories as "al" was not greater for the High mean acoustic histories ($M_{\text{High}}=51.1$, $SE=0.52$) than Low mean acoustic histories ($M_{\text{Low}}=51.0$, $SE=1.19$; $t<1$ in a paired-samples t-test).

c. Stimulus construction. Two sets of stimuli were constructed from these elements. To create the hybrid nonspeech/speech contexts preceding the speech targets, each of the nine /ga/ to /da/ target stimuli was appended to the /al/ and /ar/ speech contexts with a 50-ms silent interval separating the syllables. Each of the resulting 18 disyllables was appended to two nonspeech contexts, one an acoustic history defined by the High Mean distribution and the other an acoustic history with a Low Mean. This pairing of disyllables with acoustic histories was repeated 10 times, with a different acoustic history for each repetition. This resulted in 360 unique stimuli, exhaustively pairing /al/ and /ar/ speech contexts with High and Low mean nonspeech contexts and the nine target speech series stimuli across 10 repetitions. A second set of stimuli with only speech contexts preceding the speech targets also was created; /al/ and /ar/ stimuli were appended to each of the speech target series members with a 50-ms interstimulus silent interval for a total of 18 stimuli. These stimuli were presented 10 times each during the experiment.

3. Design and procedure

The pairing of speech and nonspeech contexts in stimulus creation yielded the two experimental conditions illustrated in Fig. 1. Stimuli making up the Conflicting condition possessed acoustic histories and speech context syllables that have been shown to have opposing effects on speech categorization (Holt, 2005; Lotto and Kluender, 1998; Mann, 1980). The Cooperating condition was made up of stimuli possessing speech and nonspeech precursor contexts that

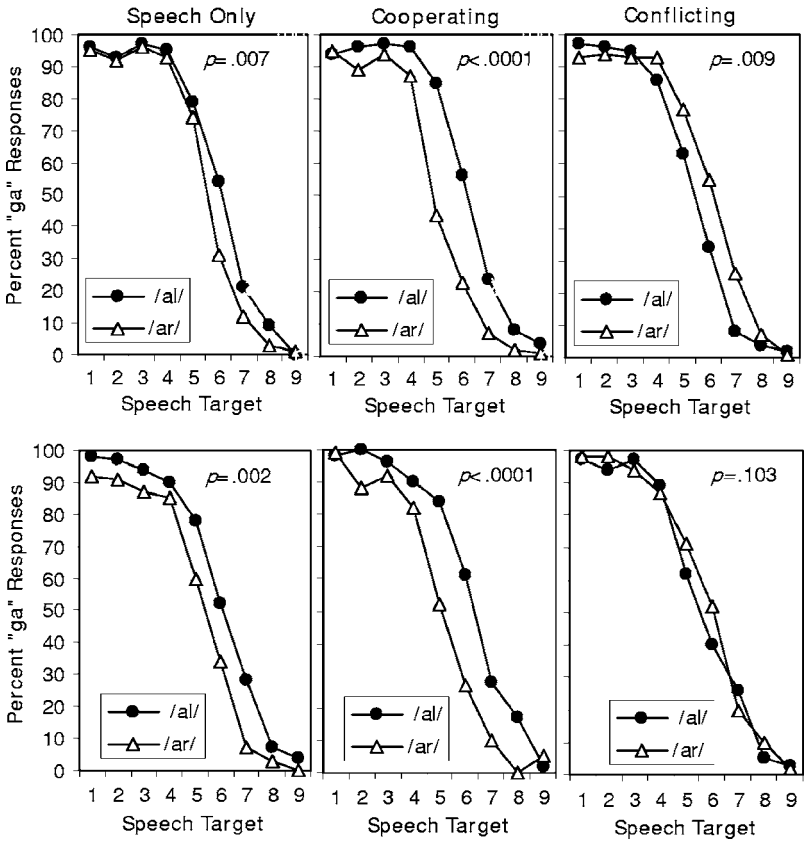


FIG. 3. Mean "ga" responses to speech series stimuli for Experiment 1 (top panel) and Experiment 2 (bottom panel). The "Speech Only" panels present categorization data for /al/ and /ar/ contexts. The other two panels illustrate categorization when the same stimuli are preceded by High and Low Mean acoustic histories and the /al/ or /ar/ precursors. In the "Cooperating" condition, speech and nonspeech precursors are expected to shift categorization in the same direction (High+/*al*/, Low +/*ar*/). In the "Conflicting" condition, acoustic histories and speech precursors exert opposite effects on speech categorization (Low+/*al*/, High+/*ar*/).

shift speech categorization in the same direction. Note that these pairings can also be described in terms of the spectral characteristics of the component context stimuli because spectral characteristics well-predict the directionality of context effects on speech categorization (e.g., Holt, 2005; Lotto and Kluender, 1998). For example, High Mean acoustic histories were matched with /al/ (also possessing greater high-frequency acoustic energy) in the spectrally matched Cooperating condition and with /ar/ (with greater low-frequency energy) in the spectrally mismatched Conflicting condition.

Seated in individual sound-attenuated booths, listeners categorized the speech target of each stimulus by pressing electronic buttons labeled "ga" and "da." Listeners completed two blocks in a single session; the order of the blocks was counterbalanced. In one block, the hybrid nonspeech plus speech contexts preceded the speech targets. In this block, stimulus presentation was mixed across the Conflicting and Cooperating conditions. In the other (Speech Only) block, participants heard only /al/ or /ar/ preceding the speech targets. Thus, each listener responded to stimuli from all three conditions.

Acoustic presentation was under the control of Tucker Davis Technologies System II hardware; stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL(A).

B. Results

Results were analyzed in terms of average percent "ga" responses across stimulus repetitions and are plotted in the top row of Fig. 3. The nonoverlapping categorization curves illustrated in each of the top panels of Fig. 3 are indicative of an influence of context for each condition (see also the mar-

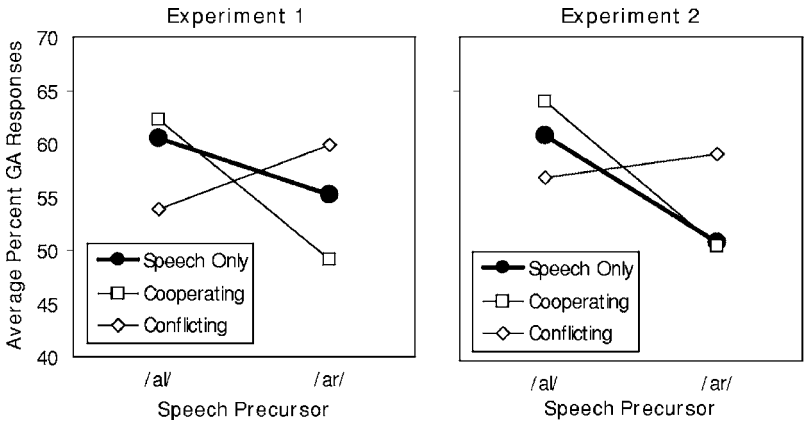


FIG. 4. Marginal means across condition and experiment.

TABLE I. Category boundaries were estimated for each participant's response to each condition of the experiment. The mean probit boundary across participants is presented in terms of the stimulus step across the nine-step /ga/ to /da/ categorization target series. The results parallel those of the ANOVA analyses across the speech stimulus series reported in the text.

Experiment	Condition	Precursor	Mean probit boundary	Standard error	*t*-test
1	Speech Only	/al/	7.0	0.21	$t(9)=3.13$, $p=0.01$
		/ar/	6.46	0.27	
	Cooperating	High+/al/	7.16	0.23	$t(9)=3.71$, $p=0.005$
		Low+/ar/	5.96	0.29	
	Conflicting	Low+/al/	6.12	0.21	$t(9)=3.76$, $p=0.005$
		High+/ar/	6.82	0.25	
2	Speech Only	/al/	6.79	0.24	$t(9)=3.59$, $p=0.01$
		/ar/	5.97	0.36	
	Cooperating	High+/al/	7.21	0.23	$t(9)=5.94$, $p<0.0001$
		Low+/ar/	5.98	0.24	
	Conflicting	Low+/al/	6.70	0.22	$t(9)=0.3$, $p=0.77$
		High+/ar/	6.64	0.19	

ginal means plotted in Fig. 4). Critically, although the immediately preceding speech context was constant across conditions, the observed context effects were not identical. Repeated-measures analysis of variance results are described in the following. Probit boundary analysis (Finney, 1971) of participants' category boundaries across conditions reveals the same pattern of results. The results of these analyses are provided in Table I.

1. Speech Only condition

The average percent "ga" responses across participants were submitted to a 2×9 (Context×Target Speech Stimulus) repeated measures ANOVA. This analysis revealed a significant effect of Context, $F(1,9)=12.12$, $p=0.007$, $\eta_p^2=0.574$. Consistent with earlier findings (Lotto and Kluender, 1998; Mann, 1980), listeners categorized speech targets preceded by /al/ as "ga" significantly more often ($M=60.44$, $SE=2.86$, here and henceforth, means refer to "ga" responses averaged across target speech stimuli and participants) than the same targets preceded by /ar/ ($M=55.22$, $SE=2.57$). These data confirm that, on their own, the speech context precursors have a significant effect on categorization of neighboring speech targets. Probit boundary values are presented in Table I.

2. Cooperating condition

A 2×9 (Context×Target Speech Stimulus) repeated measures ANOVA revealed that there was also a significant effect of Cooperating nonspeech/speech contexts on speech categorization, $F(1,9)=40.22$, $p<0.0001$, $\eta_p^2=0.817$. As would be expected from the influence that speech and nonspeech contexts elicit independently (Lotto and Kluender, 1998; Holt, 2005), the effect observed in the Cooperating condition was spectrally contrastive; categorization was shifted in the same direction as in the Speech Only condition. When listeners heard speech targets preceded by High Mean acoustic histories paired with /al/, they more often categorized the targets as "ga" ($M=62.22$, $SE=2.05$) than when the same targets were preceded by Low Mean acoustic histories paired with /ar/ ($M=49.11$, $SE=2.34$).

The primary aim of this study was to examine potential joint effects of speech and nonspeech acoustic contexts in influencing speech target categorization. A $2\times2\times9$ (Condition×Context×Target Speech Stimulus) repeated measures ANOVA of the categorization patterns of the Speech Only condition versus those of the Cooperating condition indicates that when speech and nonspeech contexts are spectrally matched such that they are expected to influence speech categorization similarly, they collaborate to produce an even greater context effect on speech target categorization ($M_{\text{High+/al/}}=62.22$ vs $M_{\text{Low+/ar/}}=49.11$) than do the speech targets on their own ($M_{/al/}=60.44$ vs $M_{/ar/}=55.22$), as indicated by a significant Context by Condition interaction, $F(1,9)=6.42$, $p=0.03$, $\eta_p^2=0.416$.

3. Conflicting Condition

A 2×9 (Context×Target Speech Stimulus) repeated measures ANOVA of responses to Conflicting condition stimuli revealed that when the spectra of speech and nonspeech contexts predicted opposing effects on speech categorization, there was also a significant effect of context, $F(1,9)=25.97$, $p=0.001$, $\eta_p^2=0.743$. Note, however, the direction of this effect. Listeners more often categorized target syllables as "ga" when they were preceded the High Mean acoustic histories paired with /ar/ speech precursors (% "ga" responses: $M_{\text{High+/ar/}}=59.89$, $SE=2.41$ vs $M_{\text{Low+/al/}}=49.11$, $SE=2.34$). In this example, the /ar/ speech context independently predicts more "da" responses (Mann, 1980) whereas the High Mean nonspeech acoustic histories independently predict more "ga" responses (Holt, 2005). Listeners more often responded "ga," following the expected influence of the nonspeech context rather than that of the speech context that immediately preceded the speech targets. These results indicate that when the spectra of nonspeech and speech contexts are put in conflict, the influence of temporally nonad-

jacent nonspeech context may be robust enough even to undermine the expected influence of temporally adjacent speech contexts.

Of note, a $2\times2\times9$ (Condition $\times$ Context $\times$ Target Speech Stimulus) repeated measures ANOVA comparing the Conflicting condition to the Speech Only condition revealed no main effect of Context, $F(1,9)=2.98$, $p=0.119$, $\eta_p^2=0.249$, but a significant Condition by Context interaction, $F(8,72)=83.17$, $p<0.0001$, $\eta_p^2=0.902$. This indicates that the context effect produced by the speech contexts plus conflicting nonspeech contexts was statistically equivalent in magnitude, although opposite in direction, to that produced by the speech contexts alone.

4. Comparison of Cooperating vs Conflicting conditions

The relative contributions of speech and nonspeech contexts can be assessed with a $2\times2\times9$ (Condition $\times$ Context $\times$ Target Speech Stimulus) repeated measures ANOVA comparing the effects of nonspeech/speech hybrid contexts across Cooperating and Conflicting conditions. This analysis reveals an overall main effect of Context (context was coded in terms of the nonspeech segment of the precursor), $F(1,9)=37.207$, $p<0.0001$, $\eta_p^2=0.805$, such that listeners more often labeled speech targets as "ga" when nonspeech precursors were drawn from the High Mean acoustic history distribution ($M=61.06$, $SE=2.01$) than the Low Mean distribution ($M=51.50$, $SE=2.09$). The contribution of the speech contexts to target syllable categorization is reflected in this analysis by the significant Condition by Acoustic History interaction, $F(1,9)=9.69$, $p=0.01$, $\eta_p^2=0.518$. With /al/ precursors, targets were somewhat more likely to be categorized as "ga" ($M=58.056$, $SE=1.9$) whereas with /ar/ precursors the same stimuli were less likely to be categorized as "ga" ($M=54.50$, $SE=2.13$). Thus, across conditions there is evidence of the joint influence of speech and nonspeech contexts. Moreover, the directionality of the observed effects is well-predicted by the spectral characteristics of the speech and nonspeech contexts.

C. Discussion

The percept created by the experiment 1 hybrid nonspeech/speech stimuli is one of rapidly presented tones preceding a bi-syllabic speech utterance. One could easily describe these nonspeech precursors as extraneous to the task of speech categorization and, indeed, listeners were not required to make any explicit judgments about them during the perceptual task. The task in this experiment was *speech perception*. Yet, even in these circumstances nonspeech contexts contributed to speech categorization. Speech does not appear to have a privileged status in producing context effects on speech categorization, even when afforded the benefit of temporal adjacency with the target of categorization.

Although general perceptual/cognitive accounts of speech perception are most consistent with these effects and can account for the directionality of the observed context effects, it is nonetheless surprising even from this theoretical perspective that the effect of nonspeech contexts is so robust. The results run counter to modular accounts that would suggest that there are special-purpose mechanisms for processing speech that are informationally encapsulated and therefore impenetrable to influence by nonlinguistic information (Liberman *et al.*, 1967; Liberman and Mattingly, 1985). The very simple sine-wave tones that comprised the nonspeech contexts are among the simplest of acoustic signals. To consider them information for speech perception by a speech-specific module would require a module so broadly tuned as to be indistinguishable from more interactive processing schemes. The results of Experiment 1 also are difficult to reconcile with a direct realist perspective on speech perception. The direct realist interpretation of the categorization patterns observed in the Speech Only condition is that the speech contexts provide information relevant to parsing the dynamics of articulation (Fowler, 1986; Fowler and Smith, 1986; Fowler *et al.*, 2000). It is unclear from a direct realist perspective why, in the presence of clear speech contexts providing information about articulatory gestures, listeners would be influenced by nonspeech context sounds at all, let alone be more influenced by the nonspeech contexts than the speech contexts in the Conflicting condition. It does not appear that context must carry structured information about articulation to have an impact on speech processing.

III. EXPERIMENT 2

The stimuli created for Experiment 1 were constructed as a compromise among stimuli used in previous experiments investigating speech and nonspeech context effects. The /ga/ to /da/ speech target series of Holt (2005) was chosen for its naturalness in an effort to provide the most conservative estimate of context-dependence (synthesized or otherwise degraded speech signals are typically thought to be more susceptible to contextual influence). The synthetic /al/ and /ar/ contexts were taken from the stimulus materials of Lotto and Kluender (1998) because they produce a robust influence on speech categorization along a /ga/ to /da/ series (see also Stephens and Holt, 2003). Nonetheless, there are stimulus differences originating from the synthetic nature of the /al/ and /ar/ speech contexts of Experiment 1 and the more natural characteristics of the speech targets. This could lead the two sets of speech materials to be perceived as originating from different sources. If this was the case, the independence of the sources should reduce or eliminate articulatory gestural information relevant to compensating for intraspeaker effects of coarticulation (a within-speaker phenomenon) via gestural parsing. Although previous research has provided evidence of cross-speaker phonetic context effects (Lotto and Kluender, 1998), it may nonetheless be argued that Experiment 1 does not provide the most conservative test of nonspeech/speech context effects because of the possible perceived difference in speech source across syllables.

Therefore, Experiment 2 was conducted in the same manner as Experiment 1, but using natural /al/ and /ar/ productions recorded from the same speaker that produced the end point stimuli of the /ga/ to /da/ speech target stimulus series. The experiment thus serves as both a replication of

the findings of Experiment 1 and an opportunity to investigate whether the influence of nonspeech context on speech categorization is robust enough to persist even when speech contexts and targets originate from the same source.

A. Methods

1. Participants

Ten adult monolingual English listeners, none of whom participated in Experiment 1, received a small payment or course credit for volunteering. All participants were recruited from the Carnegie Mellon University community and reported normal hearing.

2. Stimuli

Stimulus design was identical to that of Experiment 1, except that the speech context stimuli were digitally recorded (20-kHz sample rate, 16-bit resolution) natural utterances of /al/ and /ar/ spoken in isolation by the same speaker who recorded the natural speech end points of the target stimulus series. The 350-ms syllables were down-sampled to 10 kHz and matched in rms energy to the /da/ end point of the target stimulus series. These syllables served as the speech contexts in the stimulus construction protocol described for Experiment 1.

3. Design and Procedure

The design, procedure, and apparatus were identical to those of Experiment 1.

B. Results

The results of Experiment 2 are shown in the bottom row of Fig. 3. Marginal means are plotted in Fig. 4. Probit boundary values are presented in Table I.

1. Speech Only condition

Consistent with the findings of Experiment 1, there was a significant influence of preceding /al/ and /ar/ on speech target categorization. A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA confirmed that listeners categorized speech targets preceded by /al/ as "ga" significantly more often ($M=60.89$, $SE=2.28$) than the same targets following /ar/ ($M=51.00$, $SE=3.4$), $F(1,9)=18.426$, $p=0.002$, $\eta_p^2=0.672$. Thus, natural /al/ and /ar/ recordings matched to the target source produced a significant context effect on categorization of the speech targets.

One potential concern about the use of synthesized speech contexts in Experiment 1 was that a perceived change in talker may have reduced observed effects of speech context. However, comparison of the influence of the synthesized versus naturally produced speech contexts on categorization of the speech targets with a cross-experiment $2\times2\times9$ (Experiment$\times$Context$\times$Target Speech Stimulus) mixed model ANOVA with Experiment as a between-subjects factor, did not reveal a significant difference in the context effects produced by the /al/ and /ar/ stimuli of Experiments 1 and 2, $F(1,18)=2.88$, $p=0.11$, $\eta_p^2=0.138$.

2. Cooperating condition

The primary question of interest is whether nonspeech contexts influence speech categorization even in the presence of adjacent speech signals originating from the same source. A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA supports what is illustrated in the bottom row of Fig. 2. There was a significant spectrally contrastive effect of the cooperating, spectrally matched, speech and nonspeech contexts, $F(1,9)=76.21$, $p<0.0001$, $\eta_p^2=0.894$, such that listeners more often categorized speech targets as "ga" when High Mean nonspeech precursors and /al/ preceded them ($M=64.00$, $SE=2.19$) than when Low Mean nonspeech precursors and /ar/ preceded them ($M=50.56$, $SE=2.54$).

An additional $2\times2\times9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA examined the context effects across the Speech Only and Cooperating conditions of Experiment 2. Of note, although the mean difference between conditions was greater for the Cooperating condition ($M_{(\text{High+/al/})\text{-}(\text{Low+/ar/})}=13.44\%$) than the Speech Only condition ($M_{/\text{al/-/ar/}}=9.89\%$), this difference was not statistically reliable, $F(1,9)=2.52$, $p=0.147$, $\eta_p^2=0.219$. This differs from Experiment 1, for which speech and nonspeech contexts collaborated in the Cooperating condition to produce a greater effect of context on speech categorization than did the speech contexts alone.

3. Conflicting condition

An analogous analysis was conducted across the Speech Only and Conflicting conditions, revealing that the categorization patterns observed for the Conflicting condition were significantly different than those found for the Speech Only condition, $F(1,9)=18.63$, $p=0.002$, $\eta_p^2=0.674$. A 2×9 (Context$\times$Target Speech Stimulus) repeated measures ANOVA showed that, contrary to the robust effect of speech contexts in the Speech Only condition, there was no effect of hybrid nonspeech/speech contexts in the Conflicting condition, $F(1,9)=3.29$, $p=0.103$, $\eta_p^2=0.267$ ($M_{\text{Low+/al/}}=57.00$, $SE=2.52$ vs $M_{\text{High+/ar/}}=59.11$, $SE=2.52$). The presence of spectrally mismatched nonspeech contexts effectively neutralized the influence of the natural speech precursors.

4. Comparing Cooperating and Conflicting conditions

A comparison of the patterns of categorization for the hybrid nonspeech/speech context conditions with a $2\times2\times9$ (Condition$\times$Context$\times$Target Speech Stimulus) repeated measures ANOVA revealed a main effect of Context (entered into the analysis in terms of the nonspeech characteristics of the context) such that listeners more often labeled the speech targets as "ga" when the nonspeech context was drawn from a distribution with a High Mean frequency ($M=60.50$, $SE=2.26$) than when it was drawn from a distribution with a Low Mean frequency ($M=54.83$, $SE=2.47$), $F(1,9)=25.61$, $p=0.001$, $\eta_p^2=0.740$. In this analysis, the contribution of the speech contexts to speech target categorization was reflected by a significant Condition by Context interaction, $F(1,9)=99.10$, $p<0.0001$, $\eta_p^2=0.917$ such that listeners more often labeled targets as "ga" when the precur-

sor syllable was /al/ ($M=61.56$, $SE=2.25$) than when it was /ar/ ($M=53.78$, $SE=2.42$). Thus, both speech and nonspeech contexts contributed to the categorization responses observed in the hybrid context conditions of Experiment 2.

C. Discussion

The overall pattern of results of Experiment 2 confirms that speech and nonspeech contexts jointly influenced speech categorization, even when the natural speech contexts were matched to the categorization targets in source. Of note, however, the influence of the nonspeech contexts in the presence of the natural speech contexts was less dramatic than were the effects observed when the same nonspeech contexts were paired with synthesized speech syllables in Experiment 1. Contrary to the findings of Experiment 1, the nonspeech precursors did not collaborate with the natural speech contexts of Experiment 2 to produce a context effect significantly greater than that elicited by the natural speech syllables alone. Moreover, although there was strong evidence of joint nonspeech/speech context effects in the Experiment 2 Conflicting condition, the influence of the nonspeech was not so strong as to overpower the natural speech context and reverse the observed context effect as it did in Experiment 1. These more modest patterns of interaction may be due to the somewhat stronger effect of context elicited by the natural speech syllables. This difference, evident in the shift in mean "ga" responses across speech contexts in the Speech Only conditions (the difference in mean percent "ga" responses for /al/ vs /ar/ contexts was 5.22% for Experiment 1 and 9.89% in Experiment 2) was not consistent enough to be statistically reliable across experiments. Nonetheless, the pattern of effect sizes suggests that the natural speech syllables may have contributed a greater overall influence to target speech categorization. This is simply to say that the speech contexts of Experiment 2 may have contributed more to the resulting target percept relative to the strong influence of the nonspeech contexts than did the synthesized syllables of Experiment 1.

To more closely examine this possibility, an additional statistical analysis was conducted to determine the relative contribution of speech contexts in the hybrid nonspeech/speech conditions across experiments as speech context type (synthesized, natural) varied. A $2\times2\times2\times9$ (Experiment $\times$ Condition $\times$ Context $\times$ Target Speech Stimulus) mixed model ANOVA with Experiment as a between-subjects factor compared the relative influence of speech contexts in the Cooperating versus Conflicting conditions across experiments. A significant difference is reflected by the three-way Experiment $\times$ Condition $\times$ Context interaction, $F(1,18)=56.83$, $p<0.0001$, $\eta_p^2=0.759$. When nonspeech contexts were present, the relative influence of synthesized versus natural speech contexts differed. Computing the difference in mean "ga" responses in the hybrid nonspeech/speech conditions conditioned on the speech context illustrates why this is so. The categorization shift attributable to the synthesized speech contexts of Experiment 1 ($M_{/al/}-M_{/ar/}=58.06-54.50=3.56$) is significantly less than that of the natural speech contexts of Experiment 2 ($M_{/al/}-M_{/ar/}=61.56-53.78=7.78$). Many factors may have contributed to the relatively greater effect of context produced by the natural syllables including, but not limited to, the richer acoustic characteristics of natural speech, the closer spectral correspondence of the natural syllables with the target speech syllables, perception of the two syllables as originating from the same talker, amplitude relationships of the spectral energy from the two precursors, and auditory grouping by common acoustic characteristics. Whatever caused the natural syllables to be relatively stronger contributors to the effect on speech categorization, the results of the Conflicting condition nevertheless provide strong evidence of perceptual contributions from both nonspeech and speech contexts even for natural speech contexts. Moreover, the statistical analyses of the Experiment 2 Cooperating versus Conflicting conditions provide corroborating evidence that both the speech and nonspeech contexts contributed to the observed pattern of results.

IV. GENERAL DISCUSSION

A spectral contrast account of context-dependent speech perception makes strong directional predictions about context-dependent speech categorization in circumstances in which both speech and nonspeech contexts are present. Specifically, it is expected that the effect of joint speech/nonspeech context on speech categorization will be dictated by the spectral characteristics of each source of context such that the speech and nonspeech contexts may either cooperate or conflict in their direction of influence on speech categorization as a function of how they are paired. The results of two experiments demonstrate that speech and nonspeech contexts do jointly influence speech categorization. When hybrid nonspeech/speech context stimuli were spectrally matched in Experiment 1, they collaborated to produce a bigger effect of context on speech categorization than did the same speech contexts on their own. A context effect on speech categorization was also observed in this condition in Experiment 2 (for which natural utterances provided speech context), but this effect was not significantly greater than that observed for the natural speech contexts alone.

When the spectra of the hybrid nonspeech/speech contexts were spectrally mismatched such that they predicted opposing influences on speech categorization, the observed context effects differed from the context effect produced independently by the speech contexts. In Experiment 1, the context effect observed in the Conflicting condition was of equal magnitude, but in the opposite direction of that observed for solitary speech contexts. The direction of the context effect was predicted, not by the adjacent speech contexts, but instead by the spectral characteristics of the temporally nonadjacent nonspeech contexts. A qualitatively similar, although less dramatic, effect was observed for the spectrally conflicting speech and nonspeech contexts of Experiment 2; the nonspeech contexts neutralized the effect of speech context such that no context-dependent shift in target speech categorization was observed. Overall, the effects observed for the hybrid context conditions of Experiment 2, with natural speech contexts matched in source to the target

syllables, were relatively more modest than those observed in Experiment 1. This may have been due to the somewhat larger effect of context exerted by the natural speech contexts. Most important to the aims of the study, however, both experiments provided evidence that linguistic and nonlinguistic sounds jointly contribute to observed context effects on speech categorization. The sum of the results is consistent with general auditory/cognitive approaches with an emphasis on the shared characteristics of the acoustic signal and the general processing of these elements, in this case, spectral distributions of energy. The spectral characteristics of the context stimuli, whether the stimuli were speech or nonspeech, predicted the effects upon the speech categorization targets.

Mechanistically, an important issue that remains is whether such general auditory representations common to speech and nonspeech govern the joint effects of speech and nonspeech contexts on speech categorization or whether independent representations of the context stimuli exert an influence on speech categorization at a later decision stage. This is a thorny issue to resolve in any domain. Some theorists suggest that if processes share common resources or hardware, they can be expected to interfere or otherwise interact with one another whereas if they are distinct, they should not. The present results meet this criterion for indication of common resources or hardware, but further investigation will be required to hold this question to a strict test. Nevertheless, whether nonspeech contexts are operative on common representations or integrated at a decision stage, the information that is brought to bear on speech categorization is clearly not dependent on the signal carrying information about articulation *per se*. An account cognizant of the spectral distributions of acoustic energy possessed by the context stimuli, as postulated by a general auditory/cognitive account under the term *spectral contrast* makes the only clear predictions of what happens to speech categorization when speech and nonspeech are jointly present in the preceding input and these predictions are supported by the results.

With respect to spectral contrast, there is an element of these experiments that may seem puzzling. Considering that previous research has demonstrated that *adjacent* nonspeech context influences speech categorization (e.g., Lotto and Kluender, 1998), one may wonder why the nonspeech contexts of the present experiments exerted their influence on the nonadjacent speech *targets* rather than the adjacent speech *contexts*. To understand why this should be so, it is useful to think about speech categorization as drawing from multiple sources of information.[2] Context is merely one source of information; the acoustic signal corresponding to the target of perception is another. If the acoustic signal greatly favors one speech category alternative over another then context exerts very little effect. This is the case, for example, for the more limited effects of context that emerge (here, and in other experiments) at the end points of the target speech categorization stimulus series where acoustic information is unambiguous with respect to category membership. However, when acoustic signals are partially consistent with multiple speech categories context has a role in categorization. In the present experiments, the speech target syllables were acoustically manipulated to create a series varying perceptually from /ga/ to /da/. Thus, by their very design the intermediate stimuli along this series were acoustically ambiguous and partially consistent with both /ga/ and /da/. Context was thus afforded an opportunity to exert an influence. On the contrary, the acoustic structure of the speech context stimuli in the present experiments overwhelmingly favored either /al/ or /ar/; they were perceptually unambiguous and context therefore could exert little influence. The results of the present experiments demonstrate that when the speech contexts are acoustically unambiguous, they contribute to the effects of context rather than reflect the influence of the nonspeech precursors. Although it may seem surprising that the nonspeech context stimuli should influence perception of nonadjacent speech targets, recent research has demonstrated that the auditory -system is willing to accept context information as evidence by which to shift a categorization decision even when it occurs more than a second prior and even when multiple acoustic signals intervene (Holt, 2005). By these standards, the influence of the nonadjacency of the nonspeech contexts with the speech targets in the present experiments is relatively modest.

In sum, the joint influence of speech and nonspeech acoustic contexts on speech categorization can most simply be accounted for by postulating common general perceptual origins. Previous research has highlighted parallels between phonetic context effects and those observed between purely non-speech sounds (e.g., Diehl and Walsh, 1989), but these results have been challenged on the grounds that perception of nonspeech analogs to speech cannot be directly compared to speech perception, since speech has a clear, identifiable environmental source whereas nonspeech analogs to speech (pure tones, for example) do not (Fowler, 1990). A response to this challenge is that nonspeech contexts influence perception of *speech* (e.g., Lotto and Kluender, 1998). This is a stronger test in that it identifies the information sufficient for influencing speech categorization; when nonspeech stimuli model limited acoustic characteristics of the speech stimuli that produce context effects on speech targets these nonspeech sounds likewise elicit contexts effects on speech categorization. The present experiments introduce a new paradigm to test the joint effects of speech and nonspeech context stimuli on speech categorization. This paradigm is perhaps even stronger in that it allows investigation of the influence of nonspeech signals on speech categorization in the presence of speech context signals that also exert context effects. The present results demonstrate the utility of this tool in pursuing the theoretical question of how best to account for the basic representation and processing of speech.

ACKNOWLEDGMENTS

The author thanks Dr. A. J. Lotto, Dr. R. L. Diehl, Dr. C. Fowler, and Dr. T. Wade for helpful discussions of these results and C. Adams for her essential role in conducting the research. The work was supported by grants from the James S. McDonnell Foundation (Bridging Mind Brain and Behavior, 21st Century Scientist Award) and the National Institutes of Health (2 RO1 DC004674-04A2).

[1]Although dichotic presentation of single tones and speech targets has been shown to produce context effects (Lotto *et al.*, 2003), investigation of the influence of multiple-tone acoustic history contexts on speech categorization under dichotic presentation conditions has not been reported to date. However, the long time course (>1 s) over which effects of tonal acoustic histories on speech categorization persist and the observation that tonal acoustic histories influence speech categorization even when as many as 13 neutral tones intervene between the acoustic history and speech target argue that central (i.e., not purely sensory) auditory mechanisms play an important role (Holt, 2005).

[2]This analysis is consistent with the work of a rational Bayesian decision maker whereby the optimal policy is to combine information from different sources to assign posterior probabilities to possible interpretations of the input and choose the alternative with the highest posterior probability. This approach is amenable to speech perception in that stochastic versions of the TRACE model of speech perception (McClelland, 1991) implement optimal Bayesian inference (Movellan and McClelland, 2001). Moreover, recent theoretical discussions have highlighted how Bayesian analysis may be fruitfully applied to issues in speech perception (Geisler and Diehl, 2002, 2003).

Delgutte, B. (**1996**). "Auditory neural processing of speech," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell, Oxford), pp. 505–538.

Diehl, R. L., and Walsh, M. A. (**1989**). "An auditory basis for the stimulus-length effect in the perception of stops and glides," J. Acoust. Soc. Am. **85**, 2154–2164.

Finney, D. J. (**1971**). *Probit Analysis* (Cambridge University Press, Cambridge, MA).

Fowler, C. A. (**1986**). "An event approach to the study of speech perception from a direct-realist perspective," J. Phonetics **14**, 3–28.

Fowler, C. A. (**1990**). "Sound-producing sources as objects of perception: Rate normalization and nonspeech perception," Percept. Psychophys. **88**, 1236–1249.

Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). "Young infants' perception of liquid coarticulatory influences on following stop consonants," Percept. Psychophys. **48**, 559–570.

Fowler, C. A., Brown, J. M., and Mann, V. A. (**2000**). "Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans," J. Exp. Psychol. Hum. Percept. Perform. **26**, 877–888.

Fowler, C. A., and Smith, M. R. (**1986**). "Speech perception as 'vector analysis:' An approach to the problems of invariance and segmentation," in *Invariance and Variability in Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 123–139.

Geisler, W. S., and Diehl, R. L. (**2002**). "Bayesian natural selection and the evolution of perceptual systems," Philos. Trans. R. Soc. London, Ser. B **357**, 419–448.

Geisler, W. S., and Diehl, R. L. (**2003**). "A Bayesian approach to the evolution of perceptual and cognitive systems," Cogn. Sci. **118**, 1–24.

Holt, L. L. (**1999**). "Auditory constraints on speech perception: An examination of spectral contrast," Diss. Abstr. Int., B **61**, 556.

Holt, L. L. (**2005**). "Temporally non-adjacent non-linguistic sounds affect speech categorization," Psychol. Sci. **16**, 305–312.

Holt, L. L., and Kluender, K. R. (**2000**). "General auditory processes contribute to perceptual accommodation of coarticulation," Phonetica **57**, 170–180.

Holt, L. L., and Lotto, A. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**, 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–990.

Kluender, K. R., Coady, J. A., and Kiefte, M. (**2003**). "Sensitivity to change in perception of speech," Speech Commun. **41**, 59–69.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Lindblom, B. E. F., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.

Lotto, A. J. (**2004**). "Perceptual compensation for coarticulation as a general auditory process," in Proceedings of the 2003 Texas Linguistics Society Conference, edited by A. Agwuele, W. Warren, and S-H. Park Cascadilla Proceedings Project, Sommerville, MA, pp. 42–53.

Lotto, A. J., and Kluender, K. R. (**1998**). "General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification," Percept. Psychophys. **60**, 602–619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)," J. Acoust. Soc. Am. **102**, 1134–1140.

Lotto, A. J., Sullivan, S., and Holt, L. L. (**2003**). "Central locus for non-speech context effects on phonetic identification," J. Acoust. Soc. Am. **113**, 53–56.

Mann, V. A. (**1986**). "Distinguishing universal and language dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r"," Cognition **24**, 169–196.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

McClelland, J. L. (**1991**). "Stochastic interactive processes and the effect of context on perception," Cogn. Psychol. **23**, 1–44.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.

Movellan, J. R., and McClelland, J. L. (**2001**). "The Morton-Massaro law of information integration: Implications for models of perception," Psychol. Rev. **108**, 113–148;

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

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Smith, R. L. (**1979**). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers," J. Acoust. Soc. Am. **65**, 166–178.

Stephens, J. D. W., and Holt, L. L. (**2003**). "Preceding phonetic context affects perception of non-speech sounds," J. Acoust. Soc. Am. **114**, 3036–3039.

Sutter, M. L., Schreiner, C. E., McLean, M., O'Connor, K. N., and Loftus, W. C. (**1999**). "Organization of inhibitory frequency receptive fields in cat primary auditory cortex," J. Neurophysiol. **82**, 2358–2371.

Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.

Ulanovsky, N., Las, L., and Nelken, I. (**2003**). "Processing of low-probability sounds by cortical neurons," Nat. Neurosci. **6**, 391–398.

Wade, T., and Holt, L. L. (**2005**). "Effects of later-occurring non-linguistic sounds on speech categorization," J. Acoust. Soc. Am. **118**, 1701–1710.

Watkins, A. J., and Makin, S. J. (**1994**). "Perceptual compensation for speaker differences and for spectral-envelope distortion," J. Acoust. Soc. Am. **96**, 1263–1282.

Watkins, A. J., and Makin, S. J. (**1996a**). "Some effects of filtered contexts on the perception of vowels and fricatives," J. Acoust. Soc. Am. **99**, 588–594.

Watkins, A. J., and Makin, S. J. (**1996b**). "Effects of spectral contrast on perceptual compensation for spectral-envelope distortion," J. Acoust. Soc. Am. **99**, 3749–3757.

Processing unattended speech

Marie Rivenez[a)]
Laboratoire de Psychologie Expérimentale, CNRS, and Université René Descartes, Institut de Médecine Aérospatiale du Service de Santé des Armées, France

Christopher J. Darwin
Department of Psychology, University of Sussex, Brighton, United Kingdom

Anne Guillaume
Institut de Médecine Aérospatiale du Service de Santé des Armées, France

(Received 29 June 2005; revised 23 February 2006; accepted 3 March 2006)

Three experiments examine the effect of a difference in fundamental frequency (F0) range between two simultaneous voices on the processing of unattended speech. Previous experiments have only found evidence for the processing of nominally unattended speech when it has consisted of isolated words which could have attracted the listener's attention. A paradigm recently used by Dupoux *et al.* [J. Exp. Psychol.: Human Percept. Perform. **29**(1), 172–184 (2003)] was modified so that participants had to detect a target word belonging to a specific category presented in a rapid list of words in the attended ear. In the unattended ear, concatenated sentences were presented, some containing a repetition prime presented just before a target word. Primes speeded category detection by 25 ms when the two messages were in a difference F0 range. This priming effect was unaffected by whether the target was led to the left or the right ear, but disappeared when there was no F0 range difference between the messages. Finally, it was replicated when participants were compelled to focus on the attended message in order to perform a second task. The results demonstrate that repetition priming can be produced by words in unattended continuous speech provided that there is a difference in F0 range between the voices.
[DOI: 10.1121/1.2190162]

PACS number(s): 43.71.Bp, 43.66.Rq, 43.71.Gv, 43.66.Lj [ARB] Pages: 4027–4040

I. INTRODUCTION

The purpose of this research was to investigate the processing of unattended speech. Most studies agree that simple attributes of unattended speech are readily perceived (Cherry, 1953; Broadbent, 1958, 1971; Deutsch and Deutsch, 1963; Treisman, 1960, 1964a, c, d). For example, Cherry (1953) found that when listeners focus their attention on a message in one ear, they are able to notice simple physical changes in the unattended message in the other ear, but they fail to notice its semantic content. Furthermore, many authors have found that selection of the attended message is facilitated by a simple physical difference between the attended and unattended messages (Broadbent, 1971; Brungart and Simpson, 2002; Brungart *et al.*, 2001; Cherry, 1953; Egan *et al.*, 1954; Poulton, 1953, 1956; Spieth *et al.*, 1954; Treisman, 1964b), implying that some physical cues are processed in the unattended message.

However, there is less unanimity over whether unattended speech can be processed at the lexical level. Many studies are consistent with the idea that speech cannot be lexically processed without attention and a similar number have claimed that it can [see Holender (1986) for a review; Dupoux *et al.*, 2003]. The usual paradigm to assess the extent of lexical processing consists in presenting two messages at the same time, one in each ear. Participants are asked to focus on one message by doing a specific task (e.g., shadowing) and to ignore the other message. Lexical processing of the unattended message is assessed by its influence on performance of the primary task. Discrepancies in the results of different studies have been attributed to the material used in the unattended message: studies showing a lexical processing of unattended messages have used isolated words in the unattended ear, which might have encouraged attentional switches to the unattended message, whereas studies showing no lexical processing of the unattended message have used continuous speech (Holender, 1986).

All the studies that have shown lexical processing of the unattended message have used isolated words in the unattended ear. For example, Eich (1984) presented pairs of words in the unattended message and asked participants to perform a shadowing task on the attended message. The pairs of words were composed of a homophone and a word related to the homophone's less frequent meaning. After the shadowing, participants were asked to spell homophones, some of which had been previously presented in the unattended message, and these spellings were found to be biased towards the less frequent meaning. The unattended words would need to have been processed to a lexical level in order to give this result. MacKay's (1973) data are consistent with Eich's (1984). He asked participants to shadow ambiguous sentences while at the same time ignoring a list of isolated words, which could disambiguate the attended sentence. The sentence's interpretation was influenced by the unattended

[a)]Electronic mail: mrivenez@imassa.fr

words (see also Lackner and Garrett, 1972). In a similar vein, Lewis (1970) found that shadowing latency increased when a list of semantically related words was presented in the unattended message. Another persistent finding is that a listener's own name can be easily noticed and recorded when it is presented in a list of unattended words (Moray, 1959, 1969; Oswalt *et al.*, 1960; Wood and Cowan, 1995a, b). In all these experiments, participants' attention could have been attracted by the unattended isolated words.

In contrast to these experiments in which the unattended words were isolated, experiments that have used continuous speech or rapid presentation of isolated words in the unattended ear have failed to find evidence of it being lexically processed unless attention has been allowed to switch to the unattended message. For example, Newstead and Dennis (1979) showed that the MacKay effect disappeared when the unattended words, which had originally been presented in isolation, were embedded in sentences. Attentional switches were also implicated in a study by Wood *et al.* (1997). They found that the homophone effect found by Eich (1984) disappeared when the attended words were presented faster, providing less opportunity for listeners to switch attention. Wood and Cowan (1995a) also found that attentional switches may explain the detection of backward speech in an unattended message. Their listeners heard backward speech as the unattended continuous message and simultaneously had to perform a shadowing task on the attended ear. When listeners noticed the backward speech in the unattended message, the number of shadowing errors systematically increased.

These results all bear out the conclusion by Holender (1986) that isolated words can automatically capture attention more easily than words which are embedded in sentences. The abrupt onsets in the unattended ear, created by the isolated words surrounded by silence, may attract attention because they are one of the most relevant cues to indicate the appearance of a sound. However, a major limitation of the studies that had used continuous speech or rapid isolated words is that the lexical activation of the unattended words had never been assessed directly: unattended message processing was always assessed with off-line measures (homophone spelling, sentence interpretation, recall of what was noticed in the unattended message). Consequently, the lack of evidence for lexical processing of unattended words in experiments which had controlled switches of attention might have been a result of the particular measures of lexical activation that had been used (interference with shadowing, backward speech report). To our knowledge, only one recent study, carried out by Dupoux *et al.* (2003), has used a direct measure of lexical activation, and it reinforced Holender's distinction between isolated words and continuous speech.

Dupoux *et al.* (2003) looked for priming by unattended words (presented as lists of isolated words to one ear) of lexical decision reaction times made to lists of isolated words on the attended ear. In order to reduce the probability of attentional switches, the unattended speech was time compressed, attenuated by 12 dB relative to the attended speech, and the prime's onset was synchronized to its target. Their first experiment showed a priming effect—reaction times were faster when the prime was a repetition of the target than when it was not. However, most of the participants were conscious of the presence of the prime, indicating that they had switched their attention to the unattended prime. This effect is consistent with studies showing that isolated words attract more attention than continuous speech (Holender, 1986; Newstead and Dennis, 1979). In order to avoid these attentional switches, Dupoux *et al.* (2003) carried out a second experiment in which the prime was inserted in a carrier sentence. The carrier sentence started before the target, but target and prime were still synchronized. In this experiment no priming effect was observed, suggesting that when attention is properly controlled, lexical processing of unattended words is not possible. Finally, in order to control for possible degradation of the prime by the carrier, they replicated the second experiment but presented the prime and the target to the same ear (experiment 3), in order to allow an attentional and conscious processing of the prime. The prime and the carrier were played first, followed by the target after a 150-ms (millisecond) pause. In this experiment a priming effect was observed. The authors concluded that the carrier did not degrade the prime and thus that unattended words cannot be lexically processed if attention is well controlled.

Since in these experiments the prime was phonologically identical to the target, priming could be phonological rather than lexical. The authors argue for the priming being lexical on the grounds that it did not occur for nonword targets and primes.

However, the Dupoux *et al.* (2003) paradigm is open to criticism on three points: that the prime was degraded by temporal compression; that the prime was presented to the left ear; and that the prime and target were presented in the same voice.

First, the prime was time compressed. Although this compression did not disturb lexical processing for the isolated-word primes used in their experiment 1, it may have been at least partly responsible for the lack of priming found in their experiment 2. In their control experiment (experiment 3), they tried to demonstrate that the prime was not degraded by the carrier sentence by allowing participants to focus their attention on it and by showing a priming effect in such condition. Nevertheless in experiment 3 (where priming was found), the prime was presented alone whereas in experiment 2 (where no priming was found) the prime was presented simultaneously with the target. It is thus impossible to decide whether the lexical processing found in their control experiment was due to attention or to facilitation in the perceptual processing of the prime. Thus, we think that the lack of priming effect in the Dupoux *et al.* (2003) experiments could be due to reduced lexical processing of the prime when it is temporally compressed and embedded in a carrier sentence.

The second criticism is related to the ear to which the prime was presented. The prime was always presented to the left ear, which is preferentially connected to the right hemisphere which is not specialized for lexical processing (Kimura, 1967; Zatorre, 1989). Thus the lack of priming could be due to the poorer linguistic ability of the right than the left hemisphere.

The third criticism of this paradigm concerns the fact that the authors used the same voice for the prime and the target. When two speech messages are presented simultaneously, their intelligibility is improved when they differ in various physical attributes such as ear of presentation, pitch, and other voice characteristics such as vocal-tract length. Part of this improvement is due to better simultaneous grouping of the mixture of speech cues (Bregman, 1990; Darwin and Carlyon, 1995) and part to better tracking of a sound source over time (Darwin *et al.*, 2003). When, as in the Dupoux *et al.* (2003) paradigm, the prime is in the same pitch range as the target and had other similar vocal characteristics, the perceptual separation of the target and prime may be reduced. Such reduction may be particularly detrimental to the perception of the prime since it is being presented at a level of −12 dB relative to the target. Using the same voice for the prime and the target may thus reduce perceptual clarity[1] and hence lexical activation of the prime.

In summary, although Dupoux *et al.* (2003) used an innovative and interesting paradigm to study unattended word processing, the question of whether unattended words can be lexically processed remains open and requires control of the factors discussed above. The purpose of the present paper is to ask whether evidence for the lexical processing of unattended words can be found (a) using continuous speech in the unattended message instead of isolated words, (b) using repetition priming as a direct measurement of lexical activation, and (c) controlling for the three factors that may have been responsible for the lack of evidence for lexical priming in the second experiment of Dupoux *et al.* (2003).

The task that we used was based on that used by Dupoux *et al.* (2003), but was modified in order to make it more likely that listeners would continuously attend to the appropriate ear. Listeners were instructed to detect words belonging to a specific category in a rapid list of words presented in the attended ear and to ignore sentences played to the other ear. The prime consisted of a lexical repetition of the target embedded in the sentence and presented just before the target. In order to keep the prime perceptually salient, it was not time compressed and was spoken with a normal prosody, but, in the first experiment, the pitch of the unattended speech was increased to put it in a different pitch range from the target.

We report three experiments. The first two experiments were directly related to the three criticisms we have made of the Dupoux *et al.* (2003) paradigm. In experiment 1, we looked for evidence of lexical processing of unattended words, using the task we have just described, and assessed the effect of hemispheric asymmetries on unattended word processing. In experiment 1a, participants were right- or left-handed without any selection, in order to be able to compare our results to Dupoux *et al.* (2003); in experiment 1b, we replaced the left-handers of experiment 1a by right-handers. In experiment 2 we directly investigated the effect of a difference in the pitch range of the attended and unattended messages. Finally, in experiment 3, we introduced a secondary task into the paradigm in order to control attention more directly.

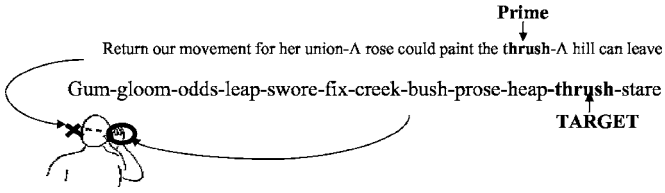


FIG. 1. Schematic of the stimuli used in experiments 1 and 2. The list of monosyllabic words was presented in the attended ear with the target in bold character. The sentences were presented in the unattended messages. They started around 800 ms after the attended list of words. They were composed of four short sentences. One of the sentences contained the prime (bold characters). Prime offset was synchronized to target onset.

The general hypothesis is that if unattended words are lexically processed, then we should get a priming effect: RTs (reaction times) should be faster when the target is preceded by a repetition prime. However, no priming effect should be observed if unattended words are not processed at a lexical level.

II. EXPERIMENT 1: UNATTENDED WORD PROCESSING WITH UNDEGRADED PRIMES: HEMISPHERIC ASYMMETRIES

The purpose of experiment 1 was to assess whether priming of a semantic category decision reaction time could be achieved from an unattended repetition prime, when the prime was embedded in continuous speech. We modified the Dupoux *et al.* (2003) paradigm in the following ways: (a) the decision was a semantic one rather than word/nonword; (b) the unattended message consisted of continuously played nonsense sentences with the prime (which was lexically identical to the target) embedded in one of the sentences; (c) the unattended sentences were not degraded by speeding up, but had their pitch raised; (d) the prime was not simultaneous with the target, but was presented just before; and (e) the prime and its surrounding sentences were presented either to the right ear or to the left and the target to the opposite ear. Lexical processing of unattended words should lead to faster reaction times to targets that are accompanied by a repetition prime than to those that occur with an unrelated word.

In experiment 1a, as in the Dupoux *et al.* (2003) study, participants were not selected for handedness whereas experiment 1b only used right-handers. If an ear difference is responsible for the failure of Dupoux *et al.* (2003) to find lexical priming from continuous speech, then we should obtain priming only for primes presented to the right ear.

A. Experiment 1a: Right- and left-handed participants: Method

1. Material

Each attended message consisted of a list of on average 18 (±1) words presented at an average rate of 2.1 (±0.1) words/s (per second) and lasting about 8.8 (±0.5) s (see Fig. 1). Monosyllabic words were used in order to avoid them being recognized substantially before their physical end, and so allowing possible attentional switches to the unattended message. For the same purpose, there were no gaps between the words of the list and we chose to use low-frequency words (average Kucera-Francis frequency=25 per million,

Kucera and Francis, 1967). Their average duration was 484 ms and they were recorded by a male native talker of British English (CJD) speaking with a monotonous intonation at an average fundamental frequency (F0) of about 140 Hz. The F0 was subsequently precisely flattened and increased to 180 Hz using the Praat (4.1) implementation of the PSOLA algorithm. Targets were also low-frequency monosyllabic words (average Kucera-Francis=20 per million) in order to make the task difficult enough to limit attentional switches toward the unattended message. The average target duration was 435 (±74) ms and their average typicality for their category was 1.635 (values ranged between 1.109 and 2.392) on a scale from 1 (very typical) to 6 (very untypical) (Hampton and Gardiner, 1983). We used 11 semantic categories (see Appendix A). Targets were initially read and subsequently processed as part of the list of attended words. There were either one or two targets in each trial. Targets were never presented within the first 3 s of a trial, but they could occur at any subsequent time in the trial.

Unattended messages were composed of nonsense sentences, which were used in order to minimize the information about the prime that listeners could hear by switching their attention to the unattended ear before the occurrence of the prime. Each unattended message, consisting of four nonsense sentences, was spoken by the same talker but with a normal reading prosody and was not digitally altered. The four sentences were concatenated with a silent interval that varied between approximatively 10 and 100 ms in order to exactly synchronize each prime offset to the onset of its relative target when the trial contained two target words. For instance, a series of four sentences could be "her fee may import the coach—his suit could land a bridge—tell her pain on the soil—the gun held the pound." The sentences were adapted from those used by Freyman *et al.* (1999) (see Appendix B). The text of the sentences was slightly modified in order to insert the prime and unprime words: we replaced a content word by a prime or an unprime word. Primes were either the same word as the target (primed condition) or an unrelated word (unprimed condition) (see Appendix B). Prime and unprime words were embedded in the unattended sentences and so were naturally coarticulated in the sentence context. Prime sentences were created by recording the sentences with their prime words. Nonprime sentences were obtained by replacing the prime words by their unprimed words in the signal, taking care that the coarticulatory cues were still coherent.

Unattended messages started on average 800 ms after the onset of attended messages in order to help participants to focus on the attended message and to ignore the unattended one. The pitch of the unattended sentences was not altered. Thus because of the monotonous pronunciation and resynthesis of the targets, prime and target were acoustically very dissimilar. Prime offset was synchronized to target onset in order to give enough time for lexical activation to occur. The prime could be at any syntactically permissible position in the sentence (second, third, or last word of the sentence).

For each participant, half of the unattended sentences were presented to the right ear (targets were in the left ear) and half were presented to the left ear (targets were in the right ear). We created four lists of items: one list was used to counterbalance the primed and unprimed items (list A and list B) and the other was used to counterbalance right ear primes and left ear primes (list 1 and list 2). The four lists were created by a combination of lists A/B and lists 1/2 (A1, A2, B1, and B2). For instance, in list A1, half of the items were primed and half unprimed and, for each priming condition, half of the primes were presented to the right ear and half to the left ear. List A2 was as list A1 but with the ears reversed. Primed items in the A lists became unprimed in the B lists and vice versa.

Sounds were digitized in 16 bits quantization at 22.05-kHz sampling rate and were presented through an Audiomedia III Soundcard in a Macintosh G3 running PsyScope under MacOS 9.2 and played binaurally through Sennheiser HD 414 headphones at an average level of 72 dB SPL (sound pressure level) for the attended ear and 60 dB SPL for the unattended ear, as measured with an artificial ear. Response times were recorded by a PsyScope button box.

2. Procedure

Participants were told to listen to the speech in one ear (left or right), ignoring the other ear (right or left), pressing a button whenever they detected a noun belonging to a specific category. They were told the categories they would be listening for and examples were given ("for instance, if you have the category Bird, then you will have to press the button if you hear Blackbird"). At the beginning of each trial, the target category for that trial appeared in the middle of the screen. After 500 ms, the trial started with the category name remaining on the screen until the end of the trial. The next trial started automatically 500 ms after the end of the preceding one. Participants first took a practice session of 12 targets (nine trials) lasting about 2 min. The test session followed the practice session and was composed of 60 targets (39 trials) lasting about 7 min. The presentation orders of primed and unprimed trials and of the different categories were randomized across trials and participants. Ear was blocked: unattended sentences were all presented in one ear for half the experiment, then, after a short break, in the other ear. Which ear was used first was counterbalanced across the listeners.

After the experiment, participants were asked if they had heard anything special in the unattended ear (which had received the unattended sentences) and were also asked specifically if they had noticed the repetition prime.

3. Participants

Forty-eight students who were native English speakers at the University of Sussex with no reported hearing, language, or attentional impairment were paid for their participation. Thirty-nine were right handed and nine were left handed. They all showed a normal audiogram for each ear with less than 10-dB difference between the two ears.[2]

B. Results

For each participant's RTs, we first discarded reaction times that were above 1500 ms or below 300 ms,[3] and then discarded reaction times above or below two standard devia-

TABLE I. Results of experiment 1a (right- and left-handed participants). RTs (in ms) and percentage of correct detections for the primed and unprimed conditions (rows) and for the ear in which the unattended sentence was presented (columns). Standard deviation (SD) is shown in parentheses: the first number refers to the SD across participants and the second refers to the SD across items.

	RTs (in ms)			% Correct detections	
	Right	Left	Average	Right	Left
Primed	756(68;58)	746(73;62)	751(70;60)	73	74
Unprimed	773(62;55)	771(58;59)	772(60;56)	75	72
Priming	17(56;52)	25(80;51)	21(51;39)		

tions from the new average, replacing the missing values by the average value for this condition. Targets detected by less than 40% of the participants (averaged across each condition) were discarded from the analysis. On average, an individual target was detected by 76% of the participants and four targets were discarded. A five-factor ANOVA for the participants analysis was computed with a priming factor contrasting primed and unprimed trials (within factor), an ear factor contrasting the ear in which unattended sentences were presented (within factor), a priming counterbalancing factor contrasting list A and list B (between factor), an ear counterbalancing factor contrasting list 1 and list 2 (between factor), and an ear order counterbalancing factor contrasting the order of ear presentation (between factor). We did a three-factor ANOVA for the item analysis with a priming factor (within factor), an ear factor (within factor), and an ear order counterbalancing factor (within factor).

Table I shows the RTs expressed in ms for the primed and unprimed condition (rows) and for the ear of prime presentation (columns). For both ears, RTs were faster for the primed condition (756 ms) than for the unprimed condition (773 ms). The magnitude of the priming effect was 21 ms and it was significant for both the participant and the item analysis $[F_1(1,40)=13.13;\ p<0.001;\ F_2(1,55)=18.22;\ p<0.0001]$. There was no main effect of ear: mean RTs were 765 ms when the prime was presented to the right ear and 759 ms when it was presented to the left ear $[F_1<1;\ F_2(1,55)=1.07;\ p=0.16]$. There was no significant interaction between priming and ear ($F_1<1;\ F_2<1$). The magnitude of the priming effect was 17 ms when unattended sentences were presented to the right ear and 25 ms when they were presented to the left ear. Thus the priming effect did not vary with the ear to which the unattended sentences were presented.

The percentages of correct detections were analyzed after having cleaned the RTs and removed targets detected by less then 40% of the participants. Although these percentages were quite low, they did not differ across the priming and the ear factor (73% with the prime in the right; 74% with the prime in the left ear; 75% with no prime and unattended sentences in the right ear; 72% with no prime and unattended sentences in the left ear) ($F<1$). In addition, a few of the participants (15%) occasionally noticed the prime (as a word from the target category in the unattended ear), but none of them noticed that there were repetitions of the same word within a trial in the unattended and attended ears. Removing these participants did not materially change the results. The overall priming effect was 22 ms $[F_1(1,33)=11.01;\ p<0.005;\ F_2(1,55)=13.44;\ p<0.001]$, which did not vary with ear $[F_1(1,33)=1.01;\ p=0.32;\ F_2(1,55)=1.71;\ p=0.20]$: it was 14 ms when the prime was presented to the right ear and 31 ms when the prime was presented to the left ear.

C. Discussion

This experiment has shown a significant repetition priming effect on RTs to a semantic category judgment by words in continuous speech. RTs were faster when the target word in the attended ear was preceded by a repetition prime than when it was preceded by an unrelated word. This result is opposed to those of Dupoux *et al.* (2003), who found that repetition priming only occurred when the primes were isolated words, not when they were embedded in continuous speech. Our result also goes against the conclusion (Dupoux *et al.*, 2003; Holender, 1986; Newstead and Dennis, 1979; Wood *et al.*, 1997) that lexical processing of nominally unattended words only occurs if there has been a switch of attention to the unattended message.

Dupoux *et al.* (2003) found a priming effect of 69 ms when participants consciously processed the prime, whereas the priming effect we found was substantially smaller (21 ms) and participants were not conscious of hearing the prime. This contradictory result may show that consciousness enhances the lexical activation of the prime, which increases the magnitude of the priming effect. When participants were not aware of hearing the prime, Dupoux *et al.* (2003) found no priming effect, while we did find one. We believe that the time-compression of the prime they used may have reduced the preactivation of the lexical representation of the target.

It is unlikely that the priming that we have found arises as a result of listeners switching their attention to the prime. When participants reported hearing the prime, they failed to notice its immediate repetition as a target. The task is thus apparently hard enough to prevent the prime being attended during a correct detection of the target. The task difficulty is also attested to by the quite low correct detection rate (73%). Since the percentage of correct detections was the same in the primed and the unprimed conditions (73%), the difference in RT between these two conditions is not due to a trade-off between reaction times and accuracy. Prime and unprimed conditions are fully comparable in terms of number and type of observations.

We found no significant interaction between the magnitude of priming and the ear to which the unattended sentences were presented. Priming was not different for primes presented to the right ear (17 ms) and to the left ear (25 ms). It is thus unlikely that the ear that the unattended message was assigned to is the reason for the difference in priming between our experiment 1 and those of Dupoux *et al.* (2003).

Failure to find any difference between the ears in the extent of priming in this experiment may be due to the fact that the participants were not selected for right-handedness. A significant effect from the right-handers might have been

TABLE II. Results of experiment 1b (right-handed participants). RTs (in ms) and percentage of correct detections for the primed and unprimed conditions (rows) and for the ear in which the unattended sentence was presented (columns). Standard deviation (SD) is shown in parentheses: the first number refers to the SD across participants and the second refers to the SD across items.

	RTs (in ms)			% Correct detections	
	Right	Left	Average	Right	Left
Primed	753(66;59)	750(68;58)	752(67;58)	74	76
Unprimed	772(71;56)	777(61;61)	774(66;58)	76	72
Priming	19(63;50)	27(73;39)	22(48;36)		

diluted by a null or opposite effect from the left-handers. This hypothesis was examined in experiment 1b.

D. Experiment 1b: Right-handed participants

The purpose of experiment 1b was to test whether a population of right-handers shows greater priming from primes presented to the right than to the left ear. We replaced the nine left-handers of experiment 1a by nine new participants who were right-handed. They took exactly the same experimental conditions as the left-handers that they replaced.

E. Results

On average, targets were detected by 71% of the participants. The six targets detected by less than 40% were discarded. The results are shown in Table II; the same ANOVAs on participants and items as in experiment 1a were conducted. The results were very similar to those of experiment 1a. Mean RTs were faster for the primed condition (752 ms) than for the unprimed condition (774 ms). This priming effect was 22 ms and it was significant [$F_1(1,40)=14.78$; $p<0.0005$; $F_2(1,53)=21.54$; $p<0.00005$]. There was no main effect of ear ($F_1<1$; $F_2<1$): mean RTs were 763 ms when unattended sentences were presented either to the right or the left ear. The priming effect was 19 ms when unattended sentences were presented to the right ear and 26 ms when they were presented to the left ear. There was no significant interaction between the priming and the ear factor ($F_1<1$; $F_2<1$).

F. Discussion

The purpose of experiment 1b was to test whether right-handed participants showed a difference in priming when the prime was channelled to either the left or the right ear. Although we replicated the priming effect found in experiment 1a with a purely right-handed population of participants, we found that the ear in which the prime was presented did not affect the magnitude of the priming effect. The priming effect is thus not sensitive to hemispheric asymmetries with our paradigm.

The lack of an ear difference in priming leaves open the question of the discrepancy between the results of experiment 1 and those of Dupoux *et al.* (2003). The main difference between our paradigm and that used by Dupoux *et al.* (2003) is that they degraded the prime by time compressing it, a manipulation that may have been responsible for their lack of priming. Another difference between the two experimental paradigms is that Dupoux *et al.* (2003) used the same voice for both targets and primes (though they temporally compressed the prime), whereas we used the same voice but manipulated the pitch of the voice. The perceptual clarity of the prime may have been greater in our experiments because the primes were not temporally compressed and also were in a different pitch range than the targets. Experiment 2 examines whether a different pitch range between target and prime influences priming.

TABLE III. Different pitch range conditions used in experiment 2. Each cell shows the average F0 of the voice for the attended and unattended messages (in row).

	Different F0	Different F0	Same F0	Same F0
Attended message	180 Hz	140 Hz	140 Hz	180 Hz
Unattended message	140 Hz	180 Hz	140 Hz	180 Hz

III. EXPERIMENT 2: EFFECT OF PITCH RANGE DIFFERENCE

The purpose of experiment 2 was to test the effect of a pitch range difference between attended and unattended messages on priming. In experiments 1a and b the F0 of the attended list of words containing the target items was fixed (180 Hz) at a substantially higher F0 than that of the unattended, normally intonated sentences that contained the prime (average F0=140 Hz). In experiment 2 we manipulated the monotone pitch of the attended list of words and of the unattended sentences, using an experimental design which made symmetric changes to the pitch of both the attended and unattended messages. This pitch manipulation factor was a between-participants factor.

A. Method

1. Material and procedure

The material was the same as in experiment 1, except that we added a new set of stimuli. We manipulated the pitch of the attended list of target words so that their monotonous pitch was either in the same pitch range as the unattended sentences (140 Hz) or in a different pitch range (180 Hz). We also manipulated the overall pitch of the unattended sentences, increasing it by 40 Hz (using Praat to maintain the original pitch contour). Thus both the attended and unattended message could have an average F0 of 140 Hz or 180 Hz depending on the pitch range condition. Table III summarizes the four different experimental conditions.

Each participant was assigned to one of the four experimental conditions. We created two lists of items in order to counterbalance the priming factor (list A/B); otherwise the procedure was the same as in experiment 1.

TABLE IV. Results of experiment 2. RTs (in ms) and percentage of correct detections for the primed and unprimed conditions (rows) and for the different and same pitch range conditions (columns) when the unattended message had an average F0 of 140 Hz. Standard deviation (SD) is shown in parentheses: the first number refers to the SD across participants and the second refers to the SD across items.

	RTs (in ms)			% Correct detections	
	Different F0	Same F0	Average	Different F0	Same F0
Primed	745(62;55)	762(61;56)	754(62;56)	72	69
Unprimed	773(66;57)	769(65;53)	771(65;54)	71	65
Priming	29(50;47)	7(51;60)	18(51;58)		

2. Participants

Eighty students at the University of Sussex were paid for their participation. They were not selected for handedness. They were native English speakers with no reported hearing, language, or attentional impairment.

B. Results

Data were cleaned in the same way as in experiment 1 and the mean values are shown in Tables IV and V. On average, targets were detected by 71% of the participants and three of them, which were detected by less than 40%, were discarded. A four-factor ANOVA for the participant analysis was conducted with a priming factor contrasting primed and unprimed trials (within factor), a pitch range difference factor contrasting the same and the different pitch range condition (between factor), an unattended message pitch range factor contrasting the 140- and the 180-Hz F0 range (between factor), and a priming counterbalancing factor contrasting list A and list B (between factor). We did a three-factor ANOVA for the item analysis with a priming factor (within factor), a pitch range difference factor (within factor), and an unattended message pitch range factor (within factor).

We found a main effect of the priming factor: mean RTs were faster for the primed condition (747 ms) than for the unprimed condition (765 ms). This priming effect (19 ms) was significant $[F_1(1,72)=12.35;\ p<0.0001;\ F_2(1,56)=22.89;\ p<0.000\,01]$. The priming effect varied significantly across the pitch range conditions: planned comparisons indicate that when the pitch range was different between the attended and unattended messages, the priming effect was 28 ms and it was significant $[F_1(1,72)=14.23;\ p<0.001;\ F_2(1,56)=25.10;\ p<0.000\,01]$, whereas, when the pitch range was the same, the priming effect was 9 ms and it was nonsignificant $[F_1(1,72)=1.43;\ p=0.23;\ F_2(1,56)=3.61;\ p=0.06]$. The interaction between the priming factor and the pitch range difference was marginally significant for the participant analysis $[F_1(1,72)=3.32;\ p<0.10]$ and significant for the item analysis $[F_2(1,56)=7.48;\ p<0.01]$. The significant priming in the different-pitch condition held for both F0 ranges of the attended message: 29 ms $[F_1(1,72)=7.23;\ p<0.01;\ F_2(1,56)=21.20;\ p<0.000\,05]$ for an unattended message F0 range of 140 Hz, and 28 ms for 180 Hz $[F_1(1,72)=7.01;\ p<0.01;\ F_2(1,56)=10.27;\ p<0.005]$. The nonsignificant priming in the same-pitch condition held also for both F0 ranges: 7 ms for an unattended message F0 range of 140 Hz ($F_1<1$; $F_2<1$) and 11 ms for 180 Hz $[F_1(1,72)=1.16;\ p=0.28;\ F_2(1,56)=3.07;\ p=0.08]$. There was neither a significant interaction between the unattended message F0 and the priming factor ($F_1<1$; $F_2<1$) nor between the other factors of the analysis.

The percentage of correct detections was homogeneous across all the experimental conditions (71% and 68%, respectively for the primed and unprimed conditions for an unattended message F0 range of 140 Hz; 74% and 71% for 180 Hz). However, the percentage of correct detections was lower (68%) when the pitch range was the same than when it was different (74%). An ANOVA conducted on the four principal factors of the experiment for the percentage of correct detections revealed a significant effect of the pitch range difference $[F(1,72)=6.41;\ p<0.05]$. However, there was neither an interaction with the priming factor nor the F0 ranges of the unattended message.

TABLE V. Results of experiment 2. RTs (in ms) and percentage of correct detections for the primed and unprimed conditions (rows) and for the different and same pitch range conditions (columns) when the unattended message had an average F0 of 180 Hz. Standard deviation (SD) is shown in parentheses: the first number refers to the SD across participants and the second refers to the SD across items.

	RTs (in ms)			% Correct detections	
	Different F0	Same F0	Average	Different F0	Same F0
Primed	729(69;62)	751(54;53)	740(62;58)	78	69
Unprimed	757(72;61)	762(45;47)	760(59;54)	75	67
Priming	28(37;66)	11(52;49)	20(46;57)		

C. Discussion

Experiment 2 has found that the priming effect is influenced by the relative pitches of the attended and the unattended messages. As in experiment 1, when attended and unattended messages were in a different pitch range, there was a significant priming effect of 28 ms on average. But when the monotonous pitch of the targets was in the same pitch range as the normally intonated sentences on the unattended ear, there was no significant priming.

We used a between-participants design that allowed substantially more data to be collected than in experiment 1 and manipulated independently both the relative F0 of the attended and unattended messages, and the absolute F0 range of the unattended message. The fact that priming only occurs when the pitch ranges are different is thus not due to a confounding effect of the absolute pitch values. Although the percentage of correct detections was homogeneous across the priming conditions, they were higher in the different than in the same pitch range condition. This difference is due to the greater difficulty of ignoring the unattended message when it is in the same pitch range as the attended one, with an increase in the probability of attentional switches leading to target omissions.

Our data offer a possible explanation for the failure of

Dupoux *et al.* (2003) to find priming when the primes were embedded in continuous speech. Two factors may have been responsible, both of which may have led to a decrease in the perceptual clarity of the unattended speech. The first factor is that the unattended speech in the experiments of Dupoux *et al.* (2003) was time compressed; the second factor is that the unattended and attended messages were in the same pitch range. In order for priming to occur, the perceptual clarity of the prime must be sufficient to give rise to the activation of a mental representation. Compression and/or interfering speech on a similar pitch may both degrade this perceptual clarity, reducing any priming.

In our final experiment we turn to the question of switches of attention to the unattended ear. In order to reduce attentional switches to a minimum, following Holender (1986) and Dupoux *et al.* (2003), our previous experiments have used continuous speech in the unattended message. However, we have not made any explicit measure of listeners' attention. Asking participants whether they noticed the prime involves a memory coding of the word. However, Cowan *et al.* (1990) and Wood and Cowan (1995a, b) have shown that only limited processing takes place without attention and short-term retention of speech is rather incomplete without attention. They have shown a rapid decay of memory for unattended spoken syllables over 10 s (Cowan *et al.*, 1990). Thus, we introduce additional attentional controls in experiment 3.

IV. EXPERIMENT 3: CONTROL OF ATTENTIONAL SWITCHES

The purpose of experiment 3 was to control attentional switches toward the unattended prime. The rationale was based on the idea that if we compel participants to focus their attention on the attended message at the same time as the prime presentation, then any priming effect that we get is unlikely to be due to attention being switched to the unattended ear. In order to compel participants to listen to the attended message at this particular time, we asked them to remember a specific word of the attended list of words, while they still had to perform the category detection task. In half of the trials, the to-be-remembered word was simultaneous with the prime (just before the target).

A. Method

1. Material

The material was the same as in experiment 1, except that a loud tone (440 Hz presented at a level of 85 dB for 400 ms) was added to the attended message (see Fig. 2) in order to announce the to-be-remembered word (the word "heap" in Fig. 2). The tone signal was mixed with the attended signal. The tone always appeared once in each trial and it ended as the to-be-remembered word started. Since the words lasted 484 ms on average, the tone always overlapped the word preceding the to-be-remembered word (the word "prose" in Fig. 2). In one half of the trials, the to-be-remembered word was just before the target word and in the other half at a neutral position (distant from the target word). When the to-be-remembered word was presented just before the target, it occurred at the same time as the prime (because prime offset was always synchronized to target onset). Thus half of the targets were preceded by a to-be-remembered word (near recall condition: the to-be-remembered word was presented just before the target word) and half were presented alone (distant recall condition: the to-be-remembered word was presented distant from the target).

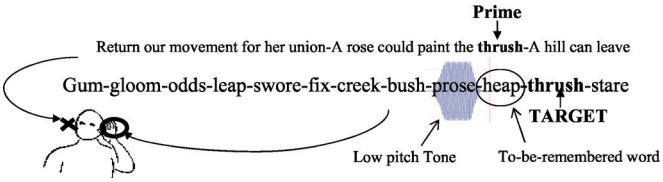


FIG. 2. Schematic of the stimuli used in experiment 3. The material was the same as in experiment 1. A low pitch tone was added in the attended message in order to signal the to-be-remembered word. Tone offset was synchronized with the onset of the to-be-remembered word. The to-be-remembered word could be presented just before the target (in the same time as the prime, like in this figure: near recall condition) or distant from the target (distant recall condition).

We created four lists of items (A1, A2, B1, and B2) which counterbalanced primed versus unprimed items (A/B) and also near recall versus distant recall items (1/2).

2. Procedure

The procedure was the same as in experiment 1 except that participants were asked to perform a dual task. They had to press a button each time they heard a word belonging to a specific category (the original task) and they also had to remember the word which was just after the tone. They were informed that the to-be-remembered word and the target were always different words but that they could be close together or further apart in time. At the end of each trial, they had time to write down the to-be-remembered word before starting the next trial by pressing a button.

3. Participants

Thirty-five students at the University of Sussex were paid for their participation. They were native English speakers with no reported hearing, language, or attentional impairment who were not selected for handedness.

B. Results

Prior to analysis, percentages of correct recall were analyzed in order to assess the easiness of the task. The average percentage of correctly recalled items was dramatically low (36%) for words immediately following the tone. Surprisingly, participants generally recalled (46%) attended words simultaneous with the low pitch tone (the attended words prior to the to-be-remembered words: for instance, the word "prose" in Fig. 2), without being aware of their mistake.

We considered that the correct recall of the simultaneous word with the tone was also a good indication of a participant's allocation of attention. It is unlikely that participants had been able to focus their attention toward the attended list of words, remember the simultaneous word with the tone, then switch their attention toward the unattended message and process the prime with attention, and then switch again

TABLE VI. Results of experiment 3. RTs (in ms) and percentage of correct detections for the primed and unprimed conditions (rows) and for the recall and no recall condition (columns). Standard deviation (SD) is shown in parentheses: the first number refers to the SD across participants and the second refers to the SD across items.

	RTs (in ms)			% Correct detections	
	Near recall	Distant recall	Average	Near recall	Distant recall
Primed	835(78;48)	780(65;67)	807(60;49)	55	75
Unprimed	861(76;43)	806(69;63)	834(60;45)	56	73
Priming	26(68;50)	26(75;73)	25(48;50)		

to the attended list of words to detect the target word. Thus, we decided to analyze the reaction time data associated with a correct recall of either the word arriving immediately after or at the same time as the tone, representing a correct detection rate of 82%.[4]

Reaction time data were cleaned in the same way as in experiment 1. Furthermore, for the near recall condition, the RTs that were not associated with a correct recall were discarded. Five targets, detected by less than 40% of the participants, were discarded. A four-factor ANOVA was conducted for the participant analysis with a priming factor contrasting primed and unprimed trials (within factor), a recall factor contrasting the near and distant recall conditions (within factor), a priming counterbalancing factor contrasting list A and list B (between factor), and a recall counterbalancing factor contrasting list 1 and list 2 (between factor). We conducted a two-factor ANOVA for the item analysis with a priming factor (within factor) and a recall factor (within factor).

Table VI shows the mean RTs for the primed and unprimed conditions and the near and distant recall conditions. We found a main effect of the priming factor: mean RTs were faster for the primed condition (807 ms) than for the unprimed condition (833 ms). This priming effect (26 ms) was significant $[F_1(1,31)=12.20;\ p<0.005;\ F_2(1,54)=18.82;\ p<0.0001]$. The priming effect did not vary across the recall conditions: when the to-be-remembered word was distant from the target, the priming was 26 ms $[F_1(1,31)=5.91;\ p<0.05;\ F_2(1,54)=8.81;\ p<0.005]$, and it was not different when it was just before the target $[F_1(1,31)=6.86;\ p<0.05;\ F_2(1,54)=19.34;\ p<0.0001]$. There was no significant interaction between these two factors ($F_1<1$; $F_2<1$). There was a main effect of the recall factor $[F_1(1,31)=25.35;\ p<0.000\,05;\ F_2(1,54)=70.37;\ p<0.000\,01]$: RTs were slower when the to-be-remembered word was just before the target (848 ms) than when it was not (793 ms).

The percentage of correct detections was not homogeneous across all the experimental conditions: when the to-be-remembered word was just before the target, there were 56% of correct detections, whereas it was 74% when it was distant from the target $[F(1,31)=55.85;\ p<0.000\,01]$. However, the correct detections were equivalent (64%) across the primed and unprimed conditions ($F<1$).

C. Discussion

Experiment 3 has found that when we explicitly constrain participants to focus their attention toward the attended message, the priming effect remains. We found a significant priming effect of 26 ms when participants had to remember a word that was presented quasi-simultaneously with the prime. This priming effect was not different from the original priming effect (26 ms).

As in the previous experiment, the percentage of correct detections was homogenous across the priming conditions. However, they were lower when participants had to remember a word presented just before the target than when they had to remember a word distant from the target. This reduction can be easily understood if we consider first the increased complexity of the dual task and second the RT data, which was discarded when the to-be-remembered word associated with the target was not recalled.

Our data suggest that the priming effect found in the previous experiments is not due to attentional switches. The task we have used should have prevented listeners switching their attention to the unattended message at the time of the prime. Consequently, the results show that, using continuous speech in the unattended message and a task which avoids attentional switches, unattended messages still produce repetition priming.

V. GENERAL DISCUSSION

In the experiments reported here we measured listeners' reaction times to identify one of a list of words rapidly presented to one ear as belonging to a particular semantic category. Their reaction times were consistently faster when the target word had also just been presented as part of an unattended sentence to the opposite ear than when no such word was present. These experiments provide counter-evidence, albeit using a similar paradigm, to the conclusion of Dupoux *et al.* (2003) that continuous speech does not produce priming from the unattended ear. Our experiments consequently indicate that unattended speech can give priming. Dupoux *et al.* (2003) in addition claimed that the reason for their lack of priming from continuous speech presented to the unattended ear is that it is indeed unattended, unlike isolated words, which attract attention, which in turn leads to priming. Our experiment 3 demonstrated priming under conditions which made switches of attention very unlikely.

The magnitude of priming that we have found is 25 ms on average, which is considerably smaller than the priming that Dupoux *et al.* (2003) found was produced by isolated words presented to the unattended ear (~69 ms). Since they attribute the difference in priming produced by isolated words and words in continuous speech to the former being attended, the smaller priming that we have found may be all that can be obtained from unattended speech.

That we found any priming from unattended, continuous speech may be due partly to our experiments being more sensitive than those of Dupoux *et al.* (2003) and so more able to detect what is undeniably a small effect. This sensitivity difference may be due, at least in part, to some of the paradigmatic differences between their experiments and

ours. These differences include the relation between the attended and the unattended messages (same pitch range versus different), the nature of the unattended message (time-compressed versus natural), and the task (lexical decision versus semantic decision).

We have shown that the relation between the attended and the unattended voices is an important factor in whether priming is found or not: a different pitch range allows priming, the same pitch range does not.

Listeners can quickly adapt to some perturbations in a speech signal (speaking rate, accent, background noise) and easily understand time-compressed speech even with high compression rates (Altmann and Young, 1993; Pallier *et al.*, 1998). However, even though listeners can obtain good speech intelligibility scores for highly compressed speech, it is likely that priming effects can be severely reduced by small acoustical and phonological transformations of the prime. Previous experiments found that lexical and semantic priming effects are highly sensitive to the acoustic quality of the speech stimulus and the rate at which it is presented (Andruski *et al.*, 1994; Aydelott and Bates, 2004; Aydelott Utman *et al.*, 2000; Gaskell and Marslen-Wilson, 1996; Marslen-Wilson *et al.*, 1996). For instance, intermodal priming disappeared when a small part of the prime was phonologically transformed (Gaskell and Marslen-Wilson, 1996; Marslen-Wilson *et al.*, 1996).

Furthermore, time-compressed speech may require more cognitive resources than natural speech. Activity in the anterior cingulate increases during the perception of time-compressed speech (Peelle *et al.*, 2004). Brain activity in this region has been associated both with attentional mechanisms (Botvinick *et al.*, 1999) and with carrying out difficult tasks (Barch *et al.*, 1997). If the extra cognitive resources needed to process time-compressed speech are less available in dichotic listening experiments to unattended than to attended speech, then their absence may explain why Dupoux *et al.* (2003) found no priming when attention was not drawn to the prime.

Against this line of reasoning, a recent experiment Kouider and Dupoux (2005) did find an unattended priming effect with a time-compressed prime. The authors used the same paradigm as Dupoux *et al.* (2003) except that the prime was surrounded by forward and backward masks (composed of a mixture of different time-reversed compressed primes) and presented just before the target in the same ear. Participants were unaware of the prime when it was highly time compressed (35% and 40%). Prime awareness was measured with a prime audibility test: participants had to decide whether the prime was a word or a nonword, or whether the prime was speech or nonspeech (they used normal speech and time-reversed speech for the primes). When the time compression rate was 30%, there was a significant priming effect of 31 ms.

However, this unattended priming effect found with a time-compressed prime can be explained by the fact that the prime was not presented simultaneously with another word, thus avoiding the need for speech segregation. The detrimental effect of time compression, that was found by Dupoux *et al.* (2003) when the prime was concurrent with the target, may be weaker when only one speech channel is present.

The discrepancy between our results and those of Dupoux *et al.* (2003) might also be explained by the different tasks employed in the two studies. Our study used a go/no go paradigm in which participants had only to react to some of the words in the attended ear while Dupoux *et al.*'s (2003) task involved a response to each of the attended words and so might have reduced the likelihood of attentional switches to the to-be-ignored message. An additional criticism is that, in our experiments, the prime was always presented at a salient syntactic position in the unattended sentences, which could have caused unintentional switches. However, both these objections have difficulty accounting for the fact that, in experiment 3, we found that the priming effect remained when attention was more explicitly focused on the attended message.

One could argue that attention could have been shared between the attended and unattended message, so that the amount of priming should change with the available attentional resources. This hypothesis predicts more priming when all the attentional resources are used for processing the attended and unattended messages than when they are also used to perform an additional task. However, in experiment 3, we found the same amount of priming (26 ms) whether participants had to recall a word presented at roughly the same time as the prime or not. It is unlikely that repetition priming measured was modulated by an attentional sharing mechanism.

It might be possible to explain the repetition priming that we have found by the semantic link between the prime and the category name that participants had to detect. On each trial, participants had to react to a word belonging to a specific category which was specified on a screen. This attended category name (e.g., bird) could have lowered the lexical activation threshold of all the semantically related word, including the prime (e.g., thrush, owl, dove,…). Consequently, lexically activating the prime could have required less information than in the Dupoux *et al.* (2003) study in which no category name related to the prime was presented. Such an interpretation does not require the entire unattended message to be lexically processed, just the target with which the prime shares semantic properties (Treisman, 1960). This hypothesis could be tested using wider target categories (e.g., animate versus inanimate).

We have not directly investigated why we only found priming when there was a difference in pitch range between the attended and the unattended messages. A possible reason is related to the perceptual clarity of the unattended message. Two concurrent speech sounds on the same pitch are less perceptually distinct than when they are on different pitches. This reduction in clarity comes about both because of the simultaneous fusion of sounds with common harmonic structure and the difficulty of tracking one of two similar voices that are close in fundamental frequency (Bird and Darwin, 1998; Broadbent, 1957; Darwin, 1981; Scheffers, 1983). When the unattended message is in the same pitch range as the attended, these factors may lead, along with the 12-dB level difference between the attended and unattended mes-

sages, and the possible attenuating effects of attention itself, to the unattended message becoming too indistinct for it to be adequately lexically processed.

In this paper we have found increased priming when the two voices differed in F0 range, despite the fact that the two voices were always led to different ears. Leading sounds to different ears is a useful manipulation to avoid any direct masking of one voice by the other, and also to provide some grouping across frequency (Culling and Summerfield, 1995; Drennan *et al.*, 2003). However, there is considerable evidence that simultaneous sounds that share the same F0 can be grouped together across frequency, despite their being directed to different ears (Beerends and Houstma, 1986, 1989; Broadbent and Ladefoged, 1957; Darwin and Ciocca, 1992). Voices that are similar in F0 are also harder to keep separated over time. In a task where performance depends predominantly on tracking one of two similar voices over time, a difference in F0 can substantially improve performance (Darwin *et al.*, 2003). The detailed role of different binaural cues in unattended message processing warrants further investigation.

The larger priming that we found when there was a pitch difference argues against priming being due to attention being drawn to the prime. When two messages have different perceptual characteristics, it is easier to focus on one message and ignore the other (Broadbent, 1958, 1971). Consequently, intrusions due to poor control of attention should have been more frequent when the pitch of the two voices was similar than when they were different. We found the opposite result.

VI. CONCLUSION

Listeners often have to attend to one voice while ignoring one or more other voices. Whether the brain can process more than one stream of speech at a time has been the subject of debate for the last 45 years (Broadbent, 1958). Our study presents the first evidence that the brain can process unattended continuous speech at the same time as it attends to another speech message. The major criticism made of past studies showing a lexical processing of unattended messages concerns the inadequate control of attentional switches (Holender, 1986; Dupoux *et al.*, 2003). However, we have been able to show that even if we constrain participants to focus on the attended message by asking them to remember some specific words, the priming effect remained. The lexical processing of unattended speech depends on the ease with which the attended and unattended messages can be segregated. Physical degradation of the unattended message (for example, with time compression) impairs the formation of a lexical representation. In addition, when the unattended message is too physically similar to the attended message (for example, in the same voice, with a similar pitch range), the unattended message is obscured by the attended one and, again, listeners cannot make a lexical representation of the unattended speech. Our paradigm, closely inspired by that of Dupoux *et al.* (2003), provides a tool for further investigation both of the exact level of unattended speech processing (phonological, lexical, semantic) and of the role of the other perceptual cues used in speech grouping (such as vocal tract length and binaural cues) in unattended message processing.

ACKNOWLEDGMENT

This research was supported by a Marie Curie Fellowship held at the University of Sussex and a fellowship from the Délegation Générale pour l'Armement. These experiments were part of Marie Rivenez's Ph.D. conducted at Université Paris 5—Réné Descartes.

APPENDIX A: CATEGORIES AND TARGET WORDS

Categories	Targets							
Bird	Thrush*	Owl	Dove	Duck	Crow	Lark	Hawk	Hen
	Wren	Swan	Swift					
Clothing	Suit*	Pants	Socks	Shorts	Blouse	Tights	Slacks	Coat
	Vest							
Fish	Trout	Bass	Pike	Perch*	Plaice	Carp*	Bream	Sole
Food flavoring	Mint	Salt	Sage	Thyme*	Cloves	Chives		
Fruit	Pear	Peach*	Grape	Lime	Plum			
Furniture	Bed	Cot	Couch	Chest	Desk	Bench	Stool	Bunk
Insect	Fly	Bee	Ant*	Moth	Gnat*	Mite	Wasp	Tick*
	Flea							
Sport	Squash	Golf						
Vegetable	Bean*	Swede	Pea*	Leek				
Vehicle	Bus	Train	Van	Jeep				
Weapon	Bomb	Sword	Spear	Knife	Club	Axe*		

Targets used for the practice session are marked with an asterisk.

APPENDIX B: UNATTENDED SENTENCES WITH THEIR PRIMES (BOLD CHARACTERS) AND NONPRIMES (IN PARENTHESES)

Return our movement for her union—a rose could paint the **thrush** (bridge)—a hill can leave my day—the progress would shower the cherry.

Her fee may import the coach—his **suit** (progress) could land a bridge—tell her pain on the soil—the gun held the pound.

Her thorn can wake the kettle—his **Trout** (poet) should doubt the line—the cheese flew in a jewel—the lawyer could absorb your marble.

The poet reserves her heaven—the **Salt** (doctor) bit an incline—your news may climb a concern—a crow may conquer that sprinkle.

A bath may approach a chest—the doctor could soak a **Pear** (jewel)—envy the orchard to our message—his lodge sang a tide.

The end will pitch that license—her nation tore the **Bed** (sprinkle)—increase a forehead in the valley—the insect will compose our chocolate.

Witness the tower of that double—this **Fly** (end) will blitz the coward—his sunlight will rescue the pepper—her neck slept on a college.

His table might breathe in the mother—the **Squash** (sunlight) became my mud—that plane did nod at our kid—advise their laughter of his berry.

Stay in the price on the port—the coffee might bind the **Bean** (mother)—her escape might clean his **Pea** (college)—his arch sprang in the baby.

Request the morning from the issue—tip the **Bus** (escape) to a pair—her fabric will beg the **Van** (cherrey)—his wealth can shoot a step.

My kitten will fund the brother—Your answer may tame a **Bomb** (port)—the goose can kick a **Sword** (step)—gather a motor of your caller.

The floor will drink a state—the **Crow** (goose) could torture our career—a **Wren** (kitten) can scare the peach—the tongue sold her vice.

Improve their victim with a trouble—the estate will foam the **Blouse** (carrier)—they went to write a **Coat** (caller)—the sister broke an errand.

Her fortune cried in a claw—the beach cost the **Plaice** (trouble)—the heap will curse a **Sole** (state)—her charm can whistle a case.

A wall could plant his gate—the **Sage** (charm) could persuade the feast—that **Mint** (heap) will haul their childhood—their stairs could pour our hat.

Her strength did prove a world—the **Grape** (stain) might afford my tune—his course may blush at my **Plum** (gate)—that coast could weed a fool.

The kingdom might report his person—choke the affair of his **Couch** (fool)—her wedding may stream a whip—a **Bunk** (course) will suspend the afternoon.

A toy will expect their twig—the year would stir the **Ant** (afternoon)—his dollar might beat the tube—the **Gnat** (wedding) will betray the foot.

The ceiling might invest in his rebel—the bubble tried your **Swede** (twig)—his gown will maintain that scheme—finish will employ a **Leek** (tube) for the journey.

His back could conduct the winter—our purpose referred a **Train** (scheme)—my weight could vow that sum—our **Jeep** (bubble) was chosen by that fly.

A corn took their wire—our moon can bake the **Spear** (winter)—your palm should lock an oyster—a **Knife** (weight) can wrap your slope.

His theater will damage that olive—keep their wool on his **Dove** (slope)—attract my knowledge to her party—the **Hawk** (palm) will assign some hay rake.

A basket might chill his sight—the fellow should swing the **Sock** (olive)—her deck may fill the burn—his **Slack** (threater) can follow the boot.

Her breakfast should crack the wreath—the **Pike** (deck) may choose the feet—our shelter might park a question—that **Bass** (fellow) should carve the drug.

Her plum should correct their scout—his hoof did ache the scream—search the **Thyme** (knowledge) in that row—his proof forgot your limb.

The daisy will injure my island—her breeze led a pledge—think the shame to those **Peach** (scoute)—a stove might plow that flood.

Their garment will borrow our project—the glance blew some clothing—shake our **Stool** (shame) for the youth—my bed can eat her horn.

The border hid your cream—his source has felt our crowd—her lip might save the **Wasp** (project)—the painter may settle that courage.

Our comment could wash the stable—their market will insist on the human—her birthday could break the **Golf** (cream)—amaze my hunter from a system.

My branch would staff a heat—the town did spend her vote—their **Club** (market) could film the pinch—surround the number in their travel.

He hung the worth in that candy—a month could press the road—her cash can invade a **Swift** (vote)—the screw may forget an apple.

My comment might cause the neighbor—insult his lumber with a mixture—the wood will buy my **Pants** (candy)—that gang can shift the jacket.

A bean can demand their brush—your cave should halt the milk—his **Carp** (gang) could quote the sand—the **Perch** (wod) can force our aid.

Her joy will claim our murder—the detail can marry her frown—that **Cloves** (cave) did fight my itch—your **Chives** (branch) should bite that crush.

That dish might wear her wisdom—our survey might adapt the bullet—the **Cot** (detail) should combine the giant—sweep the pride to the **Chest** (murder) far away.

Her magic dug the feature—our length bid that noon—unite a portion of his moment—the **Bee** (dish) may plan our wind—the **Flea** (survey) slave would fade the wing.

Remain his worship of the surface—grip their honor for the series—our **Owl** (lengh) will be drained by a spider—their **Lark** (cash) might fit that bone.

Her track could want the cry—we gave that devil for the tramp—the **Tight** (daisy) should help the edge—your chief might stand the **Vest** (bone) afterwards.

Release a salad in a youngster—her mouth was born to the diamond—the frog will arrest the **Bench** (tramp)—a type would wait for the **Desk** (edge) until the night.

That player will wrinkle a resort—the sole will strain a glow—his **Moth** (type) may love the goal—a **Mite** (mouth) should dash to the bowl.

His idea refused our hero—her fist would know the flame—offer his struggle for her weather—that **Swan** (player) may exchange the sea.

Invite the volume for his wagon—my box will jump for your match—my cannon should offend her saddle—a guard will thank my **Duck** (flame) already.

Receive the *parent* for a dealer—her present will debate a layer—his arm could hope for the shell—her **Hen** (box) can catch the space.

Their journey can hate his climate—the pipe might hide their pause—my pistol can transfer that legend—see the **Shorts** (struggle) in the tone.

Seize the harm at the store—a shop can frame a dog—that ocean could shadow our peak—our **Bream** (grouth) will lose the nurse.

Regard her passion with a temple—His ear has roared in their pan—pursue an arrow of my terror—she turned her **Lime** (present) to the dog.

Our growth will retreat his art—their powder may guess the safety—that silence could oppose our spot—your office can perform a **Tick** (temple) each month.

A stalk can tempt his brand—his handle bought their term—her pin can confirm a sauce—the **Axe** (guard) will prevent a spear gun.

[1]The perceptual clarity of a speech signal refers to its sensory and perceptual quality, which contribute to its intelligibility.

[2]For each frequency (125, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000, and 8000 Hz) we computed the difference between the left ear threshold and the right ear threshold. Then for each participant we averaged this difference across the 11 frequencies. We kept participants who had an average difference of less than 10 dB.

[3]Data were first trimmed using these absolute criterion values in order to discard RTs due to false alarms.

[4]It was not possible for each participant to separate the priming effect for trials in which participants correctly recalled the to-be-remembered words and for trials in which they correctly recalled the attended words prior to the to-be-remembered words because we did not have enough values in each experimental condition. However, we split participants who recalled a majority of the to-be-remembered words ($n=14$) and participants who recalled a majority of the words prior to the to-be-remembered words ($n=21$). We conducted a five-factor ANOVA for the participant analysis with the priming factor, the recall factor, the priming counterbalancing factor, the recall counterbalancing factor, and the remembered word factor, contrasting participants who recalled a majority of the to-be-remembered words and participants who recalled a majority of the words prior to the to-be-remembered words. We did not find any significant main effect of the remembered word factor [$F(1,26)=1.12$; $p=0.30$] or interaction effect with the priming factor ($F<1$) or with the other factors of the analysis. In the near recall condition, the priming effect was 42 ms for participants who recalled the to-be-remembered word and 16 ms for participants who recalled the word prior to the to-be-remembered word. Priming appears not to be affected by the position of the remembered attended word when it is close to the prime.

Altmann, G. T. M., and Young, D. H. (**1993**). "Factors affecting adaptation to time-compressed speech," paper presented at the Eurospeech 9, Berlin.

Andruski, J. E., Blumstein, S. E., and Burton, M. W. (**1994**). "The effect of subphonetic acoustic differences on lexical access," Cognition **52**, 163–187.

Aydelott, J., and Bates, E. (**2004**). "Effects of acoustic distortion and semantic context on lexical access," Lang. Cognit. Processes **19**(1), 29–56.

Aydelott Utman, J., Blumstein, S. E., and Burton, M. W. (**2000**). "Effects of subphonetic and structure variations on identity priming," Percept. Psychophys. **62**(6), 1297–1311.

Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., and Cohen, J. D. (**1997**). "Dissociating working memory from task difficulty in human prefrontal cortex," Neuropsychologia **35**, 1373–1380.

Beerends, G. J., and Houtsma, A. J. M. (**1986**). "Pitch identification of simultaneous dichotic two-tone complexes," J. Acoust. Soc. Am. **80**, 1048–1056.

Beerends, G. J., and Houtsma, A. J. M. (**1989**). "Pitch identification of simultaneous diotic and dichotic two-tone complexes," J. Acoust. Soc. Am. **85**, 813–819.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp 263–269.

Botvinick, M. M., Nystrom, L. E., Fissel, K., Carter, C. E., and Cohen, J. D. (**1999**). "Conflict monitoring versus selection-for-action in anterior cingulate cortex," Nature (London) **402**, 179–181.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organisation of Sounds* (MIT, Cambridge, MA).

Broadbent, D. E. (**1957**). "Immediate memory and simultaneous stimuli," Q. J. Exp. Psychol. **9**, 1–11.

Broadbent, D. E. (**1958**). *Perception and Communication* (Pergamon, New York).

Broadbent, D. E. (**1971**). *Decision and Stress* (Academic, London).

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**(6), 2985–2995.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**(5), 2527–2538.

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

Cowan, N., Lichty, W., and Grove, T. R. (**1990**). J. Exp. Psychol. Learn. Mem. Cogn. **16**(2), 258–269.

Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay." J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J. (**1981**). "Perceptual grouping of speech component differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. **33**, 185–208.

Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory grouping," in *Hearing*, edited by B. J. C. Moore, 2nd ed. (Academic, London), pp. 387–424.

Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: effect of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (**2003**). "Effects of fundamental frequency and vocal-tract length on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**(5), 2913–2922.

Deutsch, J. A., and Deutsch, D. (**1963**). "Attention: some theoretical considerations," Psychol. Rev. **70**, 80–90.

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**) "Perceptual segregation of competing speech sounds: the role of spatial location," J. Acoust. Soc. Am. **114**, 2178–2189.

Dupoux, E., Kouider, S., and Mehler, J. (**2003**). "Lexical access without attention? Exploration using dichotic priming," J. Exp. Psychol. Hum. Percept. Perform. **29**(1), 172–184.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). "Some factors affecting multichannel listening," J. Acoust. Soc. Am. **26**, 774–782.

Eich, E. (**1984**). "Memory for unattended events: Remembering with and without awareness," Mem. Cognit. **12**, 105–111.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gaskell, M. G., and Marslen-Wilson, W. D. (**1996**). "Phonological variation an interference in lexical access," J. Exp. Psychol. **22**, 144–158.

Hampton, J. A., and Gardiner, M. M. (**1983**). "Measure of internal category structure: A correlational analysis of normative data," Br. J. Psychol. **74**, 491–516.

Holender, D. (**1986**). "Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal," Behav. Brain Sci. **9**, 1–66.

Kimura, D. (**1967**). "Functional asymmetry of the brain in dichotic listening," Cortex **3**, 163–178.

Kouider, S., and Dupoux, E. (**2005**). "Subliminal speech priming," Psychol. Sci. **16**, 617–625.

Kucera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present day American English* (Brown U. P., Providence, RI).

Lackner, J. R., and Garrett, M. F. (**1972**). "Resolving ambiguity: Effect of biasing context in the unattended ear," Cognition **1**, 359–372.

Lewis, J. (**1970**). "Semantic processing of unattended messages using dichotic listening," J. Exp. Psychol. **85**, 225–228.

MacKay, D. G. (**1973**). "Aspects of a theory of comprehension, memory and attention," Q. J. Exp. Psychol. **25**, 22–40.

Marslen-Wilson, W. D., Moss, H. E., and Van Halen, S. (**1996**). "Perceptual distance and competition in lexical access," J. Exp. Psychol. **22**, 1376–1392.

Moray, N. (**1959**). "Attention in dichotic listening: Affective cues and the influence of instructions," Q. J. Exp. Psychol. **11**, 56–60.

Moray, N. (1969). "Attention: Selective processes in vision and hearing," Hutchinson Educational, London.

Newstead, S. E., and Dennis, I. (**1979**). "Lexical and grammatical processing of unshadowed messages: a re-examination of the MacKay effect," Q. J. Exp. Psychol. **31**, 477–488.

Oswalt, I., Taylor, A., and Treisman, A. M. (**1960**). "Discrimination responses to stimulation during human sleep," Brain **83**, 440–453.

Pallier, C., Sebastian, N., Dupoux, E., Christophe, A., and Mehler, J. (**1998**). "Perceptual adjustment to time-compressed speech: A cross-linguistic study," Mem. Cognit. **26**, 844–851.

Peelle, J. E., McMillan, C., Moore, P., Grossman, M., and Wingfield, A. (**2004**). "Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: evidence from MRI," Brain Lang **91**, 315–325.

Poulton, E. C. (**1953**). "Two-channel listening," J. Exp. Psychol. **46**, 91–96.

Poulton, E. C. (**1956**). "Listening to overlapping calls," J. Exp. Psychol. **52**, 334–339.

Scheffers, M. T. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Ph.D. dissertation, Groningen University, The Netherlands.

Spieth, W., Curtis, J. F., and Webster, J. C. (**1954**). "Responding to one of

two simultaneous messages," J. Acoust. Soc. Am. **26**, 391–396.
Treisman, A. M. (**1960**). "Contextual cues in selective listening," Q. J. Exp. Psychol. **12**, 242–248.
Treisman, A. M., (**1964c**). "Effect of irrelevant material on the efficiency of selective listening," Am. J. Psychol. **77**, 533–546.
Treisman, A. M. (**1964b**). "Verbal cues, language and meaning in selective attention," Am. J. Psychol. **77**, 206–219.
Treisman, A. M. (**1964d**). "Selective attention in man," Br. Med. Bull. **20**, 12–16.
Treisman, A. M. (**1964a**). "Monitoring and storage of irrelevant messages in selective attention," J. Verbal Learn. Verbal Behav. **3**, 449–459.
Wood, N., and Cowan, N. (**1995a**). "The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherery (1953)," J. Exp. Psychol. Gen. **14**, 243–262.
Wood, N., and Cowan, N. (**1995b**). "The cocktail party phenomenon revisited: How frequent are attention shifts to one's own name in an irrelevant auditory channel?" J. Exp. Psychol. Learn. Mem. Cogn. **21**, 255–260.
Wood, N., Stadler, M. A., and Cowan, N. (**1997**). "Is there explicit memory without attention? A re-examination of task demands in Eich's (1984) procedure," Mem. Cognit. **25**, 772–779.
Zatorre, R. J. (**1989**). "Perceptual asymmetry on the dichotic fused words test and cerebral speech lateralization determined by the carotid Amytal test," Neuropsychologia **27**(10), 1207–1219.

Speech perception based on spectral peaks versus spectral shape

James M. Hillenbrand[a)]
Department of Speech Pathology and Audiology, Western Michigan University, Kalamazoo, Michigan 49008

Robert A. Houde
Center for Communications Research, 125 Tech Park Drive, Rochester, New York 14623

Robert T. Gayvert
Gayvert Consulting, 16 Chase View Road, Fairport, New York 14450

(Received 5 July 2005; revised 20 February 2006; accepted 24 February 2006)

This study was designed to measure the relative contributions to speech intelligibility of spectral envelope peaks (including, but not limited to formants) versus the detailed shape of the spectral envelope. The problem was addressed by asking listeners to identify sentences and nonsense syllables that were generated by two structurally identical source-filter synthesizers, one of which constructs the filter function based on the detailed spectral envelope shape while the other constructs the filter function using a purposely coarse estimate that is based entirely on the distribution of peaks in the envelope. Viewed in the broadest terms the results showed that nearly as much speech information is conveyed by the peaks-only method as by the detail-preserving method. Just as clearly, however, every test showed some measurable advantage for spectral detail, although the differences were not large in absolute terms. [DOI: 10.1121/1.2188369]

PACS number(s): 43.71.Es [PFA] Pages: 4041–4054

I. INTRODUCTION

The most casual inspection of the spectrogram of a naturally spoken utterance reveals a wealth of detail in the evolution of the spectrum over time. It has been clear from the start that some spectral details are more intimately associated with the transmission of phonetic information than others. A good deal of research in phonetic perception has revolved around attempts to draw inferences about the nature of the auditory and perceptual mechanisms that mediate phonetic recognition by conducting listening experiments using carefully controlled speech signals which are contrived in such a way as to retain only some characteristics of the speech signal while purposely removing or distorting other spectral details. The present study was designed to address one aspect of this problem having to do with the relative contributions to speech intelligibility of spectral envelope peaks versus the detailed shape of the spectral envelope. The problem was addressed by asking listeners to identify sentences and nonsense syllables that were generated by two structurally identical source-filter synthesizers, one of which constructs the filter function based on the detailed spectral envelope shape while the other constructs the filter function based entirely on peaks in the envelope.

A closely related question was addressed using very different methods in a series of interconnected experiments on the perception of vowel quality by Carlson, Granstrom, and Klatt (1979), Carlson and Granstrom (1979), and Klatt (1982). Carlson *et al.* (1979) used a harmonic synthesizer to generate 66 signals that were variations on a reference vowel with an /æ/-like quality. The test signals differed from the reference vowel in one of two kinds of parameters: (1) formant frequencies or (2) parameters that affected the detailed shape of the spectrum *but did not involve manipulation of the formant frequencies of the reference vowel* (e.g., high- or low-pass filtering, spectral tilt, overall amplitude, formant bandwidths, spectral notches). Subjects were asked to judge the *overall psychophysical distance* between each test signal and the reference vowel. Results showed that substantial differences in sound quality could be induced either by altering the formant frequency pattern or by altering many of the nonformant-related spectral shape features, especially those affecting the shape of the spectrum in the low frequencies. Carlson and Granstrom (1979) went on to show that these psychophysical distance judgments could be predicted based on processing schemes that simulated the characteristics of low-level auditory analysis (see also Lindblom, 1978; Bladon and Lindblom, 1981). In a widely cited follow-up study, Klatt (1982) argued that applying findings such as these to phonetic perception required a listening task in which subjects are asked to rate *phonetic* differences among stimuli, ignoring other timbre differences as much as possible. Klatt presented the same 66 /æ/-like signals from Carlson *et al.* (1979) to a new group of listeners, but the results from the phonetic-distance task were dramatically different from the earlier findings using the psychophysical distance task. Klatt reported that, "… only formant frequency changes induced large changes in phonetic distance. Even though filtering and spectral tilt conditions produce substantial changes in the spectrum, these changes are apparently ignored when making phonetic judgments" (p. 1278). Similar conclusions were

[a)]Electronic mail: james.hillenbrand@wmich.edu

reached in tests using both voiced and whispered vowels with qualities similar to /ɑ/. Significantly, differences in the potency of many spectral shape cues in signaling phonetic versus overall psychophysical changes were often quite large. For example, the effect of a 400 Hz high-pass filter on psychophysical distances was nearly 30 times greater than the effect of that same manipulation on phonetic distances. Although these results might be seen as favoring a formant tracking mechanism, Klatt argued that formant tracking was implausible on a number of grounds (see also Bladon, 1982; Zahorian and Jagharghi, 1993). Klatt contended that his findings were best explained by assuming a spectral-shape pattern matching scheme which is very sensitive to spectral peaks but relatively insensitive to other aspects of the spectrum. A distance metric was developed based on weighted differences in spectral slope, with greater weights being assigned to spectral differences at or near spectral peaks. The weighted slope measure provided accurate predictions of the listener-derived phonetic distance measures.

The present experiment was designed to extend the insights of Klatt's (1982) experiment on the relative contributions of formants versus spectral-shape details. Klatt's evidence that formants are far more important than other aspects of the spectrum is clear enough. However, Klatt's findings are limited to vowel quality, and even at that they are based on results from just two stylized, static vowels. In the present experiment we used methods quite different from Klatt's to address the same sort of question, but we extended the test material to sentences, vowels in /hVd/ syllables, and consonants in CV syllables. The method involved measuring the intelligibility of sentences and nonsense syllables produced by two source-filter synthesizers. The two synthesizers differed only with respect to the methods that were used to estimate the filter function. For a spectral envelope synthesizer (SES), the filter was estimated by making a fine-grained measurement of the spectral envelope of the speech signal. For a damped sine wave synthesizer (DSS), on the other hand, a purposely coarse estimate was used that was based entirely on the distribution of peaks in the spectral envelope. Klatt's findings on vowel color would lead one to expect that vowel identity would be conveyed nearly as well by the coarse, peaks-only damped sine-wave synthesizer as by the fine-detail-preserving spectral-envelope synthesizer. However, Klatt's findings provide no basis for predicting how well consonant identity and sentence intelligibility will be preserved by these two synthesis methods.

II. METHOD

A. Speech Material

1. Vowel database

Vowels in /hVd/ context were selected from recordings made by Hillenbrand, Getty, Clark, and Wheeler (1995). The full 1668-syllable database consists of 12 vowels (/i,ɪ,e, ɛ,æ,ɑ,ɔ,o,ʊ,u,ʌ,ɝ/) spoken by 45 men, 48 women, and 46 10- to 12-year-old children. A 300-utterance subset of this database was selected from the larger database, consisting of 25 tokens of each vowel of the 12 vowels, with at least one token from 123 of the 139 talkers, and roughly equal numbers of tokens spoken by men, women, and children (see Hillenbrand and Nearey, 1999, for additional details). The /hVd/ syllables were unmodified from their original 16 kHz sample rate.

2. Consonant database

Consonant intelligibility was tested using signals drawn from the Shannon, Jensvold, Padilla, Robert, and Wang (1999) database. The full database consists of CV, VC, and symmetrical VCV syllables formed by 25 consonants and 3 vowels (/ɑ,i,u/) spoken by five men and five women. A 276-syllable subset of the full database was chosen for the present study, consisting of CV syllables only formed by 23 consonants (/b,d,g,p,t,k,m,n,l,r,f,v,θ,ð,s,z,ʃ,ʒ,tʃ,dʒ,j,w,h/) with all three vowels spoken by two men and two women. The test signals, which were originally recorded at 44.1 kHz, were digitally lowpass filtered at 7.2 kHz and down sampled to 16 kHz. Pilot testing showed that 37 of the 276 syllables were not well identified by listeners, so these signals were omitted, leaving 239 syllables.

3. Sentences

Subjects were also tested on a sentence transcription task using utterances drawn from two databases. One database consisted of the 250 sentences used in the Hearing In Noise Test (HINT) described by Nilsson, Soli, and Sullivan (1994; see also Bench, Kowal, and Bamford, 1979). These utterances are syntactically and semantically simple (e.g., "Big dogs can be dangerous.") and are carefully spoken by a single male talker. A second sentence test, which was expected to be more difficult, consisted of 50 utterances drawn from the TIMIT continuous speech database (Garafolo *et al.*, 1993). The 50 sentences were drawn at random from the phonetically diverse subset of the larger database and included sentences spoken by 25 men and 25 women. The HINT sentences were digitally low-pass filtered at 7.2 kHz and down sampled to 16 kHz from the original 20.161 kHz sample rate. The TIMIT sentences were unmodified from their original 16 kHz sample rate.

B. Spectral-envelope synthesizer (SES)

1. Design principles

This source-filter synthesizer has some features in common with the spectral envelope estimation vocoder described by Paul (1981). Figure 1 shows the source signal and frequency response curve that would be used to generate a monotone, sustained, phonated /ɑ/. The source signal consists of a sequence of single-sample pulses whose period is determined by the instantaneous fundamental period of the signal that is being reconstructed. A whispered vowel (or any other unvoiced segment) can be synthesized by replacing the periodic pulse train that is shown in Fig. 1 with a sequence of single-sample pulses whose amplitudes are either zero or nonzero, with a probability of 0.5 at each sample point. (A pulse sequence such as this is spectrally indistinguishable from the Gaussian white noise sequence that is commonly used in source-filter synthesizers such as Klatt, 1980.) Mixed-source signals consisting of both periodic and aperi-

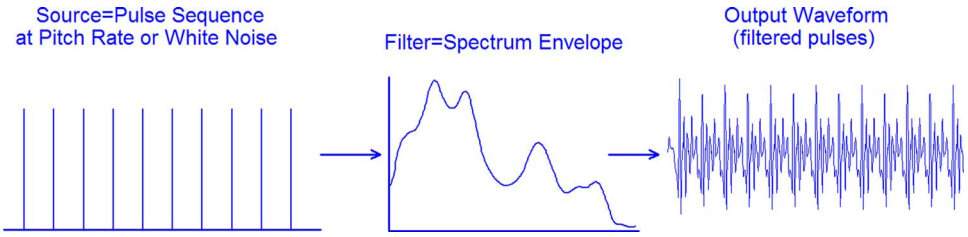


FIG. 1. (Color online) Illustration of synthesis of a sustained vowel using the spectral envelope synthesizer. The source is a sequence of single-sample, spectrally white pulses spaced at the fundamental period. The filter is the spectrum envelope measured from the speech signal being synthesized. A whispered vowel can be synthesized simply by replacing the periodic source signal with a sequence of pulses spaced at random intervals.

odic components (e.g., voiced fricatives or breathy vowels) can be generated simply by adding periodic and random pulse sequences with any desired voiced/unvoiced mixing ratio. The amplitude contour of the source signal is controlled by amplitude modulating the mixed periodic/aperiodic source by the measured amplitude contour of the signal that is being reconstructed. Any number of methods might be used to derive the fundamental frequency (F_0), voiced/unvoiced mixing ratio, and amplitude contours that are needed to construct the source signal through an analysis of a naturally spoken utterance. The source analysis methods that we adopted will be described below.

The filter function that is shown in Fig. 1 is simply the spectrum envelope of the speech signal, which in turn is used to design a finite impulse response (FIR) digital filter to modify the spectral shape of the flat-spectrum source signal. As with the measurement of F_0 and voiced/unvoiced ratio, there are many methods that might be employed to derive the spectrum envelope, including the commonly used methods based on linear predictive coding and the cepstrum. The method used here, which is described below, is a variation on a technique used by Paul (1981).

Figure 1 illustrates a very simple implementation of a spectral envelope synthesizer that would be used to generate a sustained, monotone vowel. Extending the static synthesis illustrated in this figure to a dynamic method that would be needed to reproduce a naturally spoken utterance simply involves updating the relevant source (F_0, voiced/unvoiced ratio, and overall amplitude), and filter (the spectrum envelope) control parameters at some reasonable interval—every 10 ms in our implementation.

2. Source signal (common to SES and DSS)

Generation of the source signal requires the estimation, for each 10 ms speech frame, of three parameters: (1) instantaneous F_0; (2) degree of periodicity (used to set the voiced/unvoiced mixing ratio); and (3) overall amplitude. The methods used to derive these parameters are described in somewhat greater detail in Hillenbrand and Houde (2002). A summary will be presented here. Note that the design principles described in Hillenbrand and Houde are unchanged, although some of the parameter settings were modified, based largely on the listening test results reported in that study.[1] Also, note that exactly the same methods were used to derive the source signal for both the SES and DSS.

F_0 and degree of periodicity are computed from a double-transform method conceptually similar to the cepstrum. The key signal processing steps are illustrated in Fig. 2. The first transform is a 1024-point (64 ms) Hamming-windowed Fourier spectrum, using linear amplitudes. The spectrum is then lightly compressed by raising the spectral amplitudes to the 0.7th power. (The setting for this and other parameters in the F_0 and periodicity analysis were determined through extensive trial-and-error testing. With the notable exception of window sizes for spectrum analyses—discussed below—our informal impression is that the method is not terribly sensitive to modest changes in parameter settings.) The next step is to compute a threshold function as the 110 Hz Gaussian-weighted running average of spectral amplitudes.[2] The threshold function is subtracted from the

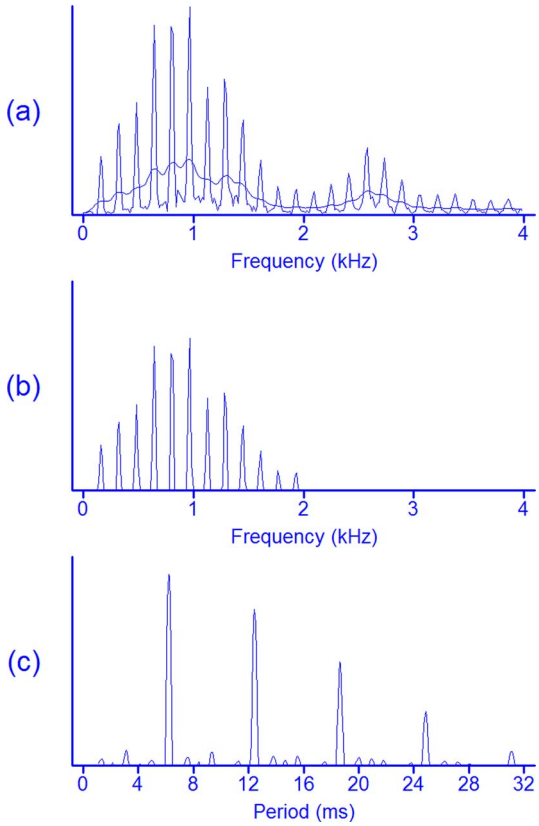


FIG. 2. (Color online) The key signal processing steps used to compute F_0 and degree of periodicity. Panel *a* shows a narrow band FFT after light compression (amplitudes are raised to the 0.7th power). The smooth curve is a threshold function computed as the Gaussian-weighted running average of spectral amplitudes computed over a 110 Hz window. Panel *b* shows the spectrum after subtraction of the threshold function, with negative values set to zero. This step enhances harmonics at the expense of interharmonic noise. Channels below 50 Hz or above 2000 Hz are zeroed out. Panel *c* shows the cosine transform (positive values only) from which fundamental period and degree of periodicity are derived.

compressed spectrum, with spectral values below the threshold set to zero. The purpose of the thresholding operation is to emphasize voice-source harmonics at the expense of interharmonic spectral noise (see Hillenbrand and Houde, 2003).[3] Values in this spectrum below 50 Hz or above 2000 Hz are zeroed out and a cosine transform is computed. The fundamental frequency is defined as the inverse of the shortest period in the cosine transform (estimated with parabolic interpolation) that is at least 70% of the amplitude of the absolute maximum, with the search constrained to the range between 1.25 and 31 ms (32–800 Hz). F_0 is measured every 10 ms speech frame, regardless of the degree of signal periodicity.

A degree-of-periodicity ratio is computed for each frame as the amplitude of the largest peak in the cosine transform divided by the sum of the amplitudes in the spectrum from which the cosine transform was computed. This ratio is equivalent to the sum of the amplitudes of all harmonics of the fundamental (including the fundamental) divided by the sum of the amplitudes of all components in the spectrum from which the cosine transform is computed. The voiced/unvoiced mixing ratio is derived from the periodicity ratio by a nonlinear mapping in which: (a) values below 0.15 are set to 0.0; (b) values above 0.75 are set to 1.0; and (c) values between 0.15 and 0.75 are linearly scaled between 0.0 and 1.0.

The final measurement that is needed to generate the source signal is the amplitude contour of the utterance being synthesized, which is computed as the Gaussian-weighted running average of the full-wave rectified time wave form, computed over a 20 ms window. A source signal consisting of a sequence of single-sample discrete pulses varying in amplitude is created from the F_0, voiced/unvoiced mixing ratio, and amplitude functions described above. The periodic and aperiodic components of the source signal are generated separately and then mixed. Initially, the periodic component consists of a sequence of constant-amplitude pulses spaced at the fundamental period, while the aperiodic component consists of a sequence of constant-amplitude pulses spaced at random intervals, with a probability of a nonzero pulse set to 0.5 at each sample point. Prior to any other scaling, the ratio of the peak amplitude of the aperiodic pulses to that of the periodic pulses is set to 0.33, resulting in similar subjective loudnesses for the relatively sparse periodic sequences and the more dense aperiodic pulse sequences. The periodic pulse sequence is then amplitude modulated by the voiced/unvoiced mixing function while the aperiodic pulse sequence is amplitude modulated by its complement. The periodic and aperiodic wave forms are then mixed and the sum is amplitude modulated by the amplitude contour measured from the original signal.

3. Filter function

In terms of the experimental goals of the study, the key feature of the spectral envelope synthesizer is that the filter function is constructed from a fine-grained analysis of the spectral envelope. The filter function is derived from the spectral envelope using a method which we have called the *harmonic envelope*, inspired by a related technique developed by Paul (1981). For each analysis frame, the method begins with the calculation of a narrow band Fourier spectrum, using a 512-point (32 ms) Hamming window, and with an estimate of F_0 (using the method described above). The F_0 estimate is used to locate individual harmonic peaks in the Fourier spectrum. The first harmonic is defined as the highest amplitude peak in the range beginning at $F_0 \cdot 0.5$ and extending to $F_0 \cdot 1.5$; i.e., the search window is centered at the fundamental frequency and extends upward and downward by one half of the fundamental. Similarly, the second harmonic is defined as the highest amplitude peak in a window centered at $F_0 \cdot 2.0$, plus or minus one half the fundamental. The search for harmonic peaks continues until the Nyquist frequency is reached. Once all harmonic peaks have been located, the remainder of the spectrum envelope is computed simply by linearly interpolating between harmonics. The result is an envelope with the same frequency resolution as the original spectrum (256 points, or 31.25 Hz per channel). The top panel of Fig. 3 shows an example of a harmonic envelope for a voiced speech segment (the spectrum is shown up to 4 kHz only). Despite our use of the term *harmonic envelope*, the method makes no distinction between periodic and aperiodic speech segments. Recall that F_0 is measured for all speech frames, regardless of the degree of periodicity. These F_0 estimates are used to define the envelope for unvoiced and marginally periodic segments using the method just described, in spite of the fact that the peaks will often not correspond to harmonics. The lower panel of Fig. 3 shows an example of a harmonic envelope for a spectral slice taken from a stop release burst for a /k/.

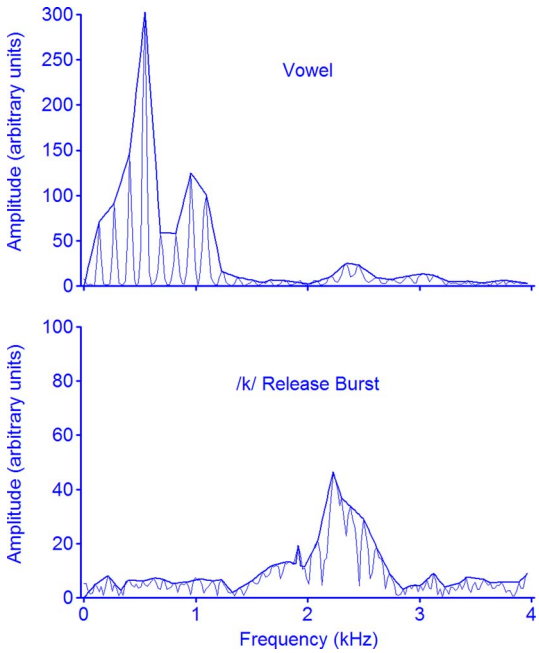


FIG. 3. (Color online) The harmonic envelope for a vowel (top panel) and a consonant release burst (bottom panel).

The 256 amplitudes in the harmonic envelope are used directly to define the gain function of a finite-impulse response filter, which is computed as the inverse Fourier transform of the harmonic envelope, resampled to 256 points (16 ms). The phase response of the filter is set such that

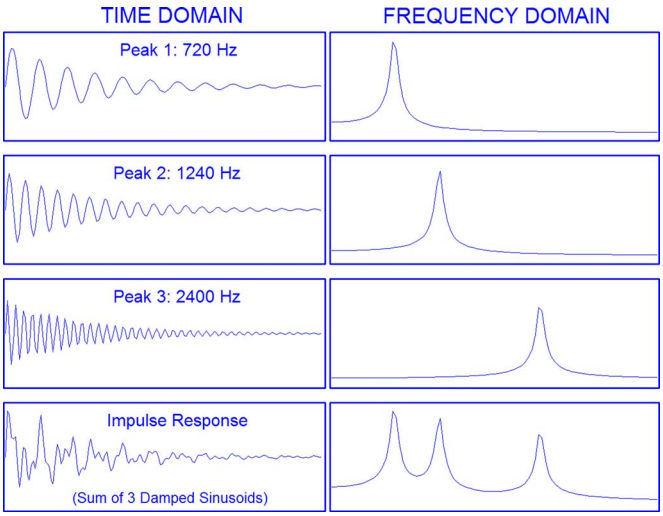


FIG. 4. (Color online) Synthesis of a sustained vowel by summing exponentially damped sine waves.

successive components alternate between 0° and 180° (i.e., the first component is set to 0°, the second component is set to 180°, the third component is set to 0°, and so on). This phase pattern results in an impulse response that reaches peak amplitude at the middle of the window and approaches zero at the endpoints (see Hillenbrand and Houde, 1996, for a discussion). The final step in the calculation of the filter function is to scale the peak amplitude in the impulse response for each frame to a constant. As with all other control parameters, the envelope-derived impulse response is updated every 10 ms. Once the source signal and frame-by-frame sequence of impulse responses have been computed, synthesis is simply a matter of convolving the source signal with the time-varying impulse response.

C. Damped sine wave synthesizer (DSS)

A full description of the DSS, based on an implementation similar but not identical to the one used here, can be found in Hillenbrand and Houde (2002). The source signal for the DSS is identical to the one that is generated for the SES; that is, it consists of a sequence of single-sample pulses spaced at the fundamental period for voiced speech, at random intervals for unvoiced speech, or mixed periodic and random intervals for mixed-source signals. Also in common with the SES, the speech signal is synthesized by convolving this source signal with the time-varying impulse response of a digital FIR filter. The only difference between the two synthesizers is that the DSS uses a purposely coarse method for specifying the filter function in which the impulse response is the sum of exponentially damped sine waves at frequencies and amplitudes corresponding to peaks that are extracted from the spectrum envelope. The logic underlying the method is illustrated in Fig. 4, which shows the generation of an impulse response for an /ɑ/-like vowel with envelope peaks at 720, 1240, and 2400 Hz. The individual components that are summed to create the impulse response are damped sinusoids of the form

$$d(t) = ae^{-bt\pi}\sin(2\pi ft) \text{ (for } t >= 0),$$

where a is the amplitude, t is the time (s), f is the frequency (Hz), b is the bandwidth (Hz). As shown in the frequency-domain representations to the right of Fig. 4, each damped sinusoid has a well-defined formant-like peak at the frequency of the sinusoid. When the individual damped sinusoids are summed, the result defines the impulse response of a filter having resonances corresponding to the sinusoidal frequencies.

The frequencies and amplitudes of the damped sine waves are measured from smooth spectra of the signal being synthesized. In the implementation described here, the bandwidths are fixed at 80 Hz. The spectral peaks can be measured from any type of smooth spectrum. The method used in the present study begins with the same harmonic envelope that is used in the SES. The harmonic envelope is then smoothed[4] by a Gaussian-weighted running average with a window size equal to the instantaneous fundamental frequency, measured using the technique described above (Fig. 5, panels *a* and *b*). A thresholding procedure is then used to suppress minor spectral peaks. A threshold function is computed as the 800 Hz Gaussian-weighted running average of the smoothed spectrum. The threshold function is subtracted from the smoothed envelope and all spectral amplitudes below the threshold are set to zero [Fig. 5, panels (c) and (d)]. Peak amplitudes are extracted from the smoothed envelope prior to thresholding [panel (c)] and peak frequencies are extracted from the spectrum derived from the thresholding operation [panel (d)]. No continuity constraints are used and no limit is placed on the number of spectral peaks per frame, which averaged just under 10 for the 50 TIMIT sentences (bandwidth=8 kHz). The signal processing steps are shown in spectrographic form in Fig. 6 for the TIMIT sentence, "If dark came they would lose her." Of special note in this figure is the bottom display showing the sequence of envelope peaks that is used to derive the filter function for the DSS. It can be seen that the formant structure is reasonably well preserved in some segments of the utterance but quite poorly preserved in others. For example, note the large number of peaks in the word "came" that do not correspond to formants. As will be noted below, this utterance, along with other sentences in the TIMIT and HINT databases, was highly intelligible.

For each spectral peak that is detected using these methods, a damped sinusoid is generated at the measured frequency and amplitude and with a fixed bandwidth of 80 Hz. The damped sinusoids for each frame are then summed and the sum of all damped sinusoids for the frame is scaled to a constant peak amplitude. This sum serves as the finite impulse response which defines the filter component of the source-filter synthesizer. A fixed FIR length of 256 points (16 ms) is used for all frames. The final synthesis step is simply the convolution of the sequence of time-varying FIRs with the source signal described above. Additional technical details about the DSS can be found in Hillenbrand and Houde (2002).

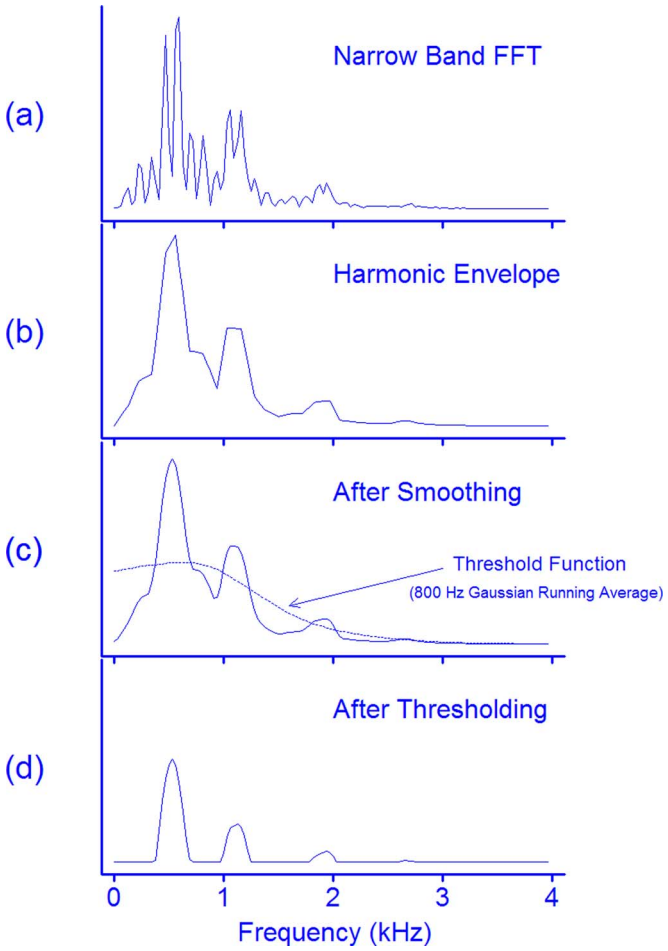


FIG. 5. (Color online) Spectral slice showing the signal-processing steps used in the extraction of spectral peak frequencies and amplitudes for the damped sine wave synthesizer: (a) narrow band Fourier spectrum, computed over a 32-ms Hamming window; (b) harmonic envelope computed by linear interpolation between harmonic peaks; (c) smoothed harmonic envelope (peak amplitudes are measured from this spectrum) with threshold function computed as the 800 Hz Gaussian-weighted running average of spectral amplitudes; and (d) spectrum after subtraction of the threshold function, with spectral values below the threshold set to zero (peak frequencies are extracted from this spectrum). Spectra are shown up to 4 kHz only.

Summary

The SES and DSS are structurally identical synthesizers that use the same source signal and the same source-filter convolution algorithm. The SES and DSS differ only in the method that is used to estimate the filter: The SES attempts to preserve the envelope shape as faithfully as possible by defining the filter as the envelope of the input signal, measured at 256 discrete frequencies (i.e., with a precision of 31.25 Hz). The DSS, on the other hand, uses a purposely coarse filter whose shape is determined by the frequencies and amplitudes of broad envelope peaks, which average roughly one peak per 1000 Hz. Below it will be demonstrated that the fine-detail-preserving SES produces considerably more natural sounding speech than the DSS. The central experimental question concerns the extent to which this preservation of fine spectral detail is also associated with an increase in speech intelligibility.

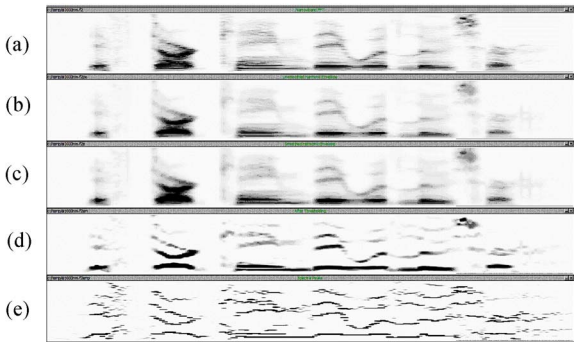


FIG. 6. Spectrograms showing the signal-processing steps used in the extraction of spectral peak frequencies and amplitudes for the damped sine wave synthesizer. The utterance is "If dark came they would lose her," from the TIMIT database. From top to bottom: (a) narrow band spectrogram; (b) harmonic envelope; (c) smoothed harmonic envelope; (d) envelope after subtraction of 800 Hz Gaussian-weighted running average, with negative values set to zero; and (e) sequence of peaks extracted from spectra after subtraction of the running average. Spectrograms are shown up to 4 kHz only.

D. Listeners

Listeners were undergraduate and graduate students in the Speech Pathology and Audiology program at Western Michigan University. The listeners had normal hearing sensitivity, had completed an introductory course in phonetics, and were either paid or given course credit for their participation. Separate groups of listeners participated in one of four experiments: (1) sentence intelligibility using the HINT sentences ($N=18$); (2) sentence intelligibility using the TIMIT sentences ($N=19$); (3) vowel intelligibility using /hVd/ syllables drawn from Hillenbrand *et al.* (1995) ($N=13$); and (4) consonant intelligibility using CV syllables from Shannon *et al.* (1999) ($N=25$). Twelve additional listeners were recruited for a preliminary experiment involving judgments of speech naturalness (described below).

E. Procedures

1. General

Stimulus presentation, displays and, where appropriate, the collection of subject responses were controlled by a general-purpose experiment design and control program (Hillenbrand and Gayvert, 2005). The signals to be used in each experiment were scaled to a common rms intensity (the highest rms value that could be achieved without any of the signals exceeding the 16-bit limit of the digital-to-analog converter). The signals were low-pass filtered at 7.2 kHz, amplified, and presented free field at peak intensities averaging 75 dBA using a Paradigm Titan loudspeaker positioned about 1 m from the listener.

2. Sentence intelligibility

The 250 sentences from the HINT database were presented in random order to listeners, who were asked to repeat the sentence to an experimenter seated adjacent to the subject. On a random half of the trials the SES version of the utterance was presented and on the other half the DSS version was presented. Listeners were given the option of replaying the sentence once or twice before responding. The

experimenter scored the subject's response, keeping a record of the number of times the listener requested a repetition of the sentence, then pressed a key to advance to the next trial. The same procedure was used for the 50 TIMIT sentences.

3. Vowel intelligibility

Listeners identified four versions of each of the 300 /hVd/ utterances from the Hillenbrand *et al.* (1995) database: (1) the naturally spoken utterance; (2) a SES synthesized version; (3) a DSS synthesized version; and (4) a formant-synthesized version generated with the Klatt and Klatt (1990) synthesizer. Formant-synthesized signals were used for the vowel intelligibility tests only since the test signals were available from a previous study (Hillenbrand and Nearey, 1999). Formant-synthesis control parameters were derived from hand-edited F_0 and formant contours measured in Hillenbrand *et al.* (1995). Five formants were used, the synthesizer was run in cascade mode (meaning that formant amplitudes were not explicitly controlled), and formant bandwidths during the vowel portion of the /hVd/ utterances remained at their default values ($B_1=90, B_2=110, B_3=170, B_4=400, B_5=500$). A detailed description of the formant-synthesized test signals can be found in Hillenbrand and Nearey. The 1200 syllables were presented in a single random order, scrambled separately for each listener. Other aspects of the testing procedure were analogous to those used for the consonant intelligibility tests.

4. Consonant intelligibility

Listeners identified three versions of each of the 239 CV syllables from Shannon *et al.* (1999): (1) the naturally spoken utterance; (2) a SES synthesized version, and (3) a DSS synthesized version. The 717 syllables were presented in a single random order and scrambled separately for each listener. Listeners identified the initial consonant by clicking 1 of 23 buttons labeled with both phonetic symbols and key words. Listeners were given the option of replaying an utterance as many times as desired before responding.

5. Speech naturalness

Although the primary focus of this study was speech intelligibility, a preliminary experiment was conducted involving judgments of speech naturalness. Our informal impression was that speech synthesized with the SES method was more natural sounding than that produced with the DSS. This would be consistent with the underlying premise that the SES method preserves more of the fine spectral detail of the original signal than the DSS method. Consequently, an experiment was run to formalize our subjective impressions about these differences in naturalness. Using the 50 TIMIT sentences, listeners heard two utterances on each trial. The naturally spoken sentence was played first, followed by either the SES or DSS version of the same sentence, determined randomly, with a probability of 0.5, and scrambled separately for each listener. Listeners were told that the first sentence was a recording of an original, naturally spoken utterance, while the second was a version of the same sentence that had been created artificially by a computer speech synthesizer. Listeners were asked to judge the naturalness of the synthesized sentence in relation to the original sentence. They were told that the original sentence had a naturalness value of 100 by definition, so if the synthesized utterance sounded half as natural as the original they were to assign a value of 50, if it sounded 1/10th as natural, they were to assigned it a value of 10, etc.

III. RESULTS

A. Naturalness

Average naturalness ratings were significantly higher for the SES signals (81.8) than for the DSS signals (59.4; $t=3.78$, $df=11$, $p<0.05$). These findings confirm an important assumption underlying the study; namely, that the SES method does a better job of preserving the detailed shape of the time-varying spectrum than does the peaks-only DSS method. It is clear, then, that preserving spectral shape detail plays a significant role in controlling the overall quality of the synthesized utterance. The key question is whether this advantage for the SES in the preservation of spectral shape detail is accompanied by a corresponding advantage in the transmission of *phonetic* information.

B. Sentence intelligibility

Intelligibility for the HINT sentences, measured as the percentage of content words correctly repeated by the listener (following Nilsson *et al.*, 1994), was slightly higher for the SES (99.8%) signals than the DSS signals (98.7%). However, variability across listeners was quite low and each of the 18 individual listeners showed higher intelligibility scores for the SES signals. Consequently, despite the small difference between the means, the difference in intelligibility for the two synthesis conditions (using arcsine transformed percent correct values) was highly significant ($t=7.6$, $df=17$, $p<0.001$). Listeners rarely asked for either version of the HINT sentences to be replayed. However, they were far more likely to request a repetition of the DSS version (median=7 per session) than the SES version (median=1). A Wilcoxon signed-rank test on these non-normally distributed data showed that this difference was highly significant ($z=3.5$, $p<0.01$).

Results for the TIMIT sentences were similar. Intelligibility was slightly but significantly higher for the SES (96.9%) signals than the DSS signals (95.7%; $t=1.7$, $df=18$, $p<0.05$). Further, listeners requested significantly more repetitions of the DSS versions (median=7) than the SES versions (median=5; $z=3.2$, $p<0.01$).

C. Vowel intelligibility

Vowel intelligibility results for the natural signals and for the three types of synthesized signals (SES, DSS, and formant synthesis), averaged across the 13 listeners, are shown in Table I. It can be seen that vowel intelligibility was nearly identical for the natural and SES signals, but there was a drop in intelligibility of some 5–6 % for either the DSS or formant-synthesized signals. The DSS and formant-synthesized signals produced very similar recognition rates.

TABLE I. Vowel intelligibility for four different versions of the 300 /hVd/ utterances from Hillenbrand *et al.* (1995): (1) naturally spoken (NAT), (2) synthesized with the spectral envelope synthesizer (SES), (3) synthesized with the damped-sine wave synthesizer (DSS), and (4) synthesized with the Klatt and Klatt (1990) formant synthesizer (Klsyn).

NAT	SES	DSS	Klsyn
Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)
95.2 (2.9)	95.0 (2.6)	89.3 (5.0)	90.0 (3.1)

A two-way repeated measures analysis of variance (ANOVA) computed on arcsine transformed percent correct scores showed highly significant effects for Stimulus Type (natural/SES/DSS/formant synthesis) [$F(3,51)=56.3$, $p<0.001$] and Vowel Category [$F(11,187)=6.2$, $p<0.001$] as well as a significant interaction [$F(33,561)=7.7$, $p<0.001$]. Bonferroni *post hoc* tests showed: (1) no difference between the natural and SES signals; (2) no difference between DSS and formant-synthesized signals; and (3) differences between both the natural and SES signals versus the DSS and formant-synthesized signals. The main effect for Vowel is unsurprising in light of the many studies show that some vowels are more readily identified than others (e.g., Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). As can be seen in Fig. 7, the effect of stimulus type was not uniform across vowels, resulting in the Stimulus-Type-by-Vowel interaction. Analysis and discussion of variations across vowels for natural versus formant-synthesized vowels can be found in Hillenbrand and Nearey (1999). Other aspects of the interaction patterns shown in Fig. 7 are rather complex and no simple explanation seems apparent to us.

The full confusion matrices for each of the four conditions are shown in Tables A1–A4. The main point worth noting in these tables is that the errors in all four matrices are unremarkable, showing the usual confusions that are seen among adjacent vowels. This is hardly surprising for the naturally spoken vowels, or for the SES vowels given that the goal was to preserve the detailed spectral shape of the original signal as closely as possible. Equally unsurprising is the well-behaved confusion matrix for the formant synthesized signals since these were generated from hand-edited formant contours. It is worth noting, however, that the confusion matrix for the DSS signals is also dominated by confusions among adjacent vowels, despite the fact that these signals were generated from unedited spectral peaks rather than hand-edited formants.

The main conclusion to be drawn from the vowel intelligibility tests is that the SES method, which preserves the fine details of the original spectrum, conveys vowel identity almost perfectly while the peaks-only methods (DSS and formant synthesis) result in a modest (5%–6%) but significant loss of information conveying vowel identity. A very similar drop in vowel intelligibility was reported by Hillenbrand and Nearey (1999) in a comparison of naturally spoken (95.4%) and formant synthesized vowels (88.5%). Also significant, and closely related to the point above regarding the general look of the confusion matrices, is the fact that the formant-synthesized signals, which were driven by carefully hand-edited formant tracks, were no more intelligible than the DSS signals, whose spectrum envelopes were created from unedited spectral peaks which sometimes corresponded to formants and sometimes did not. We will return to this point in Sec. IV.

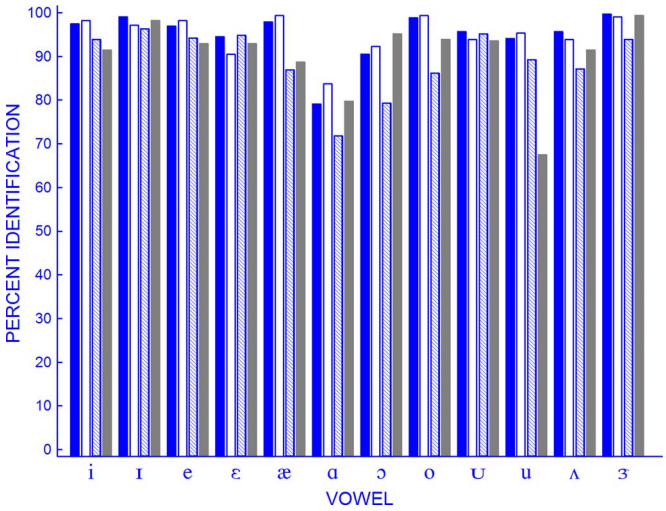


FIG. 7. (Color online) Percent identification as function of vowel category and stimulus type. Going from left to right in each group of bars, the conditions are naturally spoken signals (filled), spectral envelope synthesis (open), damped sine wave synthesis (hatched), and formant synthesis (shaded).

D. Consonant intelligibility

Consonant recognition rates, averaged across all subjects and all consonants, were 97.4% ($sd=1.7$) for the naturally spoken signals, 88.9% ($sd=3.0$) for the SES signals, and 85.9% ($sd=3.3$) for the DSS signals. Although our primary purpose was to compare the fine-detail-preserving SES representation with the peaks-only DSS method, it is clear that the largest effect that was observed was the drop in intelligibility for either of the synthesized signals as compared with the naturally spoken utterances. A two-way repeated measures ANOVA (using arcsine transformed percent correct scores) showed highly significant effects for Stimulus Type [$F(2,48)=414.1$, $p<0.001$] and Consonant [$F(22,528)=36.2$, $p<0.001$] as well as a significant interaction [$F(44,1056)=15.9$, $p<0.001$]. Bonferroni *post hoc* tests showed statistically reliable differences among all pair-wise comparisons of Stimulus Types. It is noteworthy, in our view, that the SES-DSS difference, while statistically significant, is a relatively small effect in absolute terms, amounting to an average of just seven additional signals correctly identified (from the total of 239 signals) under the SES condition versus DSS. This effect is also substantially smaller than the ~10 percentage point difference separating the naturally spoken signals from the SES and DSS signals.

The interaction patterns, which are rather complex, are summarized in Table II, which lists the consonants that showed decrements in intelligibility of 10% or greater for three different comparisons. (Full confusion matrices are given in Tables B1–B3). The first group of rows compares identification rates for naturally spoken signals with SES versions of the same signals, the middle group of rows compares naturally spoken signals with DSS versions, and the

TABLE II. Listing of the consonants that showed the largest decrements in intelligibility when comparing: (a) naturally spoken utterances with their spectral envelope synthesis counterparts (the first group of rows); (b) naturally spoken utterances with their damped sinewave synthesis counterparts (the second group of rows); and (c) the spectral envelope signals with their damped sinewave synthesis counterparts (the last group of rows). The last column lists the speech sounds with which the target sound was most commonly confused, with the percentage of such confusions given in parentheses, rounded to the nearest integer.

Speech sound	Natural	SES	Difference	Confused primarily with
b	96.7	41.2	55.5	v(50)
d	100.0	80.7	19.3	ð(15)
k	100.0	81.0	19.0	p(14)
θ	93.1	77.8	15.3	f(14), h(3)
tʃ	97.2	83.1	14.1	ʃ(16)
Speech sound	**Natural**	**DSS**	**Difference**	**Confused primarily with**
f	91.4	62.1	29.3	θ(36)
ʒ	96.7	73.9	22.8	dʒ (8), ð(7)
b	96.7	74.0	22.7	v(13) ð(5) d(2)
v	98.5	77.8	20.7	ð(10), l(5), m(5)
d	100.0	80.5	19.5	ð(13)
tʃ	97.2	78.5	18.7	ʃ(16), t(3)
r	98.0	80.9	17.1	b(7), w(4), m(4), l(3)
k	100.0	84.5	15.5	p(10)
dʒ	94.2	79.2	15.0	tʃ(7), t(5)
ʃ	98.7	84.4	14.3	s(6), tʃ(6)
g	96.0	83.9	12.1	d(7), k(3)
h	94.8	83.7	11.1	p(11)
Speech sound	**SES**	**DSS**	**Difference**	**Confused primarily with**
f	91.5	62.1	29.4	θ(36)
v	99.0	77.8	21.2	ð(10), l(5), m(5)
ʒ	88.7	73.9	14.8	ð(7), j (5), z(3)
ʃ	97.3	84.4	12.9	s(6), tʃ(6)
r	91.6	80.9	10.7	b(7), w(4), m(4)

bottom group of rows compares SES and DSS signals. For the natural versus SES comparisons, all of the sounds showing large decrements in intelligibility (/b,d,k,θ,tʃ/) are cued primarily by rapid changes in the spectrum.[5] The single most striking feature of this table is the 55.5% decrement in the intelligibility of /b/, which was heard almost exclusively as /v/ when synthesized by the SES. The most plausible explanation for the strong tendency of SES versions of /b/ to be heard as /v/ is that the 32 ms frequency analysis window that is used as the first step in the creation of the spectrum envelope smears over a time interval that is too long to preserve the rapid increase in energy (or energy in a specific band) which probably characterizes /b/ but not /v/.[6] Inadequate time resolution is also consistent with the fact that all of the sounds in the top group of rows in Table II are characterized by rapid spectral changes. The middle group of rows, comparing the naturally spoken signals with their DSS counterparts, is also dominated by sounds whose recognition is known to depend on rapid spectral changes. There is clearly more involved than temporal resolution, however, since there are a few sounds showing sizeable drops in intelligibility that are not typically thought of in terms of rapid spectral movement (e.g., /ʒ/, /r/, /ʃ/, /h/) as well as many rapid-spectral-change sounds that do not appear on either the natural/SES or natural/DSS lists. These details aside, in relation to the primarily experimental goals of the study the most important finding from the consonant recognition tests is that the fine-detail-preserving SES method conveys only slightly more information about consonant identity than the coarse peaks-only DSS method.

IV. DISCUSSION

The primary goal of this study was to determine the effects on speech intelligibility of reconstructing speech using a purposely coarse method of estimating the spectrum envelope based entirely on the distribution of spectral peaks (DSS) versus a method that attempts to preserve the fine details of the envelope shape (SES). Taken in the broadest possible terms the results showed that nearly as much speech information is conveyed by the peaks-only DSS method as by the detail-preserving SES method. Just as clearly, however, every test showed some measurable advantage for spectral detail, although the differences were not large in absolute terms.

Sentence intelligibility for both the TIMIT and HINT signals was slightly (~1 percentage point) but reliably higher for the SES condition, and listeners requested significantly more repetitions of the DSS versions. Sentences from both databases were presented in quiet under good listening conditions. It is quite possible that the intelligibility advantage for the SES sentences would be greater under adverse listening conditions. In general, though, the fact that sentence intelligibility scores were so similar for the two methods tends to provide some support for cochlear implant processor strategies such as MPEAK (Multipeak) and SPEAK (Spectral Peak) that rely on transmitting primarily the high energy components of the spectrum. The results also suggest that accurate formant tracking is not necessary, at least for quiet signals, since the DSS is driven by unedited spectral peaks rather than edited formants. There is, however, evidence indicating that MPEAK, which relies on peak picking, does not perform as well in noise as SPEAK which, despite the name, transmits information about the highest amplitude spectral components, independent of whether the component constitutes a spectral peak (Skinner, Holden, Holden, Dowell, Seligman, Brimacombe, and Beiter, 1991).

The major findings from the vowel intelligibility tests are (1) there was no difference in vowel intelligibility between the SES signals and the naturally spoken signals, (2) there was an intelligibility drop of some 5–6 percentage points for both of peaks-only methods (DSS and formant synthesis), and (3) there was no difference in vowel intelligibility between the DSS and formant synthesis conditions. Taken together we believe these findings argue for an underlying pattern matching mechanism that is very much in line with Klatt's (1982) interpretation of his findings on phonetic distance judgments with static vowels. While Klatt's results showed that formant frequencies were by far the most important determinant of vowel quality, he argued that his findings were best explained not by an underlying formant tracking mechanism but rather by a spectral shape pattern matching process which, "… attends to the locations of prominent energy concentrations, but does not attend greatly to their relative intensities, nor to the shape of the spectrum in the valleys between energy concentrations" (p. 1281). Three aspects of the present vowel intelligibility results are consistent with this interpretation. First, a modest but measurable loss of phonetically relevant information occurs when the original spectral shape is reduced to either formants or unedited spectral peaks, indicating that formants/spectral peaks alone are not entirely sufficient to specify vowel identity [see Bladon's (1982), *reduction* argument in his critique of formant representations]. Second, formants *per se* do not appear to be required since vowel identity was conveyed just as well by unedited spectral peaks as by hand-edited formant frequencies. Third, the confusion matrix for the DSS vowels remains quite well behaved, showing almost exclusively listening errors involving vowels that are phonetically quite similar to the vowel intended by the talker. Given that the DSS is driven by raw spectral peaks, this suggests to us that this synthesis method conveys vowel identity as well as the formant synthesizer not by implicitly conveying formant information but rather by doing a reasonable enough job of preserving the shape of the spectrum in high energy regions.

There remains considerable uncertainty about the precise nature of the peak-dominated spectral shape pattern matching scheme that might meet Klatt's (1982) requirements. Klatt proposed a *weighted spectral slope* metric (WSM) based on the comparison of spectral slopes, but with greater weight given to slope differences in and around spectral peaks. The metric provided good predictions of perceived phonetic distances among static synthetic vowels. Klatt's WSM metric was used by Nocerino *et al.* (1985) in a study of automatic recognition of naturally spoken words (alphabet and digits). Surprisingly, WSM was found to perform best when the peak-emphasizing weighting factors were completely removed, a finding that runs counter to the notion that spectral shape differences in the vicinity of peaks are of greater importance. A similar result was reported in a study of concurrent vowel recognition by Assmann and Summerfield (1989), where the WSM showed only a small increase in recognition performance over a non-peak-weighted version of the same metric.

Hillenbrand and Houde (2003) proposed a *narrow band pattern matching model* of vowel perception based on city block spectral distances between narrow band input spectra and smoothed spectral shape templates derived empirically by averaging the harmonic spectra of like vowels spoken by a panel of talkers. The information-bearing spectral peaks were enhanced by a "thresholding" procedure that zeroes out spectral values below a threshold function consisting of a center-weighted running average of spectral amplitudes. The pattern matching model recognized vowels from a 1668-token database consisting of 12 vowels in /hVd/ context spoken by men, women, and children. Model accuracy approached but did not quite reach that of human listeners. Of special relevance to the present findings, the effect of the peak-enhancing operation was dramatic, with overall recognition accuracy falling from 91.4% with the thresholding operation all the way to 59.9% without it. Similarly, Liénard and Di Benedetto (2000) developed a peak-enhancement method to recognize French vowels from "bump vectors"—smooth spectra resulting from an independently developed peak-enhancing operation that is quite similar to the one used in our pattern matching model. Recognition experiments showed a substantial advantage for the bump vector over a variety of alternative smoothed spectral shape representations that did not incorporate a peak-enhancement operation.

The pattern of results for the consonant intelligibility tests differed in some important ways from the vowel results. There was a small but statistically reliable advantage of about 3 percentage points for the detail-preserving SES method as compared to the peaks-only DSS method. In numerical terms, however, the SES-DSS difference for consonants was only about half of that for vowels. Consonant spectra tend to be rather ragged and irregular in shape, and it appears that in most but not quite all cases the more coarse DSS method conveys the relevant phonetic information as well as the SES method. However, unlike the vowel tests, which showed that the SES signals were just as intelligible as the naturally spoken signals, the SES versions of the consonants were, on average, 8.5 percentage points less intelli-

gible than their naturally spoken counterparts. Especially large decrements in intelligibility tended to be seen for consonants that are cued by rapid changes in the spectrum, suggesting that inadequate temporal resolution was at least in part responsible for the intelligibility deficit. Follow up experiments are underway to explore this possibility.

Appendix A

TABLE A1. Confusion matrix for naturally spoken /hVd/ utterances (iy=/i/, ih=/ɪ/, ei=/e/, eh=/ɛ/, ae=/æ/, ah=/ɑ/, aw=/ɔ/, oa=/o/, oo=/ʊ/, uw=/u/, uh=/ʌ/, er=/ɝ/).

	iy	ih	ei	eh	ae	ah	aw	oa	oo	uw	uh	er
iy	97.5	0.9	0.9				0.3					0.3
ih	0.3	99.1		0.3					0.3			
ei		0.3	96.9	0.6	1.8					0.3		
eh				94.5	4.3		0.3			0.3	0.6	
ae				1.8	97.9	0.3						
ah				0.3	15.7	79.1	4.3		0.3		0.3	
aw	0.3				0.3	7.4	90.5	0.3			1.2	
oa				0.3			0.6	98.8	0.3			
oo			0.3				0.3		95.7	0.6	2.5	0.6
uw								2.2	3.7	94.1		
uh		0.3				1.2	0.6	0.3	1.5		95.7	0.3
er										0.3		99.7

TABLE A2. Confusion matrix for utterances generated with the spectral envelope synthesizer (iy=/i/, ih=/ɪ/, ei=/e/, eh=/ɛ/, ae=/æ/, ah=/ɑ/, aw=/ɔ/, oa=/o/, oo=/ʊ/, uw=/u/, uh=/ʌ/, er=/ɝ/).

	iy	ih	ei	eh	ae	ah	aw	oa	oo	uw	uh	er
iy	98.2		1.5	0.3								
ih	0.9	97.2		1.5				0.3				
ei		0.3	98.2	0.6		0.6				0.3		
eh			0.9	90.5	8.6							
ae				0.6	99.4							
ah			0.3		14.7	83.7	0.6	0.3	0.3			
aw						6.2	92.3				1.2	0.3
oa								99.4		0.6		
oo		0.3					0.3	0.3	93.9	0.3	4.0	0.9
uw								1.5	2.5	95.4	0.3	0.3
uh					0.3	1.5	1.2	0.3	1.8	0.6	93.9	0.3
er	0.3	0.3									0.3	99.1

TABLE A3. Confusion matrix for utterances generated with the damped sine wave synthesizer (iy=/i/, ih=/ɪ/, ei=/e/, eh=/ɛ/, ae=/æ/, ah=/ɑ/, aw=/ɔ/, oa=/o/, oo=/ʊ/, uw=/u/, uh=/ʌ/, er=/ɝ/).

	iy	ih	ei	eh	ae	ah	aw	oa	oo	uw	uh	er
iy	93.9	3.7	2.1	0.3								
ih	1.5	96.6		1.8								
ei	5.2	0.6	94.2									
eh			0.6	94.8	4.6							
ae		0.3	0.3	12.2	86.9					0.3		
ah					19.0	71.8	3.4	0.6			4.9	0.3
aw					0.3	15.5	79.3	0.6	0.6		3.7	
oa							0.3	86.2	3.4	5.2	4.9	
oo	0.3								95.1	2.5	1.8	0.3
uw	0.3							5.2	4.6	89.2	0.6	
uh				0.6		0.3	0.6		10.8		87.1	0.6
er				1.5	0.3	0.3		0.3	2.8	0.6	0.3	93.9

TABLE A4. Confusion matrix for utterances generated with the Klatt and Klatt (1990) formant synthesizer (iy=/i/, ih=/ɪ/, ei=/e/, eh=/ɛ/, ae=/æ/, ah =/ɑ/, aw=/ɔ/, oa=/o/, oo=/ʊ/, uw=/u/, uh=/ʌ/, er=/ɝ/).

	iy	ih	ei	eh	ae	ah	aw	oa	oo	uw	uh	er
iy	91.4	3.7	4.6			0.3						
ih	0.6	98.2		0.9		0.3						
ei	3.7	1.9	92.9	0.3	0.9	0.3						
eh		0.9		92.9	6.1							
ae		0.3		11.0	88.7							
ah				0.6	13.5	79.7	3.7	0.3			1.8	0.3
aw	0.3					4.3	95.1		0.3			
oa						0.6	0.6	93.9	2.5	1.5	0.6	0.3
oo			0.3	0.6					93.6	0.9	4.3	0.3
uw							0.3	18.1	14.1	67.5		
uh		0.3				0.9	1.5		5.8		91.4	
er	0.3								0.3			99.4

Appendix B

TABLE B1. Confusion matrix for the consonant intelligibility tests using naturally spoken utterances (th=/θ/, sh=/ʃ/, dh=/ð/, zh=/ʒ/, dz=/dʒ/,ch =/tʃ/). Percentages are rounded to the nearest integer.

	b	d	g	p	t	k	f	th	s	sh	h	v	dh	z	zh	dz	ch	m	n	l	r	w	j
b	97	1						1				1	1										
d		100																					
g		1	96			3																	
p				96		3																	
t					100																		
k						100																	
f							91	8	1														
th							3	93	2				2										
s								3	97														
sh									1	99							1						
h				2							95											3	
v												99	2										
dh		1						4				1	93										
z														99									
zh															97	3							
dz		1	5													94	1						
ch					1					2							97						
m																		100					
n																			99				
l	2																			98			
r	1		1																		98		
w												1										98	1
j																							100

TABLE B2. Confusion matrix for the consonant intelligibility tests using sentences generated with the spectral envelope synthesizer (th=/θ/, sh=/ʃ/, dh=/ð/, zh=/ʒ/, dz=/dʒ/, ch=/tʃ/). Percentages are rounded to the nearest integer.

	b	d	g	p	t	k	f	th	s	sh	h	v	dh	z	zh	dz	ch	m	n	l	r	w	j
b	41	1						1				50	5					2					
d		81	1					1					15			1							
g		2	87	7		2						1											1
p				93			1				6												
t					92											1	7						
k				14	1	81					3					1							
f							92	8	1			1											
th							9	78	2			1	10										
s								2	98														
sh									1	97					1		1						
h				6			1	1			89			1								3	
v								1				99	1										
dh								3				11	86										
z												1	6	94									
zh															89	7							4
dz			6												5	88	2						
ch										16							83						
m																		100					
n						1												3	96				
l	5												2							92			
r	8		1																		92		
w												4										96	
j	1																		1				98

TABLE B3. Confusion matrix for the consonant intelligibility tests using sentences generated with the damped sine wave synthesizer (th=/θ/, sh=/ʃ/, dh=/ð/, zh=/ʒ/, dz=/dʒ/, ch=/tʃ/). Percentages are rounded to the nearest integer.

	b	d	g	p	t	k	f	th	s	sh	h	v	dh	z	zh	dz	ch	m	n	l	r	w	j
b	74	2	3					1				13	5										1
d		81	2		1			1					13			1							
g		7	84	1		3		1				1	1										2
p				88		7					4							1					
t					97	1											1						
k				10	1	85					3												
f							62	36			1	1	1										
th							6	86	1		1		6										
s								1	98	1													
sh							1		6	84	1				1		6						
h				11		1		1			84			1								3	
v	4											78	10					4		5	1		
dh		1						3				2	91							3			
z													5	94									
zh			1										7	3	74	8				1			5
dz			2		5			1						1	5	79	7					1	
ch					3				1	16						1	79						
m	1											2						95		1			1
n																		7	92	1			
l	4			1									2					1		90			
r	7		1															4		3	81	4	
w												2						2				95	1
j			1																5	1			93

[1]The most important difference between the implementation used here and that described in Hillenbrand and Houde (2002) involves the FFT window size for envelope analysis, which was shortened from 64 to 32 ms, based on findings showing poor identification of some phonetic contrasts that depend on rapid spectral change. As will be discussed below, time resolution remains a problem even with the 32 ms window.

[2]Throughout the text, we will use the term *Gaussian-weighted running average* to refer to an approximation implemented with three passes of a rectangular (i.e., unweighted) running average. In a Gaussian weighted running average, each spectral amplitude is replaced by the weighted average of n neighbors of higher and lower frequency, with n being determined by the width of the smoothing window. Greater weight is assigned to spectral values at the center of the averaging window than to values nearer to the edge of the window. In a true Gaussian-weighted average, the distribution of weights follows a Gaussian function. A simple-to-implement, close approximation to a Gaussian-weighted average can be achieved by running three passes of a rectangular (i.e., unweighted) average: The output of an initial running average operation becomes the input to a second running average, whose output in turn becomes the input to a third running average. (See Hillenbrand and Houde (2003), for an explanation of the end-correction scheme that is used.).

[3]Without a smoothing operation of some kind, the peak frequencies that are measured from the harmonic envelope of a voiced speech segment would always correspond to the highest amplitude harmonic in the collection of harmonics comprising the peak. The evidence is quite clear that the phonetic quality associated with speech signals does not correspond to the strongest harmonic (Klatt, 1986). A peak frequency measured from the harmonic envelope after F_0-dependent smoothing corresponds closely with the method based on the weighted average of harmonic amplitudes used in studies such as Peterson and Barney (1952); i.e., for a peak composed of three harmonics, the peak frequency will be displaced from the central harmonic in the direction of the second strongest harmonic.

[4]What is referred to here as a "thresholding" operation is referred to in some of our earlier writings as a masking operation, a term which remains appropriate, in our view, but was dropped to avoid creating the impression that we were attempting a faithful simulation of simultaneous masking in the auditory system.

[5]The weak fricative /θ/, which was confused primarily with /f/, might seem out of place with the stops and the affricate which form the remainder of this list, but there is clear evidence that the /f/-/θ/ distinction is cued primarily by brief-duration formant transitions rather than differences in the spectral content of the quasistationary fricative noise (Harris, 1958).

[6]The word *probably* is used here because we are unaware of any published experimental work that has systematically examined the cues to distinctions between stops and similar-place weak fricatives such as /b/-/v/ and /d/-/ð/. Extensive pilot work of our own suggests that the voiced stops /b/ and /d/ typically show a more abrupt increase in high-frequency energy (e.g., above 1 kHz) than /v/ and /ð/. Further, because of the 32 ms analysis window size, this abrupt high-frequency onset that appears to characterize stops is often not observed in the SES versions of /b/ and /d/. Informal listening also suggests to us that SES versions of /b/ are rendered much more faithfully with short analysis window sizes such as 8 ms, although other features of the synthesis are severely compromised when this short window is used.

Assmann, P. F., and Summerfield, Q. (**1989**). "Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency," , **85**, 327–338.

Bench, J., Kowal, A., and Bamford, J. (**1979**). "The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children," Br. J. Audiol., **13**, 108–112.

Bladon, A. (**1982**). "Arguments against formants in the auditory representation of speech," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granstrom (Elsevier Biomedical Press, Amsterdam), pp. 95–102.

Bladon, A. and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am., **69**, 1414–1422.

Carlson, R., and Granstrom, B. (**1979**). "Model predictions of vowel dissimilarity," Speech Transmission Laboratory Quarterly Progress and Status Report No. STL-QPSR 3-4/1979 (Royal Institute of Technology, Stockholm, Sweden), pp. 84–104.

Carlson, R., Granstrom, B., and Klatt, D. H. (**1979**). "Vowel perception: The relative perceptual salience of selected acoustic manipulations," Speech Transmission Laboratory Quarterly Progress and Status Report No. STL-QPSR 3-4/1979 (Royal Institute of Technology, Stockholm, Sweden), pp. 73–83.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G.,Pallett, D. S., and Dahlgren, N. L. (**1993**). "DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM," Gaithersburg, MD, National Institute of Standards and Techonology.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech, **1**, pp. 1–7.

Hillenbrand, J. M., and Gayvert, R. T. (**2005**). "Open source software for experiment design and control," J. Speech Lang. Hear. Res., **48**, 45–60.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am., **97**, 1300–1313.

Hillenbrand, J. M., and Houde, R. A. (**1996**). "A method for creating filters with arbitrary response characteristics for use in hearing and speech research," J. Speech Hear. Res., **39**, 390–395.

Hillenbrand, J. M., and Houde, R. A. (**2002**). "Speech synthesis using damped sinusoids," J. Speech Lang. Hear. Res., **45**, 639–650.

Hillenbrand, J. M., and Houde, R. A. (**2003**). "A narrow band pattern-matching model of vowel perception," J. Acoust. Soc. Am., **113**, 1044–1055.

Hillenbrand, J. M., and Nearey, T. N. (**1999**). "Identification of resynthesized /hVd/ syllables: Effects of formant contour," J. Acoust. Soc. Am., **105**, 3509–3523.

Klatt, D. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am., **67**, 971–995.

Klatt, D. H. (**1982**). "Prediction of perceived phonetic distance from critical-band spectra: A first step," *IEEE ICASSP*, 1278–1281.

Klatt, D. H. (**1986**). "Representation of the first formant in speech recognition and in models of the auditory periphery," in *Proceedings of the Montreal Satellite Symposium on Speech Recognition*, edited by P. Mermelstein, McGill University, Montreal, pp. 5–7.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am., **87**, 820–857.

Liénard, J-S., and Di Benedetto, M-G. (**2000**). "Extracting vowel characteristics from smoothed spectra," J. Acoust. Soc. Am., **108**, Suppl. 1, 2602.

Lindblom, B. (**1978**). "Phonetic aspects of linguistic explanation," Studia Linguistica, **32**, 137–153.

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.

Nocerino, N., Soong, F. K., Rabiner, L. R. and Klatt, D. H., (**1985**). "Comparative study of several distortion measures for speech recognition," Speech Commun., **4**, 317–331.

Paul, D. B. (**1981**). "The spectral envelope estimation vocoder," IEEE Trans. Acoust., Speech, Signal Process., **29**, 786–794.

Peterson, G., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am., **24**, 175–184.

Shannon, R. V., Jensvold, A. Padilla, M., Robert, M., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am., **106**, 71–74.

Skinner, M., Holden, L., Holden, T., Dowell, R., Seligman, P., Brimacombe, J., and Beiter, A. (**1991**). "Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant," Ear Hear., **12**, 3–22.

Zahorian, S. A., and Jagharghi, A. J. (**1993**). "Spectral shape versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am., **94**, 1966–1982.

Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments[a)]

Nao Hodoshima[b)] and Takayuki Arai
Department of Electrical and Electronics Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554, Japan

Akiko Kusumoto
Department of Electrical and Electronics Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
and Center for Spoken Language Understanding, OGI School of Science and Engineering, Oregon Health and Science University, 20000 NW Walker Road, Beaverton, Oregon 97006

Keisuke Kinoshita[c)]
Department of Electrical and Electronics Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan

(Received 25 June 2004; revised 13 February 2006; accepted 13 March 2006)

Overlap-masking degrades speech intelligibility in reverberation [R. H. Bolt and A. D. MacDonald, J. Acoust. Soc. Am. **21**(6), 577–580 (1949)]. To reduce the effect of this degradation, steady-state suppression has been proposed as a preprocessing technique [Arai *et al.*, Proc. Autumn Meet. Acoust. Soc. Jpn., 2001; Acoust. Sci. Tech. **23**(8), 229–232 (2002)]. This technique automatically suppresses steady-state portions of speech that have more energy but are less crucial for speech perception. The present paper explores the effect of steady-state suppression on syllable identification preceded by /a/ under various reverberant conditions. In each of two perception experiments, stimuli were presented to 22 subjects with normal hearing. The stimuli consisted of mono-syllables in a carrier phrase with and without steady-state suppression and were presented under different reverberant conditions using artificial impulse responses. The results indicate that steady-state suppression statistically improves consonant identification for reverberation times of 0.7 to 1.2 s. Analysis of confusion matrices shows that identification of voiced consonants, stop and nasal consonants, and bilabial, alveolar, and velar consonants were especially improved by steady-state suppression. The steady-state suppression is demonstrated to be an effective preprocessing method for improving syllable identification by reducing the effect of overlap-masking under specific reverberant conditions. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2198191]

PACS number(s): 43.71.Es, 43.66.Dc, 43.55.Hy [KWG] Pages: 4055–4064

I. INTRODUCTION

Understanding speech in a large auditorium is difficult. One reason for this difficulty is reverberation, which is the superposition of reflected sounds having various delays and amplitudes. Although reverberation adds richness to music, longer reverberation makes speech more difficult to understand. Since the optimal reverberation time for speech is shorter than that for music, speech intelligibility in multiple-purpose auditoria, which are used for different types of events that incorporate both speech and music, is sometimes poor (Knudsen, 1929). People with impaired hearing (Nábělek and Pickett, 1974; Finitzo-Hieber and Tillman 1978), children, elderly people (Elliot, 1982; Nábělek and Robinson, 1982; Crandell and Smaldino, 2000), and non-native listeners (Nábělek and Donahue, 1984; Takata and Nábělek, 1990) have been reported to have more difficulty understanding speech in reverberation than people with normal hearing.

There are two general approaches to improving speech intelligibility in reverberant environments: postprocessing techniques and preprocessing techniques. Postprocessing methods are applied to speech signals that are captured in reverberant environments. Microphone arrays (Flanagan *et al.*, 1985), inverse filtering (Miyoshi and Kaneda, 1998), and modulation filtering (Langhans and Strube, 1982; Avendano and Hermansky, 1996) are examples of postprocessing methods. Microphone arrays take advantage of the spatial information related to a sound source. Modulation filtering alters the modulation spectrum of a signal, which is derived from frequency analysis of the temporal envelope of the band

[a)]Portions of this work were presented in "Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments," Proc. China-Japan Joint Conference on Acoustics, Nanjing, November 2002, and "Improving speech intelligibility by steady-state suppression as pre-processing in small to medium sized halls," Proceedings of Eurospeech, Geneva, September 2003.
[b)]Electronic mail: n-hodosh@sophia.ac.jp
[c)]Present address: NTT Communication Science Laboratories, 2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan.

passed signal. Postprocessing methods are widely applied in fields such as automatic speech recognition and teleconferencing (Kinoshita *et al.*, 2005).

Preprocessing approaches process speech signals before they are affected by reverberation. This approach reduces the influence of reverberation on the transmission path. Langhans and Strube (1982) applied the same modulation filtering in preprocessing as they had used in postprocessing, but intelligibility of the processed signals was lower than that of the unprocessed signals. Kitamura *et al.* (2000) applied a data-derived IMTF filter (a modulation filtering method) calculated from stimuli that were used in their perceptual tests. The results showed that for a reverberation time (RT) of 2.1 s the filter improved phoneme identification scores by 4.8% for people who have sensorineural hearing losses and improved the identification scores by 5.0% for words, 6.2% for morae, and 5.5% for phonemes for people with normal hearing.

Using a modulation filter, Kusumoto *et al.* (2000) enhanced a particular modulation frequency region (around 4 Hz) that had been reported as being important for speech perception. Four subjects who have sensorineural hearing losses rated reverberant processed speech signals as being easier to hear than reverberant unprocessed speech signals for an RT of approximately 3 s. Kusumoto *et al.* (2005) evaluated the effect of the two modulation filters that Kusumoto *et al.* (2000) and Kitamuta *et al.* (2000) used with native speakers of English with normal hearing. The results of a modified rhyme test showed that both filters improved word-final consonant discrimination scores by approximately 6% for an RT of 1.1 s. These studies reported above show modulation filtering, as a preprocessing approach, improved speech intelligibility in reverberant environments.

Overlap-masking and self-masking are contributors to the reduction of speech intelligibility under reverberant conditions (Bolt and MacDonald, 1949; Nábělek and Robinette, 1978; Nábělek *et al.*, 1989). Overlap-masking occurs when reverberation tails of previous portions of a sound affect subsequent segments, whereas self-masking is internal, temporal smearing within a phoneme. Knudsen (1929) conducted recognition tests with consonant-vowel-consonant (CVC) syllables in a reverberant environment and reported a greater degradation of speech intelligibility in final consonants than in initial consonants. He showed that the preponderance of errors among final consonants was the result of overlap-masking produced by the reverberation of the preceding vowel. Furthermore, Bolt and MacDonald (1949) reported that self-masking was relatively unimportant since intelligibility could be greatly increased by speaking slowly. Therefore, it can be surmised that overlap-masking contributes more than self-masking to the degradation of speech intelligibility by reverberation.

To decrease the effect of overlap-masking, a first attempt might be to decrease the energy of all preceding phonemes so that the energy of the reverberation components overlapping to subsequent portions is attenuated. The problem is that reducing the energy of all components might degrade the speech intelligibility. Therefore, consideration of which components should be reduced appears to be necessary in order to minimize the effect of overlap-masking without degrading speech intelligibility. This approach is applied to the steady-state suppression described below.

Spectral transitions of speech signals are important features for speech perception. Strange *et al.* (1983) reported that dynamic spectral information was sufficient for accurate identification of vowels even when the vowel nuclei were attenuated to zero. They showed that the information in steady-state portions of a speech signal was relatively insignificant compared with that in transient segments. Furui (1986) examined the relationship between dynamic spectral features and the identification of syllables. The identification tests using syllables of which the initial and/or final portions were truncated at various positions showed that spectral transitions are crucial for syllable and vowel perception, whereas vowel nuclei are not necessary for either vowel or syllable perception. Both "delta" processing of cepstral features (Furui, 1986) and RelAtive SpecTrAl (RASTA) processing (Hermansky and Morgan, 1994) enhance transitions of speech signals and contribute to increased recognition rates in automatic speech recognition.

A number of studies have reported that enhancing consonant energies in noise and reverberation improve speech intelligibility. Nábělek *et al.* (1989) predicted that a signal modification that enhances the intensity of consonants should be beneficial in reverberant environments. That is, if the intensity of a high-energy vowel is reduced or if the intensity of the following consonant is increased, the effect of overlap-masking should be reduced. Clearly articulated speech is characterized by relatively more intense consonants than conversational speech. Payton *et al.* (1994) showed clear speech to be less degraded by reverberation and noise than conversational speech, both for young people with normal hearing and for people who have sensorineural hearing losses. Gordon-Salant (1986) reported that increasing the energy of a consonant relative to that of a vowel improved consonant recognition in noise, both for young people and for elderly people aged 65 to 72 years with normal hearing. On the other hand, Uchanski *et al.* (2002) reported that when the envelope of a speech signal is modified to amplify fast elements of the speech in noise, speech perception was degraded for both children with normal hearing and children who have sensorineural hearing losses. They do not report any perceptual experiments in reverberation to test the reduction of overlap-masking.

Arai *et al.* (2001, 2002) proposed steady-state suppression to reduce the overlap-masking caused by the reverberation components of previous portions of speech. They accomplished this by automatically suppressing steady-state portions of speech that had more energy but were less crucial for speech perception. This method reduced the effect of overlap-masking without degrading speech intelligibility. They demonstrated improved consonant identification for an RT of 1.1 s. In conventional modulation filtering methods, temporal envelopes are filtered linearly. In contrast, steady-state suppression consists of nonlinear processing of temporal envelopes.

The ultimate goal of the present research is to provide a preprocessing filter that is suitable for a given auditorium

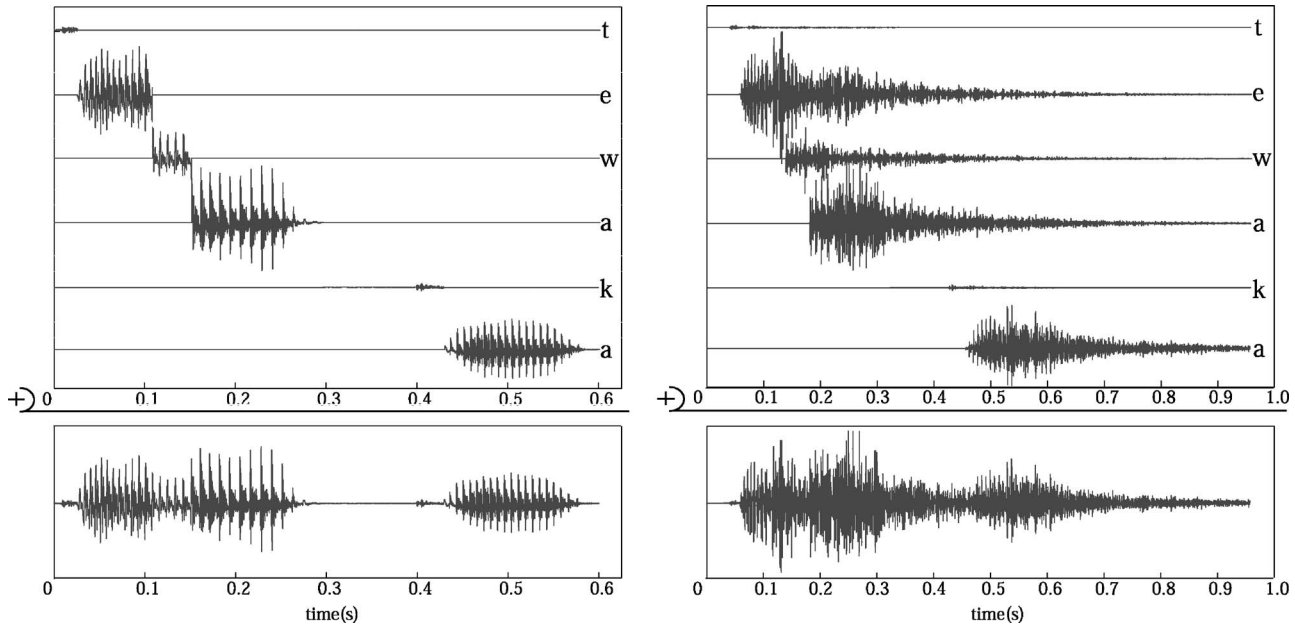


FIG. 1. An illustration of overlap-masking on the segments "tewaka." (a) The waveforms on the left are the original signals. The left bottom panel shows the original whole utterance, while the top left panel shows the waveforms for the phonemes /t/, /e/, /w/, /a/, /k/, and /a/. (b) The reverberant signals shown on the right correspond to the original signals on the left. The reverberant signals are obtained by taking the convolution of the original signals with the impulse response of the room with an RT of 1.1 s. This figure shows that when a previous segment has strong energy, as in the case of a vowel, the maskee of the reverberant following segment is smeared to a much greater degree by the masker of the reverberant previous segment.

that has distinct acoustic properties. In order to achieve this, it is necessary to investigate the effects of both the preprocessing filter and RT on speech intelligibility. It would be difficult to examine both aspects simultaneously because it is not clear which aspect contributes more to speech intelligibility. Therefore, syllable identification tests were conducted under various RTs while applying a single preprocessing filter (modulation filter) (Hodoshima *et al.*, 2002). The results showed that the effect of modulation filtering varied with RT and most notably that modulation filtering prevented the degradation of syllable identification under specific conditions.

The purpose of the present paper was to explore the effect of steady-state suppression (a preprocessing method) on syllable identification preceded by /a/, under various reverberant conditions. Since the effects of steady-state suppression have only been examined for an RT of 1.1 s (Arai *et al.*, 2001, 2002), various reverberant conditions (RT of 0.4–1.3 s) were prepared according to Hodoshima *et al.* (2002) to evaluate the effect of steady-state suppression using the approach described by Arai *et al.* (2001, 2002). By fixing the suppression parameters of steady-state portions of speech, two perceptual tests were conducted with a set of artificial impulse responses having different RTs. The effects of steady-state suppression were examined for RTs of 0.9–1.3 s in experiment I. In order to investigate the effects of steady-state suppression under shorter reverberant conditions than those in experiment I, experiment II was conducted for RTs of 0.4–1.0 s.

II. STEADY-STATE SUPPRESSION

A. Effect of overlap-masking

Overlap-masking is the main reason reverberation degrades speech intelligibility (Knudsen, 1929; Bolt and MacDonald, 1949; Nábělek and Robinette, 1978; Nábělek *et al.*, 1989). Figure 1 shows overlap-masking on the Japanese segment /te wa ka/. The left part of Fig. 1 shows the original speech signals, with no reverberation. The bottom left panel shows the original utterance, while the top left panel shows the individual waveforms for the phoneme /t/, /e/, /w/, /a/, /k/, and /a/, which are manually segmented. The right part of Fig. 1 shows the reverberant speech signals corresponding to the original signals in the left part of this figure. The reverberant speech segments are obtained by taking the convolution of the original speech segments with the impulse response of the room (RT of 1.1 s). The entire waveform of this utterance (in the bottom panel) is the summation of the top five waveforms, each of which corresponds to a constituent. The right part of Fig. 1 shows that consonants that have weak energies, such as /k/, are masked by the reverberation tails of the previous vowels. This indicates that when a previous segment has strong energy (as in the case of a vowel) and the environment is reverberant, the following segments can be significantly smeared.

B. Signal processing

The steady-state suppression used in the present paper was proposed by Arai *et al.* (2001, 2002) to suppress the steady-state portions of speech signals. In this method, steady-state portions are detected and suppressed automatically. Figure 2 shows a block diagram of the steady-state suppression. The top panel of Fig. 2 shows the general flow of signals, while the bottom panel shows the computation of the D parameter which represents spectral transitions. The steady-state suppression defines a speech portion as steady state when D is less than a certain threshold, where D is essentially identical to a parameter proposed by Furui

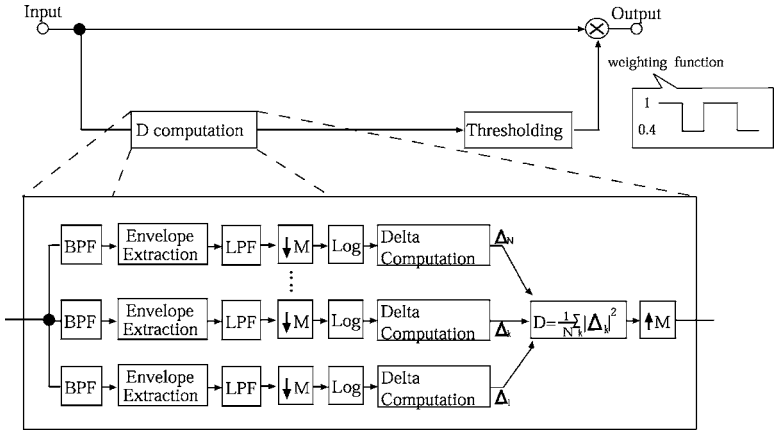


FIG. 2. Block diagram of the steady-state suppression. The top panel shows the general flow of signals. The steady-state suppression defines a speech portion as steady state when D is less than a certain threshold. Once a portion is considered to be steady state, the amplitude of the portion is multiplied by the factor 0.4. The bottom panel shows the D computation. First, the original signal is split into $\frac{1}{3}$ octave bands. In each band, the envelope is extracted and then smoothed by a low-pass filter. After down-sampling ($M=160$), the regression coefficients, *delta*, are calculated from the five adjacent samples of the time trajectory of the logarithmic envelope of each subband. The mean square of the regression coefficients over all bands, D, is then calculated. Finally, the time trajectory of D is up-sampled to the original sampling rate.

(1986). In the present paper, D is calculated as the mean square of the regression coefficients for each time trajectory of the logarithmic envelope of a subband, rather than that of the regression coefficients for each time trajectory of a cepstrum. Once a portion is considered to be steady state, the amplitude of the portion is multiplied by the factor 0.4 (the 40% suppression rate). This factor is used because Arai *et al.* (2001, 2002) showed that 40% was a reasonable suppression rate for an RT of 1.1 s.

The bottom part of Fig. 2 shows the D computation. First, an original signal, sampled at 16 kHz, is split into $\frac{1}{3}$ octave bands by bandpass filters (BPFs) that approximate critical bands. In each band, the envelope is extracted by the Hilbert transform and then smoothed by a low-pass filter (LPF) at a cutoff frequency of 35 Hz. After down-sampling ($M=160$), the regression coefficients, *delta*, are calculated from the five adjacent samples of the time trajectory of the logarithmic envelope of each subband. The mean square of the regression coefficients over all bands, D, is calculated every 10 ms. Finally, the time trajectory of D is up-sampled to the original sampling rate.

III. REVERBERATION CONDITIONS

The artificial impulse responses were synthesized as described by Hodoshima *et al.* (2002) to obtain the desired reverberant conditions. An impulse response h_o, which is measured in an enclosure, can be modeled as a product of an exponential decay with time constant τ_o and a stationary noise $w(t)$ (Schroeder, 1981). That is,

$$h_o(t) = e^{-t/\tau_o} w(t). \tag{1}$$

Then, a modified impulse response h_n with a new time constant τ_n is defined as

$$h_n(t) = e^{-t/\tau} h_o(t), \quad \text{where } \frac{1}{\tau} = \frac{1}{\tau_n} - \frac{1}{\tau_o}. \tag{2}$$

Thus, the desired RT can be obtained as a function of τ. The original impulse response h_o that is used in the present study was measured in Hamming Hall, Higashi-Yamato City, Tokyo. A reflection board was not used in the measurement. Table I lists the reverberant conditions used in these experiments (h6 in Table I is identical to h_o). The impulse responses h1–h5, h7, and h8 are obtained by adjusting τ in Eq. (2) to achieve the desired reverberant conditions.

Here, the RT is defined as the time for the decay curve of the impulse response to decrease 60 dB from the steady state. The early decay time (EDT), which is the time required for 10 dB of reverberation decay, was multiplied by 6 to estimate the RT. The values of the RTs in Table I are the average RTs derived from the EDT at the center frequencies of 0.5, 1, and 2 kHz of the 1-oct bandpassed impulse response.

IV. EXPERIMENT I

A. Stimuli

The original speech samples consisted of nonsense consonant-vowel (CV) syllables embedded in the Japanese carrier phrase, "Daimoku to shite wa _ to iimasu" (It is called _ as a title.). Table II shows the 24 CVs (targets) that were used in the experiment. All possible CV combinations were selected, excluding those that do not meet Japanese phonotactics. The original speech samples were obtained from the ATR speech database of Japanese. The speaker who was used in the experiment was MAU, a 40-year-old male. The CV syllables were selected from the monosyllable dataset. The carrier phrase was a combination of two partial sentences that were taken from the sentence data set. The same carrier sentence was used for all targets. The beginning position of the target vowel was adjusted to 150 ms from the conclusion of the pretarget carrier phrase in order to control the amount of energy overlapping to the target from previous portions. The beginning position of the target vowel of 150 ms was selected because mean durations of Japanese

TABLE I. Reverberant conditions used in the experiments. The impulse responses h1–h5, h7, and h8 were obtained by multiplying the exponential decay by the original impulse response h6 to achieve the desired reverberant conditions. The RT values are the average RTs derived from early decay time (EDT) at the center frequencies of 0.5, 1, and 2 kHz of the 1-oct bandpassed impulse response.

Impulse response	h1	h2	h3	h4	h5	h6	h7	h8
RT (s)	0.4	0.5	0.7	0.9	1.0	1.1	1.2	1.3
	Experiment II				Experiment I			

TABLE II. Twenty-four nonsense consonant-vowel syllables (CVs) used in the experiments.

	Voiceless C +Vowel	Voiced C +Vowel
Stop C +Vowel	/pa/ /ta/ /ka/ /pi/ /ki/	/ba/ /da/ /ga/ /bi/ /gi/
Fricative C +Vowel	/sa/ /ʃa/ /ha/ /ʃi/ /hi/	
Affricate C +Vowel	/tʃa/ /tʃa/	/dza/ /dʒa/ /dʒi/
Nasal C +Vowel		/ma/ /na/ /mi/ /ni/

syllables are between 150 and 200 ms (Arai and Greenberg, 1997). The root-mean-square (rms) energy in the CVs that had the same vowel was normalized, and then the ratio of the rms in the carrier phrase relative to the rms energy in the CVs was normalized.

The stimuli consisted of the original (unprocessed) signals with reverberation (Org_rev) and the processed signals with reverberation (Proc_rev). RTs of 0.9, 1.0, 1.1, 1.2, and 1.3 s were used in experiment I. A total of 240 stimuli were prepared (5 impulse responses $\times$24 CVs $\times$2 processing conditions).

B. Subjects

Twenty-four subjects with normal hearing (14 males and 10 females, aged 18 to 26 years old) participated in the experiment. All were native speakers of Japanese.

C. Procedure

The experiment, controlled by a computer, was conducted in a sound-treated room. The stimuli were presented through headphones (STAX SR-303). Five practice sessions were held to familiarize the subjects with the procedure before the experiment. The sound level was adjusted to the comfort level of each subject beforehand, and this level was maintained throughout the experiment. A stimulus was presented at each trial during the main session, and then 24 CVs in Kana orthography were shown on a PC screen. While the stimuli were presented to the subjects, the letters representing the 24 CVs were not visible. (They were shown on the screen only after the stimuli had been presented.) Subjects were forced to choose 1 of 24 CVs (syllables) by clicking a button on the screen using a computer mouse. Once they had selected a CV, the next stimulus was presented. For each subject, the stimuli were presented randomly.

D. Experimental results

Subject's responses were scored in each reverberation and processing condition. The results of 22 subjects (two outliers were excluded) were then analyzed. Circles in Fig. 3 show the mean percent correct for each reverberation and processing condition. Table III lists the types of errors and changes in confusions by the steady-state suppression, classified according to manner and place of articulation feature, as well as the overall results, for both processing conditions in experiment I.

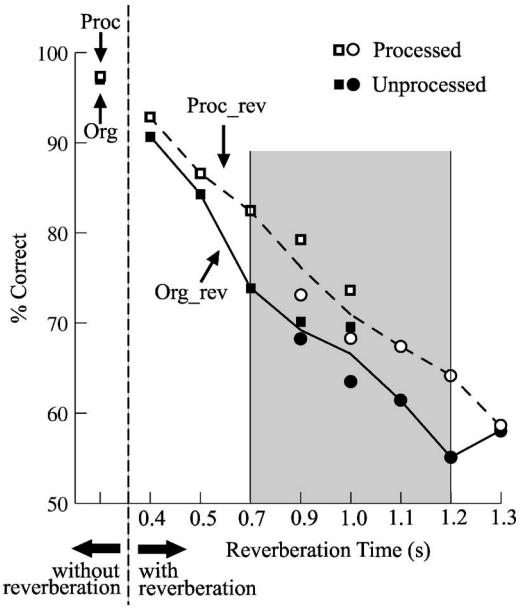


FIG. 3. Mean percent correct for identification of 24 CV syllables by 22 subjects with or without the steady-state suppression in experiments I and II. Open circles represent the mean performance of processed signals with reverberation and closed circles represent the mean performance of unprocessed signals with reverberation for five RTs (0.9, 1.0, 1.1, 1.2, and 1.3 s) in experiment I. Open squares indicate the mean performance of processed signals and closed squares indicate the mean performance of unprocessed signals for five RTs (0.4, 0.5, 0.7, 0.9, and 1.0 s) and without reverberation in experiment II. The dashed line separates the results for the signals with reverberation (right) from those without reverberation (left). Lines drawn through the data pass through the mean performance for two reverberant conditions (RTs of 0.9 and 1.0) repeated in both experiments. The shaded region shows the RTs where significant differences were found.

A 2×5 ANOVA for repeated measures was performed with processing and impulse response, confirming the main effects of processing ($p<0.01$) and the impulse response ($p<0.01$) to be statistically significant. The interaction effect was significant ($p=0.02$). For the comparison of the mean percent correct values between the with and without processing conditions, t tests were performed for each impulse response. In these tests, the probability value was modified from $p=0.05$ to $p=0.021$ in 2×5 conditions (Kirk, 1995). A statistically significant difference was obtained for RTs of 1.0–1.2 s [RT=1.0 s:Org_rev (63.4%), Proc_rev (68.3%), $p=0.02$; RT=1.1 s:Org_rev (61.3%), Proc_rev (67.4%), $p<0.01$, and RT=1.2 s:Org_rev (55.1%), Proc_rev (64.2%), $p<0.01$].

V. EXPERIMENT II

The results of experiment I showed statistically significant improvements on syllable identification to the steady-state suppression for the shorter RTs (1.0–1.2 s), which may suggest that steady-state suppression can improve syllable identification for RTs shorter than 0.9 s. On the other hand, the amount of overlap-masking is expected to be reduced as RT decreases. Therefore, as the RT is shortened, the difference between syllable identification scores in experiments

TABLE III. Types of errors and changes in confusions by the steady-state suppression, classified according to manner and place of articulation feature, as well as the overall results, for both processing conditions in experiment I.

	Overall	
Proc_rev > Org_rev	/ka/ /na/ /ga/ /da/ /ma/ /bi/ /pa/ /ba/ /pi/ /gi/ /dza/ /ta/ /sa/ /mi/ confusions from /ka/ to /pa/, from /ga/ to /ta/, and from /da/ to /na/ were greatly reduced	
Proc_rev < Org_rev	/ni/ /ki/ /tʃi/ /ʃa/ /dʒi/ /dʒa/ /ʃi/	
Proc_rev = Org_rev	/ha/ /hi/ /tʃa/	
Types of errors	Org_rev had correct rates below 50% in /ka/, /ba/, /ga/, /ha/, /bi/, /gi/, /da/, /na/ /ni/, /ta/, and /tʃi/ (in order of number of errors)	
	Place	Manner
Proc_rev > Org_rev	velars, bilabials, and alveolars confusions from velars to bilabials were greatly reduced	stops, affricates, and nasals confusions from stops to fricatives and from stops to nasals were greatly reduced
Proc_rev < Org_rev	palatals and postalveolar	fricatives
Proc_rev = Org_rev	glottal	
Types of errors	in both Org_rev and Proc_rev, confusions with bilabials were the most frequent	in Org_rev, nasals and stops were confused most frequently

using Org_rev and Proc_rev is predicted to decrease due to reduced overlap-masking. When the RT is increased, this difference also appears to decrease because steady-state suppression is less effective for excessively long RTs. Above a high RT threshold or below a low threshold, the expected difference in syllable identification is small. In order to investigate the lower limit of RT at which steady-state suppression improves syllable identification, experiment II is conducted with shorter RTs than experiment I, whereas the other parameters of the steady-state suppression, such as the suppression rate, were left unchanged.

A. Stimuli

The speech samples from experiment I were also used in experiment II. However, the stimuli consisted of the original (unprocessed) signals (Org), the processed signals (Proc), the original signals with reverberation (Org_rev), and the processed signals with reverberation (Proc_rev). RTs of 0.4, 0.5, 0.7, 0.9, and 1.0 s in Table I were used in experiment II. Therefore, a total of 288 stimuli were prepared, i.e., 5 (impulse responses + original condition) $\times$24 CVs $\times$2 (with or without) processing conditions.

B. Subjects

Twenty-two subjects with normal hearing (11 males and 11 females, aged 19 to 27 years old) participated in this experiment. All were native speakers of Japanese. Ten of the subjects had also participated in experiment I.[1]

C. Procedure

The procedure was the same as in experiment I, except for the total number of stimuli that were prepared (288 in experiment II). First, 240 stimuli (Org_rev and Proc_rev) were presented randomly, and then 48 stimuli (Org and Proc) were presented in random order.

D. Experimental results

Subjects' responses were scored for each reverberation and processing condition. Squares in Fig. 3 show the mean percent correct of 22 subjects for each reverberation and processing condition. Table IV lists the types of errors and changes in confusions by the steady-state suppression, classified according to manner and place of articulation feature, as well as the overall results, for both processing conditions in experiment II.

A 2×5 ANOVA for repeated measures was performed, confirming the main effects of processing ($p<0.01$) and impulse response ($p<0.01$) to be statistically significant. The interaction effect was significant ($p=0.006$). For comparison of the mean percent correct values obtained with and without processing, t tests were performed for each RT. In these tests, the probability value was modified from $p=0.05$ to $p=0.021$ in 2×5 conditions (Kirk, 1995). A statistically significant difference was obtained for an RT of 0.7 and an RT of 0.9 s [RT=0.7 s: Org_rev (73.9%), Proc_rev (82.4%), $p<0.01$; RT=0.9 s: Org_rev (70.1%), Proc_rev (79.2%), $p<0.01$].

VI. DISCUSSION

The main effect of reverberation was statistically significant for the mean percent correct of the targets (CV syllables) in the two experiments. In addition, syllable identification scores for an RT of 0.9 s (70.7%) was higher than that for an RT of 1.3 s (58.3%) in experiment I, and syllable identification scores for an RT of 0.4 s (91.8%) were higher than that for an RT of 1.0 s (71.7%) in experiment II. Therefore, it is confirmed that syllable identification scores decreases as RT increases, with or without processing.

The main effect of processing was statistically significant in the two experiments, and Proc_rev (66.3% in experiment I and 85.3% in experiment II) performed better than Org_rev (61.2% in experiment I and 80.9% in experiment II). The interaction for reverberation and processing was also statistically significant in the two experiments. This implies

TABLE IV. Types of errors and changes in confusions by the steady-state suppression, classified according to manner and place of articulation feature, as well as the overall results, for both processing conditions in experiment II.

	Overall	
Proc_rev > Org_rev	/ka/ /ta/ /da/ /bi/ /ba/ /ga/ /pi/ /na/ /pa/ /ma/ /ʃa/ /hi/ /tʃi/ /dza/ /dʒa/ /mi/ confusions from /ka/ to /pa/, from /ta/ to /pa/, from /ba/ to /ma/, and from /da/ to /na/ were greatly reduced	
Proc_rev < Org_rev	/ni/ /sa/ /ki/ /gi/ /dʒi/ /ha/	
Proc_rev = Org_rev	/ʃi/ /tʃa/	
Types of errors	Org_rev had correct rates below 50% in /ka/, /da/, /ha/, and /ta/ (in order of number of errors)	
	Place	**Manner**
Proc_rev > Org_rev	velars, bilabials, and alveolars confusions from velars to bilabials and from alveolars to bilabials were greatly reduced	stops and nasals confusions from stops to nasals and from stops to fricatives were greatly reduced
Proc_rev < Org_rev	palatals and glottal	fricatives and affricates
Proc_rev = Org_rev		
Types of errors	in both Org_revand Proc_rev, confusions with bilabials were the most frequent	in Org_rev, nasals and stops were confused most frequently

that the effect of steady-state suppression depends on RT. In order to compare syllable identification scores of Org_rev and Proc_rev in each reverberant condition, *t* tests (corrected for multiple comparison, see Secs. IV D and V D) were conducted. The results of these *t* tests showed that Proc_rev had significantly higher syllable identification scores than Org_rev for RTs of 1.0–1.2 s in experiment I and RTs of 0.7 and 0.9 s in experiment II. The results indicate that steady-state suppression improves syllable identification in specific reverberant environments.

In the present paper, the range of reverberant conditions in which the steady-state suppression (Arai *et al.*, 2001, 2002) prevents the degradation of syllable identification was investigated. Two perceptual tests were conducted with a set of artificial impulse responses: RTs of 0.9–1.3 s in experiment I and RTs of 0.4–1.0 s in experiment II. Steady-state suppression yielded statistically significant improvement in the ranges of 1.0–1.2 s in experiment I and 0.7–0.9 s in experiment II when the 40% suppression rate was used.

In addition, a close to statistically significant difference was obtained for an RT of 0.9 s in experiment I [Org_rev (66.5%), Proc_rev (73.1%), $p=0.03$]. The results obtained from experiments I and II indicate that the steady-state suppression is effective for an RT of 0.9 s because a statistically significant effect of processing may appear continuously in a certain range of RTs. Therefore, the effect of steady-state suppression with a 40% suppression rate prevents the degradation of syllable identification for RTs of 0.7–1.2 s.

VII. GENERAL DISCUSSION

A. Effect of the steady-state suppression

The main reason that steady-state suppression improved syllable identification is that steady-state suppression reduces the effect of overlap-masking. The results obtained in experiments I and II show that the amount of overlap-masking on identification of the following consonants was reduced using the 40% suppression rate in the steady-state suppression when the preceding vowel is /a/ under a certain RT. This supports the notion that for a range of RTs syllable identification is degraded by the effect of the overlap-masking caused by reverberation.

Overall, syllable identification scores for the processed condition were higher than that of the original (unprocessed) condition, but statistically significant improvement was not observed for RTs less than 0.7 s or greater than 1.2 s under the current experimental conditions using the 40% suppression rate. The steady-state suppression improved syllable identification by 0.4%–9.1% for RTs in the range of 0.4 to 1.3 s in the experiments. However, syllable identification may be unaffected by the steady-state suppression when the RT is extremely long or extremely short, i.e., less than 0.1 s or greater than 10 s, because of the flooring effect that results from too much overlap-masking for longer RTs and the ceiling effect that results from too little overlap-masking for shorter RTs. This means that steady-state suppression with particular parameters is effective only for a specific range of reverberant conditions. The decrease in the difference between syllable identification scores of Org_rev and Proc_rev at an RT of 1.3 s in Fig. 3 shows that 1.2 s is the upper limit of RT at which statistically significant improvements are observed under the current experimental conditions. The decrease in the difference between syllable identification scores of Org_rev and Proc_rev as RT is shortened in Fig. 3 indicates that 0.7 s is the lower limit of RT at which statistically significant improvements are observed under the present experimental conditions. The upper and lower limits can be varied as parameters of steady-state suppression are modified.

Figure 3 shows that syllable identification scores of Org and Proc (on the left side of the dashed line in Fig. 3) were approximately the same, indicating that suppressing steady-state portions of a speech signal do not degrade the intelligibility of a clean speech signal. Figure 3 also shows that syllable identification decreased approximately 6% from an RT of 0.4 to an RT of 0.5 s. The relatively large difference might be caused by the relatively weak direct sound in the modified impulse response that was used in the present study. Real

rooms with RTs of 0.4 or 0.5 s would typically have a stronger direct sound, and consequently smaller relative changes in the early part of the impulse response.

B. Comparison of experiments I and II

In addition to the fact that Proc_rev performed better than Org_rev in both experiments, syllable identification scores were higher in experiment II than in experiment I in general. The mean percent correct in experiment II (85.3% in Proc_rev and 80.9% in Org_rev) was higher than that in experiment I (66.3% in Proc_rev and 61.2% in Org_rev) with or without processing. Shorter reverberant conditions in experiment II account for this, because shorter RTs result in less overlap-masking and therefore higher syllable identification scores.

Whether or not syllable identification was affected significantly by contextual vowels (those preceding and following target C) was investigated, whether the vowels were the same or different. The subject's responses to targets that contain the same consonant C (e.g., /ka/ and /ki/) were compared. The recognition rate was, in many cases, higher for C paired with /i/ than C paired with /a/ in both experiments I and II, although the difference was less noticeable in experiment II. The reason for the higher recognition rate for C paired with /i/ may be that the previous vowel of the target was /a/. When a target vowel is /a/, a target consonant lies between the same vowels. The frequency regions of the target consonant and the contextual vowels /a/ may be more overlapped than those of the target consonant and different contextual vowels. Therefore, the energy of the previous vowel of the target would overlap the target consonant more and syllable identification scores for the target would be reduced.

The target identification may change as different preceding vowels are used, because different vowels have different intensities and energy distributions across frequency. However, in the present study, only /a/ was used as a preceding vowel, in order to control the experimental environment. The effect of steady-state suppression on consonant-vowel combinations with other preceding vowels remains for future investigation.

C. Modulation spectra

The important modulation frequencies for speech perception have been reported to be 1–16 Hz, especially 4 Hz, where the modulation spectrum usually reaches its maximum value (Drullman *et al.*, 1994; Arai *et al.*, 1999). When reverberation is present, the peak of the modulation spectrum shifts to the lower modulation frequency and the modulation index is reduced (Houtgast and Steeneken, 1985). Thus, there is a strong relationship between speech intelligibility and the modulation spectrum.

The effect of steady-state suppression could be expressed as a change in the modulation spectrum of a speech signal. Thus, the modulation spectra of signals with and without steady-state suppression were compared. The modulation spectrum is derived based on a frequency analysis of the temporal envelope of a bandpassed signal. First, speech signals were divided into four bands as described in Arai and Greenberg (1998) and a modulation spectrum was calculated in each band. These modulation spectra were then averaged over 24 sentences (the target CVs with the carrier sentence) used in the experiments. The averaged modulation spectra were then smoothed by a three-point moving average on the modulation frequency axis. Figure 4 shows the modulation spectra of original signals (ms_org) and steady-state suppressed signals (ms_proc). When ms_org and ms_proc were compared in Fig. 4, their modulation indices around 4 Hz and above 10 Hz are intensified by the steady-state suppression in all bands. Statistically significant improvements in identification scores by the steady-state suppression can be observed because of the strong increase in the modulation index around 10 Hz in ms_proc, which is not observed in ms_org.

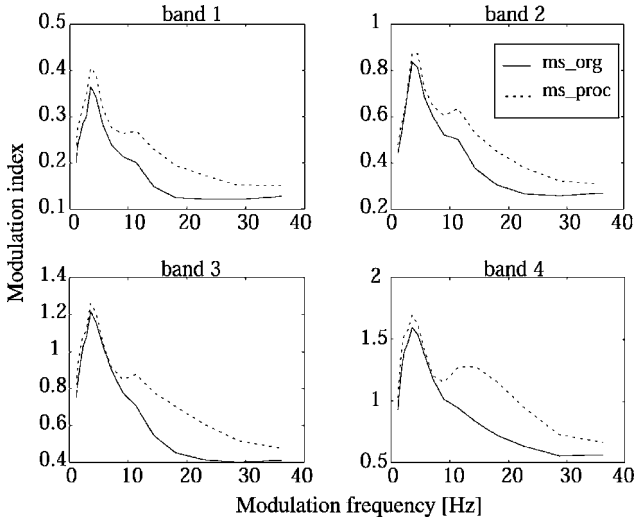


FIG. 4. Average modulation spectra of 24 speech sentences used in the experiments with and without steady-state suppression [solid line: without processing (ms_org), dashed line: with processing (ms_proc)] for each frequency band. Frequency regions are constituted with band 1 (0–800 Hz), band 2 (800–1600 Hz), band 3 (1600–3200 Hz), and band 4 (3200–8000 Hz).

These findings indicate that syllables and transitions are enhanced by the steady-state suppression. As Greenberg and Arai (2004) pointed out, as initially reported by Drullman *et al.* (1994) and later confirmed by Arai *et al.* (1999), speech intelligibility depends on low-frequency modulation of signal energy between 3 and 10 Hz, and this low-frequency modulation reflects the variation in syllable emphasis. It has also been shown that a modulation frequency of approximately 4 Hz reflects the syllabic rate of speech, whereas modulation frequencies above 10 Hz reflect the phonemic rate of speech (Duquesnoy and Plomp, 1980). In other words, the steady-state suppression preemptively enhances modulation indices around frequencies that are important for speech perception, so that the modulation index (that is, intelligibility) is prevented from being reduced by reverberation. The syllable identification scores of approximately 100% without reverberation might be a result of the ceiling effect.

When reverberation is present, a correlation exists between reverberation and the modulation spectrum of

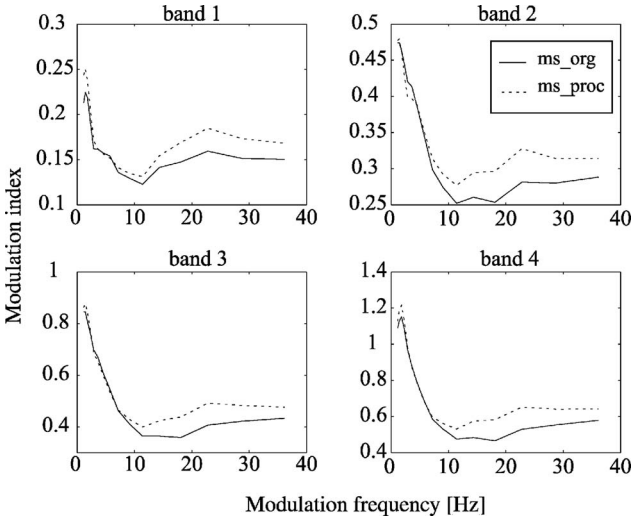


FIG. 5. Average modulation spectra of 24 speech sentences used in the experiments with and without steady-state suppression after reverberation (RT of 1.2 s) [solid line: without processing (ms_orgrev), dashed line: with processing (ms_procrev)] for each frequency band. Frequency regions are constituted with band 1 (0–800 Hz), band 2 (800–1600 Hz), band 3 (1600–3200 Hz), and band 4 (3200–8000 Hz).

Org_rev. This correlation can be explained by the theory of speech transmission index (STI), where the modulation transfer function depends on reverberation. Figure 5 shows the modulation spectra of original signals after reverberation (ms_orgrev) and steady-state suppressed signals after reverberation (ms_procrev). An RT of 1.2 s was used in Fig. 5 as a reference. The difference between ms_orgrev and ms_procrev is relatively small so that a change in this difference could not be seen across RTs. However, modulation indices above 10 Hz in ms_procrev (after reverberation) are still enhanced compared to those in ms_orgrev (after reverberation) in Fig. 5.

VIII. CONCLUSIONS

In the present paper, the effect of the steady-state suppression proposed by Arai *et al.* (2001, 2002) on syllable identification was explored under various reverberant conditions. By fixing the parameters of the steady-state suppression (e.g., the 40% suppression rate), two perceptual tests were conducted, with 22 subjects with normal hearing in each experiment. In order to simulate reverberant conditions in small- to medium-sized halls, nonsense monosyllables in a Japanese carrier phrase with and without the steady-state suppression were convolved with various artificial impulse responses of a room (RTs of 0.4–1.3 s). The results showed that statistically significant improvements by steady-state suppression were obtained for RTs of 0.7–1.2 s. In particular, steady-state suppression enabled subjects to distinguish voiced consonants, stop and nasal consonants, and bilabial, alveolar, and velar consonants more accurately in all experiments. These findings show that steady-state suppression is an effective preprocessing method for mitigating the degradation of syllable identification by reducing the effect of overlap-masking under specific reverberant conditions.

ACKNOWLEDGMENTS

This research was supported by Grants-in-Aid for Scientific Research (A-2, 16203041) and Grants-in-Aid for JSPS Fellows (176911), both from the Japan Society for the Promotion of Science. The authors would like to thank Hideki Tachibana, Kanako Ueno, and Sakae Yokoyama for offering the use of the impulse response data, and Chikashi Michimata and Hirofumi Kamata for their guidance regarding the statistical analysis. The authors are grateful to Peter Svensson, Dawn Behne, Paul Calamia, Terri Lander, and Michael Shire for their useful comments and for proofreading the manuscript.

[1]No learning evidence was observed in these ten subjects, since syllable identification scores in each processing and reverberant condition were approximately the same for the 10 subjects and the entire set of subjects. (Syllable identification scores for an RT of 0.9 s was 71.2% for the ten subjects and 70.1% for the entire set of 22 subjects in Org_rev and was 77.1% for the ten subjects and 79.2% for the entire set of 22 subjects in Proc_rev. For an RT of 1.0 s, syllable identification scores was 66.7% for the ten subjects and 69.5% for the entire set of 22 subjects in Org_rev and was 72.5% for the ten subjects and 73.9% for the entire set of 22 subjects in Proc_rev.)

Arai, T., and Greenberg, S. (**1997**). "The temporal properties of spoken Japanese are similar to those of English," *Proc. ESCA Eurospeech*, Vol. **2**, pp. 1011–1014.

Arai, T., and Greenberg, S. (**1998**). "Speech intelligibility in the presence of cross-channel spectral asynchrony," *Proc. IEEE ICASSP*, Vol. **2**, pp. 933–936.

Arai, T., Pavel, M., Hermansky, H., and Avendano, C. (**1999**). "Syllable intelligibility for temporally filtered LPC cepstral trajectories," J. Acoust. Soc. Am. **105**(5), 2783–2791.

Arai, T., Kinoshita, K., Hodoshima, N., Kusumoto, A., and Kitamura, T. (**2001**). "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," Proc. Autumn Meet. Acoust. Soc. Jpn., Vol. **1**, pp. 449–450 (in Japanese).

Arai, T., Kinoshita, K., Hodoshima, N., Kusumoto, A., and Kitamura, T. (**2002**). "Effects of suppressing steady-state portions of speech on intelligibility in reverberant environments," Acoust. Sci. & Tech. **23**(4), 229–232.

Avendano, C., and Hermansky, H. (**1996**). "Study on the dereverberation of speech based on temporal envelope filtering," *Proc. ESCA ICSLP*, Vol. **2**, pp. 889–892.

Bolt, R. H., and MacDonald, A. D. (**1949**). "Theory of speech masking by reverberation," J. Acoust. Soc. Am. **21**(6), 577–580.

Crandell, C. C., and Smaldino, J. J. (**2000**). "Classroom acoustics for children with normal hearing and with hearing impairment," J. Lang. Speech Hear. Services Schools **31**(4), 362–370.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**(2), 1053–1064.

Duquesnoy, A. J., and Plomp, R. (**1980**). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. **68**(2), 537–544.

Elliot, L. L. (**1982**). "Effects of noise on perception of speech by children and certain handicapped individuals," J. Sound Vib. **16**(12), 10–14.

Finitzo-Hieber, T., and Tillman, T. (**1978**). "Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children," J. Speech Hear. Res. **21**(3), 440–458.

Flanagan, J. L., Johnston, J. D., Zahn, R., and Elko, G. W. (**1985**). "Computer-steered microphone arrays for sound transmission in large rooms," J. Acoust. Soc. Am. **78**(5), 1508–1518.

Furui, S. (**1986**). "On the role of spectral transition for speech perception," J. Acoust. Soc. Am. **80**(4), 1016–1025.

Gordon-Salant, S. (**1986**). "Recognition of natural and time/intensity altered CVs by young and elderly subjects with normal hearing," J. Acoust. Soc. Am. **80**(6), 1599–1607.

Greenberg, S., and Arai, T. (**2004**). "What are the essential cues for understanding spoken language?" IEICE Trans. Inf. Syst. **87-D**(5), 1059–1070.

Hermansky, H., and Morgan, N. (**1994**). "RASTA processing of speech," IEEE Trans. Speech Audio Process. **2**(4), 578–589.
Hodoshima, N., Arai, T., and Kusumoto, A. (**2002**). "Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments," *Proc. Forum Acusticum Sevilla*.
Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**(3), 1069–1077.
Kinoshita, K., Nakatani, T., and Miyoshi, M. (**2005**). "Efficient blind dereverberation framework for automatic speech recognition," *Proc. ISCA Interspeech*, pp. 3145–3148.
Kirk, R. E. (**1995**). *Experimental Design: Procedures for the Behavioral Sciences* (Brooks/Cole, Pacific Grove, CA).
Kitamura, T., Kinoshita, K., Arai, T., Kusumoto, A., and Murahara, Y. (**2000**). "Designing modulation filters for improving speech intelligibility in reverberant environments," *Proc. ICSLP*, Vol. **3**, pp. 586–589.
Knudsen, V. O. (**1929**). "The hearing of speech in auditoriums," J. Acoust. Soc. Am. **1**(1), 56–82.
Kusumoto, A., Arai, T., Takahashi, M., and Murahara, Y. (**2000**). "Modulation enhancement of speech as a preprocessing for reverberant chambers with the hearing-impaired," *Proc. IEEE ICASSP*, Vol. **2**, pp. 853–856.
Kusumoto, A., Arai, T., Kinoshita, K., and Hodoshima, N. (**2005**). "Modulation enhancement of speech by preprocessing for improving intelligibility in reverberant environment," Speech Commun. **45**(2), 101–113.
Langhans, T., and Strube, H. W. (**1982**). "Speech enhancement by nonlinear multiband envelope filtering," *Proc. IEEE ICASSP*, Vol. **7**, pp. 156–159.
Miyoshi, M., and Kaneda, Y. (**1988**). "Inverse filtering of room acoustics," IEEE Trans. Acoust., Speech, Signal Process. **36**(2), 145–152.
Nábělek, A. K., and Donahue, A. M. (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**(2), 632–634.
Nábělek, A. K., and Pickett, J. M. (**1974**). "Monaural and binaural speech perception through hearing aids under noise and reverberation," J. Speech Hear. Res. **17**, 724–739.
Nábělek, A. K., and Robinette, L. (**1978**). "Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **63**(1), 187–194.
Nábělek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**(4), 1242–1248.
Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap- and self-masking in consonant identification," J. Acoust. Soc. Am. **86**(4), 1259–1265.
Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**(3), 1581–1592.
Schroeder, M. R. (**1981**). "Modulation transfer functions: Definition and measurement," Acustica **49**(3), 179–182.
Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**(3), 695–705.
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**(2), 663–666.
Uchanski, R. M., Geers, A. E., and Protopapas, A. (**2002**). "Intelligibility of modified speech for young listeners with normal and impaired hearing," J. Speech Lang. Hear. Res. **45**, 1027–1038.

Auditory-visual speech perception and synchrony detection for speech and nonspeech signals

Brianna Conrey[a] and David B. Pisoni
Department of Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Bloomington, Indiana 47405

(Received 17 November 2004; revised 27 January 2006; accepted 13 March 2006)

Previous research has identified a "synchrony window" of several hundred milliseconds over which auditory-visual (AV) asynchronies are not reliably perceived. Individual variability in the size of this AV synchrony window has been linked with variability in AV speech perception measures, but it was not clear whether AV speech perception measures are related to synchrony detection for speech only or for both speech and nonspeech signals. An experiment was conducted to investigate the relationship between measures of AV speech perception and AV synchrony detection for speech and nonspeech signals. Variability in AV synchrony detection for both speech and nonspeech signals was found to be related to variability in measures of auditory-only (A-only) and AV speech perception, suggesting that temporal processing for both speech and nonspeech signals must be taken into account in explaining variability in A-only and multisensory speech perception. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2195091]

PACS number(s): 43.71.Rt [KWG] Pages: 4065–4073

I. INTRODUCTION

The temporal coordination of signals from multiple sensory modalities is a necessary condition for multisensory integration (Stein and Meredith, 1993). However, multisensory signals do not have to be precisely physically synchronous in order to be perceived as single temporal events. For example, previous research using detection, discrimination, syllable identification, and sentence intelligibility tasks has consistently identified an "intersensory synchrony window" (Lewkowicz, 1996) of several hundred milliseconds during which human observers perceive desynchronized auditory and visual signals as synchronous. Table I summarizes previous studies that have attempted to define the temporal limits of a window for auditory-visual (AV) synchrony perception in normal-hearing adults.

Three characteristics of the AV synchrony window are relatively consistent across studies, despite differences in the stimuli, tasks, and statistical procedures used by different investigators. First, the synchrony window is on the order of several hundred milliseconds wide. The width of the window may reflect general information processing constraints that link events perceptually when they occur within around 200 ms of each other (Munhall *et al.*, 1996). A multisensory interaction window that is several hundred milliseconds wide is also consistent with the estimates of the temporal window for multisensory enhancement and depression reported in electrophysiological studies in animals (King and Palmer, 1985; Meredith, 2002; Meredith *et al.*, 1987; Stein and Meredith, 1993). Second, the synchrony window is asymmetrical: Visual-leading asynchronies are more difficult to detect than auditory-leading asynchronies. Finally, the width of the synchrony window varies a great deal across individuals, although judgments of AV synchrony appear to be highly consistent within individuals (Stone *et al.*, 2001).

The present study focused on the individual variability in the AV synchrony window. It has been proposed that greater attunement to AV synchrony in speech may be associated with better visual-only speech perception ("lipreading" or "speechreading") skills (McGrath and Summerfield, 1985), or alternatively with greater multisensory benefit from having both auditory and visual speech signals available (Grant and Seitz, 1998). In particular, individuals who are better at utilizing visual-only or auditory-plus-visual information to understand speech may also be more attuned to detailed temporal relationships between auditory and visual speech signals.

Although previous studies suggest that the perception of AV synchrony in speech is related to AV speech perception abilities, it is possible that the variability in the synchrony window is also related to the perception of multisensory nonspeech signals. For instance, variability in sensory processing, perceptual learning, or cognitive abilities such as visual or auditory attention could all lead to individual differences in the size of the synchrony window for both speech and nonspeech signals. However, because most previous studies investigating AV synchrony perception have used either speech or nonspeech signals but not both, it is difficult to judge the extent to which the AV synchrony windows for speech and nonspeech signals are similar or related and also the extent to which each of these synchrony windows is related to measures of speech perception. Addressing these issues may lead to new insights about the relationship between general multisensory perception and multisensory perception of speech.

The primary aim of this study was to further examine the individual variability in the synchrony window for speech and nonspeech signals and the relationship of this variability to auditory-only, visual-only, and AV speech per-

[a]Electronic mail: bconrey@indiana.edu

TABLE I. Previous research on AV synchrony perception.

Study	*N*	Stimulus type	Visual stimulus	Auditory stimulus	Task[a]	Asynchrony levels[b]	Window (ms)
Dixon and Spitz (1980)	18	Nonspeech	Hammer hitting nail	Hammering sound	D (MOA)	n/a	A75V-V188A
Dixon and Spitz (1980)	18	Speech	Passages	Passages	D (MOA)	n/a	A131V-V258A
Grant and Greenberg (2001)	9	Speech	Sentences	Sentences	Int	A400, 300, 200, 160, 120, 80, 40V, V40, 80, 120, 160, 200, 300, 400A	A40-80V-V200A
Grant *et al.* (2003)	26	Speech	Sentences	Sentences	D (FCT)	n/a	A35V-V230A
Lewkowicz (1996)	10	Nonspeech	Bouncing green disk	Descending tone	D (MCS)	A50, 80, 110, 140V V110, 140, 170, 200A	A65V-V112A
Massaro *et al.* (1996)	28	Speech/ speechlike	Synthetic syllables	Natural and synthetic syllables	M Int	Exp. 1: A267, 167, 67V V67, 167, 267A Exp. 2: A533, 267, 133V V133, 267, 533A	A250V-V250A
McGrath and Summerfield (1985)	12	Speechlike	Liplike Lissajous pattern	Triangular wave approximating a CV syllable	D (FCT)	n/a	A79V-V138A
Munhall *et al.* (1996)	19	Speech	Natural syllables	Natural syllables	M Int	A360, 300, 240, 180, 120, 60V, V60, 120, 180, 240, 300, 360A	A60V-V240A
van Wassenhove *et al.* (2002)	36	Speech	Natural syllables	Natural syllables	M Int; D (MCS)	A467V-V467A	~A67V-~V167A

[a]D=synchrony detection; MCS=method of constant stimuli; MOA=method of adjustment; FCT=forced choice to threshold; Int=speech integration; M Int=McGurk syllable integration.
[b]All numbers are in milliseconds. The synchronous condition was always tested. A#V indicates that the auditory signal preceded the visual signal by # ms, and V#A indicates that the visual signal led by # ms.

ception. A secondary aim was to characterize the relationship or lack of a relationship between speech and nonspeech synchrony windows within participants.[1] We measured normal-hearing adults' ability to detect AV synchrony in both speech and nonspeech signals. We also measured the performance of the same participants on a sentence intelligibility task using auditory-only (A-only), visual-only (V-only), and AV presentation formats. Based on previous empirical results (Grant and Seitz, 1998; McGrath and Summerfield, 1985; Pandey *et al.*, 1986), we hypothesized that participants with better speech perception scores would be more accurate at detecting AV synchrony. In addition, because a preliminary study indicated that the synchrony window is highly similar for our particular speech and nonspeech signals (Conrey and Pisoni, 2003), we predicted that better synchrony detection for both speech and nonspeech signals would also be related to higher speech perception scores.

II. METHODS

A. Participants

Thirty-nine Indiana University undergraduates participated in the experiment. Data from 9 subjects were excluded from analysis for the following reasons: 3 participants did not follow directions on one or the other of the synchrony detection task conditions (they reversed their response hand), and 6 participants responded "synchronous" more than 50% of the time at all asynchrony levels and their data could not be fit with the same curve-fitting procedures used for the other participants. The remaining 30 participants included 25 females and 5 males who ranged in age from 18 to 22 years (mean age=19.67 years).

B. Procedure

Each participant completed both speech and nonspeech conditions in an AV synchrony detection task (described later). The participants also completed a modified version of the City University of New York (CUNY) Sentences Test (Boothroyd *et al.*, 1985). The CUNY test evaluates the perception of spoken sentences presented in the A-only, V-only, and AV modalities. Participants always completed the CUNY test first, followed by the AV synchrony detection task. The conditions in the CUNY test were presented in the order A-only, V-only, and then AV, and the order of the speech and nonspeech conditions of the AV synchrony detection task was counterbalanced across participants.

1. CUNY sentences

As administered in this experiment, the CUNY test consisted of three sets of 12 sentences. A different set of sentences was used for each of the A-only, V-only, and AV conditions for a given observer. Before beginning each condition, the participants were given three example trials in which they viewed and/or heard sentences but were not required to respond. During the test itself, each sentence was presented, then the participant was asked to respond by typing what he or she thought the speaker had said. The sentences were all spoken by the same female talker originally recorded by Boothroyd *et al.* (1985) and later digitized by Theresa Hnath-Chisolm and her graduate students at the University of South Florida. In our version of the CUNY test, the sentences were presented using SuperCard 4.1.1 (copyright 1989–2003, Solutions Etcetera) running a program created in SuperEdit (copyright 1989–2003, Solutions Etcetera) for the MacIntosh.

The participants were all normal-hearing adults. In order to avoid ceiling effects in the A-only and AV conditions, the auditory signal was degraded to make the test more difficult. Specifically, following the methods for locally time-reversed speech (TRS) described by Saberi and Perrott (1999), the signal was divided into 80-ms long segments. Each segment was time-reversed, and then the segments were reassembled in the original order. TRS was chosen as the method to degrade performance because reversing the time segments forced participants to perceptually integrate over long time scales in order to understand what was said. The choice of TRS as it pertains to our results will be addressed in Sec. IV. Eighty ms was chosen for the reversal interval because pilot testing in our lab found that this level was more difficult than the 50-ms interval, at which participants performed at near-ceiling levels, and less difficult than the 100-ms interval, which was nearly impossible for most participants. The 80-ms interval seemed likely to produce a range of variability in results without unduly frustrating participants.

2. AV synchrony detection

a. Stimuli. The nonspeech stimuli were modeled after those used in a PET study by Bushara *et al.* (2001) that investigated the neural correlates of AV asynchrony detection for several asynchrony levels. The present study used a 4-cm-diam red circle paired with a 2000-Hz tone. As in the earlier Bushara *et al.* study, the visual and auditory stimuli were both 100 ms in duration. The nonspeech stimuli were not intended to be analogous to the speech stimuli. Instead, they were chosen as a simple static stimulus to contrast with the complex dynamic speech stimuli and to test for the possibility that multisensory processing even for simple nonspeech signals might be related to variability in AV speech perception.

For the speech condition, 10 familiar English words were chosen from the Hoosier Audiovisual Multitalker Database (Lachs and Hernandez, 1998; Sheffert *et al.*, 1996), which contains digitized AV movies of isolated monosyllabic words spoken by single talkers. The most intelligible of the eight talkers in the database was determined in a previous study, and A-only, V-only, and AV intelligibility data had been collected for all her utterances (Lachs, 1999; Lachs and Pisoni, 2004a, 2004b). In this study, all 10 words were spoken by this talker. All of the words used in the present experiment had 100% intelligibility scores for both auditory-alone and AV presentation.

The experimental stimuli used in this study were created using Final Cut Pro 3 (copyright 2003, Apple Computer, Inc.). The visual and auditory stimuli were combined beforehand into precompiled movies rather than being assembled "on the fly" by the computer during the experiment. Precompiling the movies ensured reliable timing of the auditory and visual signals, essential for this experiment. For the asynchronous speech stimuli, the portions of the audio and video tracks that did not overlap with each other were edited from the stimulus movie. The removed portions did not contain any speech sounds, and the finished movies started and ended with the mouth in a neutral, nonarticulating position. The movies were edited in this way so that the participants would be unable to rely on any global temporal cues such as the audio track playing while the screen was blank to determine if the movie was synchronous. Instead, the participants had to make their judgments about synchrony based on whether the presented auditory and visual information was temporally coincident across both modalities.

Previous research on AV synchrony detection (Dixon and Spitz, 1980; Lewkowicz, 1996; Massaro and Cohen, 1993; Massaro *et al.*, 1996; McGrath and Summerfield, 1985; Pandey *et al.*, 1986) and pilot studies in our lab indicated that most normal-hearing young adult participants were able to judge AV stimuli as asynchronous with close to 100% accuracy when the auditory signal leads the visual signal by 300 ms (A300V ms) and when the visual signal leads the auditory signal by 500 ms (V500A ms). These asynchrony levels served as the upper limits of the stimuli presented in our study, and many individual participants could detect smaller levels of asynchrony with close to 100% accuracy. Because the videos used were recorded at a rate of 30 frames/s, each successive stimulus could differ by 33.33 ms. This resulted in 25 asynchrony levels covering a range of 800 ms, from A300V to V500A. Nine stimuli had auditory leads, one was synchronous, and fifteen had visual leads.

b. Procedure. The visual stimuli were presented on an Apple Macintosh G4 computer. Auditory stimuli were presented over Beyer Dynamic DT headphones at 70 dB SPL. PsyScope version 1.5.2 (Cohen *et al.*, 1993) was used for stimulus presentation and response collection. All participants were tested in both speech and nonspeech conditions. The conditions were blocked and were counterbalanced across participants.

The stimuli were presented in a single-interval synchrony detection task using the method of constant stimuli. On each trial, the participants were asked to judge whether the AV stimulus was synchronous or asynchronous ("in sync" or "not in sync"). Participants were instructed to press one button on a response box if the stimuli were synchronous and another if they were asynchronous. Response hand was counterbalanced across participants but kept constant for each participant on both conditions of the experiment so as to minimize confusion about the instructions. Before beginning each condition, the participants received instructions and were presented with examples of synchronous and asynchronous movies for that condition.

Each of the two conditions consisted of a block of 250 randomized trials, 10 for each of the 25 asynchrony levels. In the nonspeech condition, all trials used the same visual and auditory stimuli, the red circle and the 2000-Hz tone described earlier. In the speech condition, each of the 10 words was presented once at each asynchrony level. At the onset of each trial, a fixation mark ("+") was displayed on the computer screen for 200 ms and was followed by 300 ms of blank screen before the test stimulus was presented. The participant's response cued the onset of the next trial.

III. RESULTS

A. CUNY sentences

Participants' responses were printed out from an output file and scored by comparison with a master list of sentences. Responses were scored using a "whole-word" method: Each word was given a score of 1 or 0 points based on whether it was completely correct or had any errors, respectively. The reversal of two letters in a word, as in *appel* for *apple*, was

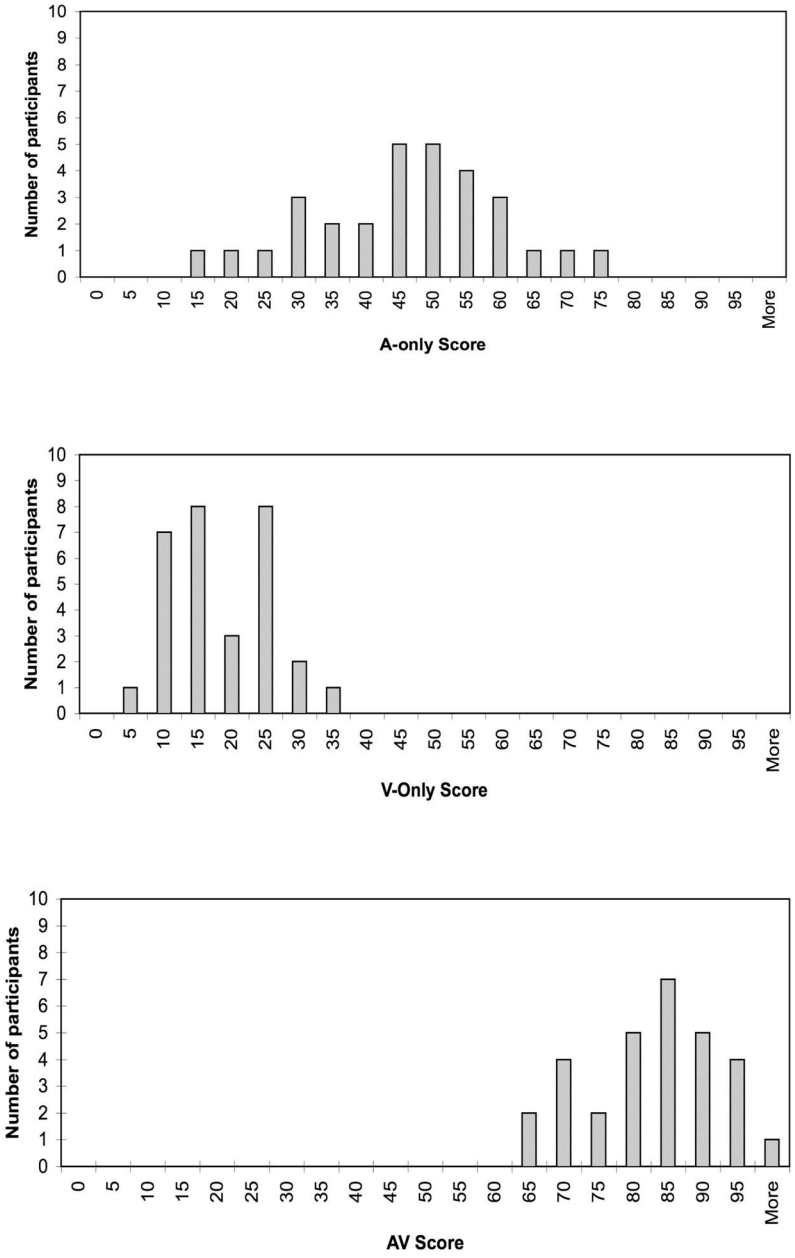


FIG. 1. Distribution of A-only, V-only, and AV scores on the CUNY sentences. All scores are out of a possible 102 whole words correct.

counted as correct as long as the reversal did not form a new English word. Similarly, one-letter typographical errors that did not result in the formation of a new word were also counted as correct.

As expected, the participants' performance was worst overall with the visual-only sentences, with a mean score of 16 words correct (s.d.=7) out of 102. A-only scores were next, with a mean of 44 (s.d.=15). Participants were best overall at the AV condition, with a mean score of 81 (s.d. =10). All but 3 individual participants had higher A-only than V-only scores, and all were best at the AV condition. The distributions of A-only, V-only, and AV scores are shown in Fig. 1.

B. AV synchrony detection

Throughout this report, we will refer to synchronous AV stimuli as the 0 condition, for 0-ms delay/lead. Because our figures represent auditory leads to the left side of 0 on the abscissa and visual leads to the right, "lower" will indicate further toward the auditory-leading side of the figure, and "higher" will indicate further toward the visual-leading side of the figure. Similarly, negative numbers will refer to the auditory signal leading the visual signal in time, and positive numbers to the visual signal leading the auditory signal.

The proportion of synchronous responses at each level of asynchrony was determined for each participant. The average proportions of synchronous responses are plotted for the speech and nonspeech conditions in Fig. 2. In looking at the figure, two major findings are apparent. First, the average range of asynchronies identified as synchronous was quite large, on the order of several hundred milliseconds. Second, this range was not centered at 0 ms, the physically synchronous condition, but was shifted to the right and centered on the visual-leading side of the continuum.

To quantify these findings and examine individual vari-

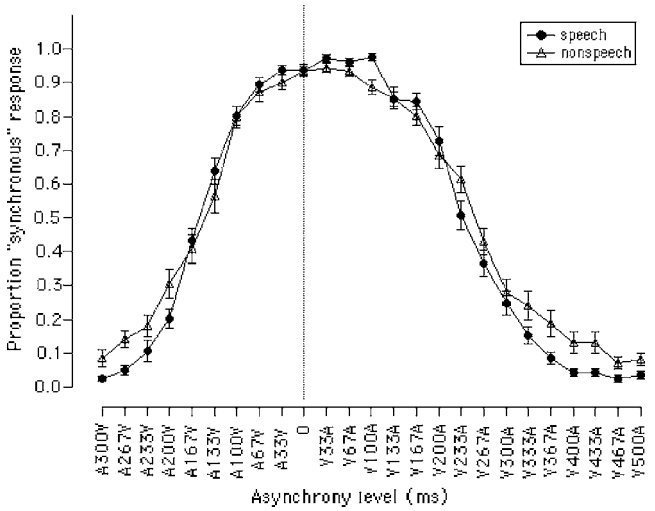


FIG. 2. Average "synchronous" response for all participants on the AV synchrony detection task, in the speech and nonspeech conditions. The dotted vertical line is at 0-ms asynchrony. Error bars represent one standard error of the mean.

ability in the results, we fit each condition from the individual participants' data with a Gaussian curve using Igor Pro 4.05A Carbon (copyright 1988-2002, WaveMetrics, Inc.). This analysis resulted in two curves for each participant, one for speech and one for nonspeech. The curve-fitting procedure provided estimates of the center and width of the synchrony window; from the center and width, the auditory- and visual-leading end points of the window were also calculated. The center was defined to be the mean point of synchrony (MPS), which indicated how far in ms the synchrony window was from being centered on the physically synchronous condition. For example, a MPS of 50 ms would indicate that a participant's synchrony window was shifted from physical synchrony by 50 ms of visual lead, whereas a MPS of −50 ms would indicated that the participant's synchrony window was shifted from physical synchrony by 50 ms of auditory lead. Although there were several possibilities for how to define the AV synchrony window's width, here width was defined as the range of asynchronies over which subjects responded that the signals were synchronous more than half the time. This width was estimated using the full width at half maximum (FWHM) of the fitted Gaussian curve. The auditory-leading threshold was the largest amount of auditory lead at which signals were judged synchronous more than half the time. It was estimated as the MPS minus the half width at half maximum. Similarly, the visual-leading threshold was the largest amount of visual lead at which signals were judged synchronous more than half the time. It was calculated as the MPS plus the half width at half maximum. The auditory- and visual-leading thresholds are referred to as "derived" measures in the rest of the report, because they are derived from the MPS and FWHM and are not independent of those measures.

The order of presentation of the speech and nonspeech conditions did not produce any significant effects on the MPS or FWHM for either condition [speech MPS: $t(28)=0.430$; speech FWHM: $t(28)=-0.379$; nonspeech MPS: $t(28)=0.226$; nonspeech FWHM: $t(28)=0.846$; all p's >0.05], so the speech-first and nonspeech-first participant groups were combined and analyzed together. Table II presents a summary of the MPS, FWHM, and auditory- and visual-leading threshold estimates obtained from a fit of the average response data weighted by the standard error. All statistical analyses used estimates from fitting curves to individual subject data.

TABLE II. Curve fits for averaged responses.[a]

Condition	MPS	FWHM	A-Lead	V-Lead
Speech	47 (15)	357 (61)	−131 (31)	225 (36)
Nonspeech	47 (43)	400 (66)	−153 (46)	247 (61)

[a]Standard deviations given in parentheses. All numbers are in milliseconds. Negative numbers indicate that the auditory signal led the visual signal. MPS=mean point of synchrony; FWHM=full width at half maximum; A-lead=auditory-leading threshold; V-lead=visual-leading threshold.

The distributions of the MPS and FWHM for the speech and nonspeech conditions are shown in Figs. 3 and 4, respectively. The distributions of the auditory- and visual-leading thresholds are not shown because these measures were derived directly from the MPS and FWHM.[2]

C. Correlational analyses

Correlational analyses were performed to examine the relationship between CUNY sentence scores (A-only, V-only, and AV) and performance on the AV synchrony detection task as measured by the MPS, FWHM, and auditory- and visual-leading threshold estimates. Visual inspection of the data revealed that some of the measures, particularly for the nonspeech condition of the AV synchrony detection task, displayed skewed distributions and outliers, and so correlational analyses were performed using Spearman's rho (r_s) rank-order correlations rather than Pearson's r correlations. Multiple regression analyses were generally redundant with the correlational analyses presented in the following because many of the measures were highly intercorrelated, so the results of regression analyses are not reported here.

1. CUNY sentences

Table III shows the bivariate correlations among the raw scores—A-only, V-only, and AV—on the CUNY sentences task. A-only and AV scores were positively correlated, $r_s=+0.53$, $p<0.01$. Higher A-only scores were associated with higher AV scores. The correlations between V-only and A-only or AV scores were small and not statistically significant.

2. AV synchrony detection

A summary of the AV synchrony detection correlations among MPS, FWHM, and auditory- and visual-leading threshold measures is given in Table IV. The FWHMs were positively correlated for the speech and nonspeech conditions, $r_s=+0.65$, $p<0.01$. Larger speech synchrony windows were associated with larger nonspeech synchrony windows. Also, the MPSs were positively correlated for the speech and nonspeech conditions, $r_s=+0.46$, $p<0.01$. The

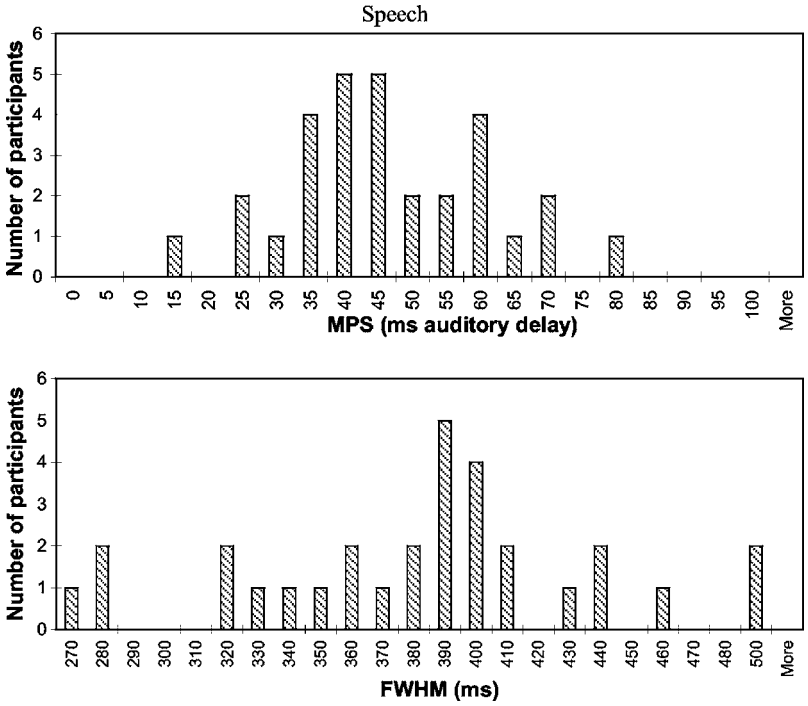


FIG. 3. Distribution of MPS (top panel) and FWHM (bottom panel) measures for the speech condition of the AV synchrony detection task.

distance of the center of the synchrony window from physical synchrony was positively associated for the speech and nonspeech conditions.

Among the measures derived from the MPS and FWHM, the visual-leading thresholds were positively correlated for the speech and nonspeech conditions, $r_s = +0.50$, $p < 0.01$. Also, for the speech condition only, the auditory-leading and visual-leading thresholds of the synchrony window were negatively correlated, $r_s = -0.47$, $p < 0.05$. Because auditory-leading thresholds were coded as negative numbers and visual-leading thresholds were coded as positive, this finding indicates that larger ("lower") auditory-leading thresholds were related to larger ("higher") visual-leading thresholds. The relationship between auditory- and visual-leading thresholds was not significant for the nonspeech condition, however ($r_s = +0.17$, $p > 0.05$). Several of the correlations between "primary" and "derived" synchrony detection performance measures were statistically significant, although this finding may not have much practical significance because the derived measures were calculated using the primary measures and so were not statistically independent from them.

3. AV synchrony detection and CUNY sentences

Table V summarizes the intercorrelations among the AV synchrony detection and CUNY sentence measures. Several

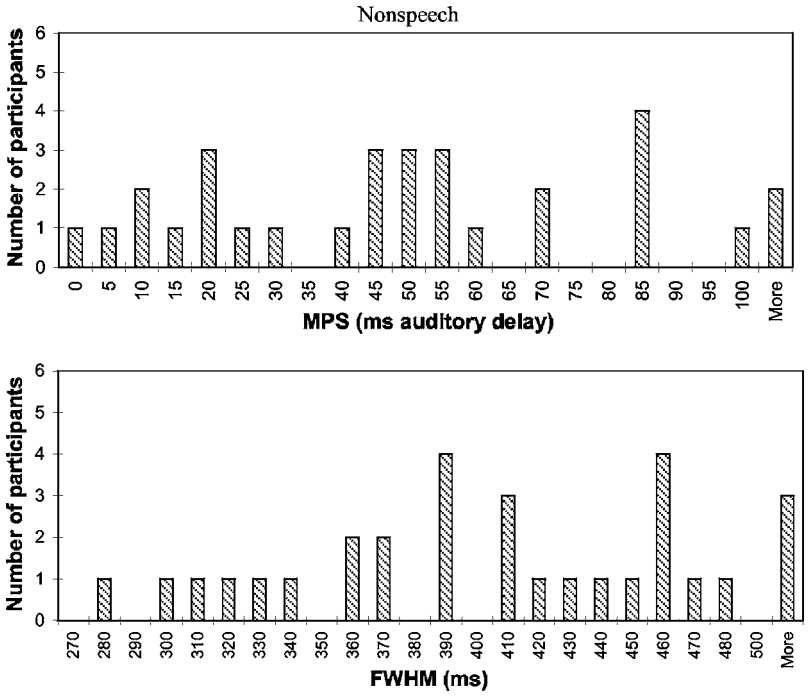


FIG. 4. Distribution of MPS (top panel) and FWHM (bottom panel) measures for the nonspeech condition of the AV synchrony detection task.

TABLE III. Correlations among measures of AV speech perception (CUNY sentences).

	A-Only	V-Only	AV
A-Only	⋯		
V-Only	−0.06	⋯	
AV	0.53[a]	0.03	⋯

[a]$p<0.01$.
[b]$p<0.05$.

of these correlations were statistically significant. CUNY AV raw scores were negatively correlated with the FWHM for both the speech and nonspeech conditions ($r_s=-0.47$, $r_s=-0.45$; $p<0.01$, $p<0.05$, respectively). CUNY A-only scores were also negatively correlated with the FWHM for the nonspeech condition, $r_s=-0.41$, $p<0.05$. Higher CUNY AV scores were thus associated with smaller synchrony windows, or lower tolerance for asynchrony, in both speech and nonspeech conditions. Higher A-only scores were also associated with smaller synchrony windows, but in the nonspeech condition only. Finally, AV and A-only scores were positively correlated with the auditory-leading threshold for the speech condition only ($r_s=+0.49$, $r_s=+0.39$, respectively; p's <0.05). This pattern indicates that higher CUNY AV and A-only scores for sentence intelligibility were associated with auditory-leading thresholds that were closer to physical synchrony.

IV. DISCUSSION

The results of this study demonstrate relationships between measures of AV sentence intelligibility and the ability to detect AV synchrony in both speech and nonspeech signals. First, significant within-participant correlations were found between AV synchrony detection for speech and nonspeech signals. In addition, participants who obtained higher AV sentence intelligibility scores tended to have smaller windows over which they identified AV signals—either speech or nonspeech—as synchronous and thus were better at detecting small differences in the relative timing between auditory and visual signals.

Similar to Grant and Seitz (1998), who did not find a relationship between V-only speech perception scores and the effect of AV asynchrony on speech perception, the present study failed to find a relationship between V-only

TABLE V. Correlations among measures of AV synchrony detection and AV speech perception.[a]

	CUNY sentences		
Synchrony Detection	A-Only	V-Only	AV
Primary			
Speech			
MPS	0.10	−0.29	0.08
FWHM	−0.31	−0.04	−0.47[b]
Nonspeech			
MPS	−0.12	−0.17	−0.09
FWHM	−0.41[c]	0.20	−0.45[b]
Derived			
Speech			
A-Lead	0.39[c]	−0.14	0.49[b]
V-Lead	−0.10	−0.15	−0.31
Nonspeech			
A-Lead	0.14	−0.17	0.21
V-lead	−0.31	0.14	−0.14

[a]Abbreviations as in previous tables.
[b]$p<0.01$.
[c]$p<0.05$.

TABLE IV. Correlations among measures of AV synchrony detection.[a]

	Primary				Derived			
	Speech		Nonspeech		Speech		Nonspeech	
	MPS	FWHM	MPS	FWHM	A-Lead	V-Lead	A-Lead	V-Lead
Primary								
Speech								
MPS	⋯							
FWHM	0.18	⋯						
Nonspeech								
MPS	0.46[b]	0.30	⋯					
FWHM	−0.04	0.65[b]	0.24	⋯				
Derived								
Speech								
A-Lead	0.28	−0.85[b]	−0.01	−0.68[b]	⋯			
V-Lead	0.65[b]	0.82[b]	0.47[c]	0.41[c]	−0.47[b]	⋯		
Nonspeech								
A-Lead	0.37[c]	−0.12	0.67[b]	−0.47[c]	0.37[c]	0.15	⋯	
V-Lead	0.35	0.47[b]	0.81[b]	0.68[b]	−0.31	0.51[b]	0.17	⋯

[a]Abbreviations as in Table II. "Derived" means derived from the MPS and FWHM, which were the "primary" measures obtained directly from the curve-fitting procedure.
[b]$p<0.01$.
[c]$p<0.05$.

speech perception scores and AV synchrony detection. However, the present results are consistent on the whole with previous findings reported in the literature that AV synchrony perception for speech is correlated with some measures of AV speech perception ability (Grant and Seitz; McGrath and Summerfield, 1985). The present study used TRS to degrade the auditory signal, rather than more ecologically valid degradations such as noise that were used in previous studies. The TRS manipulation may have encouraged different strategies of integration than are typically employed when perceiving AV speech, which could potentially have strengthened the association between speech perception scores and the synchrony detection for nonspeech signals. Although this issue deserves further investigation, the present results for synchrony detection in speech signals are quite similar to previous results obtained with other forms of auditory degradation. In addition, the similarity of synchrony windows for our speech and nonspeech signals, which is independent of the auditory degradation used for the sentences, suggests that our findings linking speech perception scores with synchrony detection in nonspeech signals would hold for other forms of auditory degradation as well.

The present findings suggest new avenues of inquiry into the relationship between AV synchrony perception for nonspeech signals and AV speech perception abilities. Specifically, because AV speech perception ability is related to AV synchrony detection for both speech and nonspeech signals, it may be necessary to further examine variability in perceptual and cognitive mechanisms that are not specific to speech—such as sensory processing, perceptual learning, and auditory and visual attention—for further insights into individual differences in AV speech perception ability. In addition, future work examining the neural bases of the temporal integration of AV speech may also want to consider including nonspeech conditions to analyze the contribution of neural areas or mechanisms not specific to AV speech integration. Finally, the directionality of the link between explicit AV synchrony detection and speech perception should be further explored to determine for theoretical and clinical reasons whether explicit training in either synchrony detection or speech perception under degraded conditions affects performance on the other task.

Although some measures of AV synchrony detection for both speech and nonspeech signals were significantly correlated with AV speech perception measures, the positive correlation of A-only and AV scores with auditory-leading thresholds that were closer to physical synchrony was observed only for the speech and not the nonspeech signals. This result suggests that subjects who performed better on AV sentence intelligibility measures were more accurate at identifying the asynchrony in auditory-leading speech, but not in auditory-leading nonspeech signals. Long-term perceptual learning and experience may underlie this finding. A recent report suggests that after adaptation to asynchronous auditory- or visual-leading signals, AV synchrony judgments for simple nonspeech stimuli are recalibrated in the direction of the adapted asynchrony (Fujisaki *et al.*, 2004). Perhaps consistent exposure to speech signals, in which visual articulatory motion often precedes the production of sound, causes long-term adaptation effects so that participants who are better at integrating auditory and visual information in speech perception are also able to more easily recognize the unnatural situation in which auditory speech precedes visual speech. On the other hand, the lack of expectations about which signal should lead in the nonspeech signal condition may have prevented the use of naturally occurring cues in that condition.

In conclusion, the positive correlations between AV synchrony detection measures for speech and nonspeech signals suggest that the asynchrony detection for the two types of signals may rely on some common properties of multisensory processing. Furthermore, the finding that accuracy at detecting AV synchrony in unfamiliar, semantically meaningless nonspeech signals is positively correlated with AV speech perception measures suggests that variation in AV speech perception abilities may be affected, in part, by variation in domain-general multisensory processing. In future studies it will be necessary to examine in greater detail the relationship between general multisensory processing and multisensory processing for speech signals, including the extent to which they engage shared neural mechanisms.

ACKNOWLEDGMENTS

This work was supported by NIH NIDCD R01 Research Grant No. DC00111 and NIH NIDCD T32 Training Grant No. DC00012. The first author was also supported by an Indiana University Chancellor's Fellowship and an NSF Graduate Research Fellowship. The authors would like to thank Luis Hernandez for technical support and Sara Phillips and Angelique Horace for help with data collection. We also thank Ken Grant and the two anonymous reviewers for their helpful comments. We gratefully acknowledge the audience at the 2003 Audio-Visual Speech Processing workshop for their valuable suggestions on a preliminary version of this work.

[1]It was not of interest here to determine whether the speech and nonspeech synchrony windows had precisely the same parameters, because these can be expected to vary somewhat based on the particular stimuli employed; rather, the relationships among the parameters were examined here.

[2]An earlier study (Conrey and Pisoni, 2003) reported smaller synchrony windows for words with high versus low visual-only intelligibility; subsequent and more detailed analyses, however, indicated large item-by-item differences that were not always related to visual-only intelligibility.

Boothroyd, A., Hanin, L., and Hnath, T. (**1985**). "A sentence test of speech perception: Reliability, set equivalence, and short term learning," Speech and Hearing Science Report No. RC110, City University of New York.

Bushara, K. O., Grafman, J., and Hallett, M. (**2001**). "Neural correlates of auditory-visual stimulus onset asynchrony detection," J. Neurosci. **21**(1), 300–304.

Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). "PsyScope: A new graphic interactive environment for designing psychology experiments," Behav. Res. Methods Instrum. Comput. **25**(2), 257–271.

Conrey, B., and Pisoni, D. B. (**2003**). "Audiovisual asynchrony detection for speech and nonspeech signals," paper presented at the ISCA 2003 International Conference on Auditory-Visual Speech Processing.

Dixon, N., and Spitz, L. (**1980**). "The detection of audiovisual desynchrony," Perception **9**, 719–721.

Fujisaki, W., Shinsuke, S., Kashino, M., and Nishida, S. (**2004**). "Recalibration of audiovisual simultaneity," Nat. Neurosci. **7**, 773–778.

Grant, K. W., and Greenberg, S. (**2001**). "Speech intelligibility derived from

asynchronous processing of auditory-visual information," paper presented at the ISCA International Conference on Auditory-Visual Speech Processing.
Grant, K. W., and Seitz, P. F. (**1998**). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438–2450.
Grant, K. W., van Wassenhove, V., and Poeppel, D. (**2003**). "Discrimination of auditory-visual synchrony," paper presented at the ISCA 2003 International Conference on Auditory-Visual Speech Processing.
King, A. J., and Palmer, A. R. (**1985**). "Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus," Exp. Brain Res. **60**, 492–500.
Lachs, L. (**1999**). "Use of partial stimulus information in spoken word recognition without auditory stimulation," in Research on Spoken Language Processing Progress Report No. 23, Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 82–114.
Lachs, L., and Hernandez, L. R. (**1998**). "Update: The Hoosier Audiovisual Multitalker Database," in Research on Spoken Language Processing Progress Report No. 22, Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 377–388.
Lachs, L., and Pisoni, D. B. (**2004a**). "Crossmodal source identification in speech perception," Ecological Psychol. 16, 159–187.
Lachs, L., and Pisoni, D. B. (**2004b**). "Crossmodal source information and spoken word recognition," J. Exp. Psychol. Hum. Percept. Perform. **30**(2), 378–396.
Lewkowicz, D. J. (**1996**). "Perception of auditory-visual temporal synchrony in human infants," J. Exp. Psychol. Hum. Percept. Perform. **22**, 1094–1106.
Massaro, D., and Cohen, M. (**1993**). "Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables," Speech Commun. **13**, 127–134.
Massaro, D., Cohen, M. M., and Smeele, P. M. T. (**1996**). "Perception of asynchronous and conflicting visual and auditory speech," J. Acoust. Soc. Am. **100**(3), 1777–1786.
McGrath, M., and Summerfield, Q. (**1985**). "Intermodal timing relations and audio-visual speech recognition by normal-hearing adults," J. Acoust. Soc. Am. **77**(2), 678–684.
Meredith, M. A. (**2002**). "On the neuronal basis for multisensory convergence: A brief overview," Cognit. Brain Res. **14**, 31–40.
Meredith, M. A., Nemitz, J. W., and Stein, B. E. (**1987**). "Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors," J. Neurosci. **7**(10), 3215–3229.
Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (**1996**). "Temporal constraints on the McGurk effect," Percept. Psychophys. **58**(3), 351–362.
Pandey, C. P., Kunov, H., and Abel, M. S. (**1986**). "Disruptive effects of auditory signal delay on speech perception with lip-reading," J. Aud. Res. **26**, 27–41.
Saberi, K., and Perrott, D. R. (**1999**). "Cognitive restoration of reversed speech," Nature (London) **398**, 760.
Sheffert, S. M., Lachs, L., and Hernandez, L. R. (**1996**). "The Hoosier Audiovisual Multitalker Database," in Research on Spoken Language Processing Progress Report No. 21, Speech Research Laboratory, Indiana University, Bloomington, IN, pp. 578–583.
Stein, B., and Meredith, M. A. (**1993**). *The Merging of the Senses* (MIT, Cambridge, MA).
Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., *et al.* (**2001**). "When is now? Perception of simultaneity," Proc. R. Soc. London, Ser. B **268**, 31–38.
van Wassenhove, V., Grant, K. W., and Poeppel, D. (**2002**). "Temporal integration in the McGurk effect," poster presented at the annual meeting of the Society for Cognitive Neuroscience, San Francisco.

Measuring the off-axis angle and the rotational movements of phonating sperm whales using a single hydrophone

Christophe Laplanche,[a] Olivier Adam, Maciej Lopatka, and Jean-François Motsch
Laboratoire Images, Signaux et Systèmes Intelligents, Groupe Ingénierie des Signaux Neuro-Sensoriels, Université Paris XII, France

(Received 20 July 2005; revised 11 February 2006; accepted 14 February 2006)

The common use of the bent-horn model of the sperm whale sound generator describes sperm whale clicks as the pulse series $\{p0,p1,p2,p3,\ldots\}$. Clicks, however, deviate from this standard when recorded using off-axis hydrophones. The existence of additional pulses within the $\{p0,p1,p2,p3,\ldots\}$ series can be explained still using the bent-horn model. Multiple reflections on the whale's frontal and distal sacs of the $p0$ pulse lead to additional sets of pulses detectable using a farfield, off-axis hydrophone. The travel times of some of these additional pulses depend on the whale's orientation. The authors propose a method to estimate the off-axis angle of sperm whale clicks. They also propose a method to determine the nature of the movement (if it is pitch, yaw, or roll) of phonating sperm whales. The application of both methods requires the measurement of the travel time differences between pulses composing a sperm whale click. They lead, using a simple apparatus consisting of a single hydrophone at an unknown depth, to new measurements of the underwater movements of sperm whales. Using these methods shows that sperm whales would methodically scan seawater while searching for prey, by making periodic pitch and yaw movements in sync with their acoustic activity. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2184987]

PACS number(s): 43.80.Ka, 43.30.Sf, 43.66.Qp [WWA] Pages: 4074–4082

I. INTRODUCTION

Sperm whales (*Physeter macrocephalus*) mostly feed on meso- and bathypelagic squid and fish of various sizes and physical constitution (Kawakami, 1980; Whitehead, 2003). A 15-ton adult sperm whale would need to catch around half a ton of pelagic food in a day, assuming a daily consumption of 3% of its body weight (Lockyer, 1981). Sperm whales spend most of their time underwater and undertake long, deep dives to catch such an amount of prey (Miller *et al.*, 2004; Watkins *et al.*, 1999). However, the low-light, high-pressure conditions reigning at these great depths make difficult the visual observation of foraging sperm whales. Consequently, one does not know precisely the technique(s) sperm whales would use to successfully seek and catch their food.

Nevertheless, scientists have developed many methods of studying the underwater behavior of deep-diving marine mammals (e.g., sperm whales), notably electronic tagging and passive acoustics. By using such methods, scientists have been collecting over the last decades many clues regarding the foraging behavior of sperm whales.

Sperm whales have a sound generator located in their snout (Madsen *et al.*, 2002). The snout represents one quarter to one third the length of the whale (Clarke, 1978b; Nishiwaki *et al.*, 1963), and the whale uses it primarily for biosonar purposes (Jaquet *et al.*, 2001); Møhl *et al.*, 2000; Whitehead and Weilgart, 1990). Sperm whales are indeed the most acoustically active toothed whales. They emit a series of transient sounds (*clicks*) when undertaking foraging dives (Backus and Schevill, 1966). Such clicks would be of a duration, frequency content, source level, and directionality suited to echolocate their prey (Møhl *et al.*, 2003, 2000).

Sperm whale dives are composed of a vertical descent, a hunt in a prey layer at depth, and a reascent (Miller *et al.*, 2004). Sperm whales emit a long series of clicks at a slow rate (*usual clicks*, $0.45 \leqslant \mathrm{ICI} < 2$ s, ICI denoting the *inter-click interval*) during the descent. At depth, sperm whales emit usual clicks followed by clicks of increasing rate (*creaks*, $\mathrm{ICI} < 0.45$ s) and/or silences (Mullins *et al.*, 1988; Zimmer *et al.*, 2005b). It has been advanced that sperm whales emit usual clicks when searching for prey (denoted *searching phase*) and creaks when closing in on prey (denoted *approach* and *terminal phases*). They repeat this searching/approach/terminal pattern at depth (Laplanche *et al.*, 2005; Madsen *et al.*, 2005). According to Miller *et al.* (2004), sperm whales would keep crossing the prey layer during the searching phase and actively swim during the approach/terminal phases.

The sperm whale movements and precise use of their biosonar during these searching/approach/terminal phases are not perfectly understood yet. Laplanche *et al.* (2005) have proposed a model describing the sperm whale vertical movements and biosonar use during the searching phase. This model is, however, only partially correct and the authors offer some clarifications below. They propose a passive acoustic technique, requiring a single hydrophone, to determine the nature of the movements of diving sperm whales (i.e., if it is a pitch, yaw, or roll movement). They also present a new technique to estimate the off-axis angle of sperm whale clicks. The application of both techniques is grounded on a close description of the waveform of sperm

[a]Electronic mail: laplanche@univ-paris12.fr

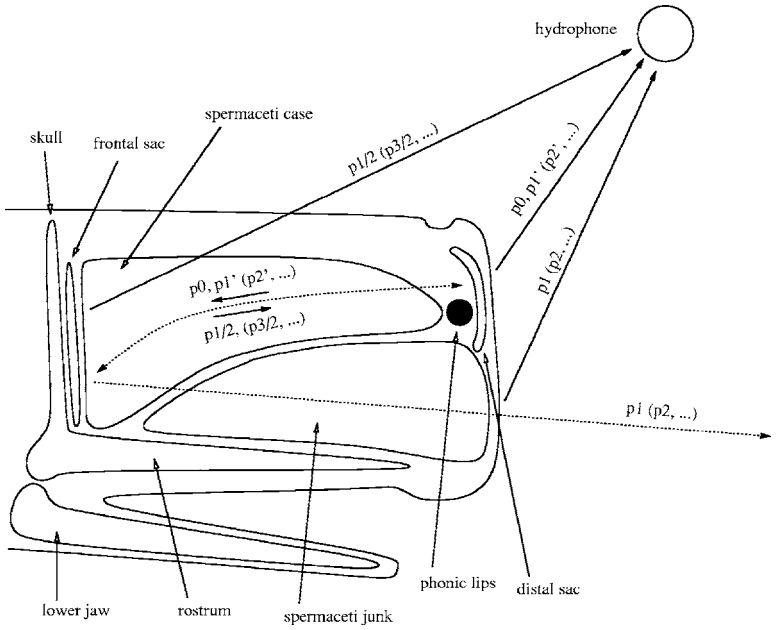


FIG. 1. The initial pulse $p0$ is generated by the phonic lips, transmits through the spermaceti case, and reflects on the frontal sac into the $p1/2$ pulse. The $p1/2$ pulse transmits (1) through the junk and leads to the echolocation pulse $p1$, (2) back through the spermaceti case, reflects on the distal sac, and leads to the distal pulse $p1'$, and (3) through the whale's body to the receiver. The pi' pulses ($i \geq 1$) then recursively lead to the $p(2i+1)/2$, $p(i+1)$, and $p(i+1)'$ pulses.

whale clicks, as predicted by the standard model of the sperm whale sound generator. The authors show that sperm whales synchronize their movements with their clicking activity. The authors then propose a more accurate model describing how sperm whales move and use their biosonar while searching for prey.

II. METHODS

A. The leaky bent-horn model

Today's hypothesis is that sperm whales generate their clicks with their snout (Norris and Harvey, 1972) according to the *bent-horn model*. This model, proposed by Møhl *et al.* (2003) and later confirmed by Zimmer *et al.* (2005b), successfully predicts the function and location of the various organs composing the sperm whale's snout. The sperm whale's snout is composed of the nasal circuitry (*left* and *right nares*), a connective organ (*phonic lips*), two air sacs (*distal* and *frontal sacs*), two spermaceti compartments (*spermaceti case* and *junk*), muscles, and blubber, as described by Clarke (1978b) and illustrated in Fig. 1. According to the bent-horn model, and in the notation of Møhl *et al.* (2003), clicks are generated by the phonic lips (leading to the $p0$ pulse), roughly directed backward by the distal sac, focused by the spermaceti case, and reflected by the frontal sac. The frontal reflected pulse (labeled $p1/2$) is channeled and projected forward by the spermaceti junk, leading to the outgoing $p1$ pulse.

The use of this model also predicts that the sperm whale click waveform is multipulsed, i.e., composed of a series of pulses, $\{p0, p1, p2, \ldots\}$ (Fig. 2, top). Indeed, the frontal reflection $p1/2$ travels back in the spermaceti case and reflects on the distal sac (leading to a $p1'$ pulse). The distal reflection $p1'$ follows an acoustic path inside the snout of the whale similar to the one followed by $p0$, leading to a frontal reflection $p3/2$, the junk pulse $p2$, and a distal reflection $p2'$, the distal reflection $p2'$ itself leading to the triplet $\{p5/2, p3, p3'\}$, and so on. The standard use of the bent-horn model predicts that the initial $p0$ and the junk $\{p1, p2, p3, \ldots\}$ pulses exit from the sperm whale's snout, explaining the observed multipulse structure of sperm whale clicks. The bent-horn model also successfully predicts the measured variations of the apparent source level of the $\{p0, p1\}$ pulses with the orientation of the hydrophone relative to the whale (Møhl *et al.*, 2003; Zimmer *et al.*, 2005b).

Using this model (or its Norris and Harvey ancestor), some researchers have described the $\{p0, p1, p2, \ldots\}$ series as regular, the time of arrival difference between successive pulses $\{pi, pi+1\}$ ($i \geq 0$) being a function of the whale's length (Goold, 1996; Gordon, 1991; Rhinelander and Dawson, 2004). However, the irregularity of the multipulse structure (i.e., variations of such time of arrival differences as well as the existence of additional pulses within the $\{p0, p1, p2, \ldots\}$ frame) has been underlined by different re-

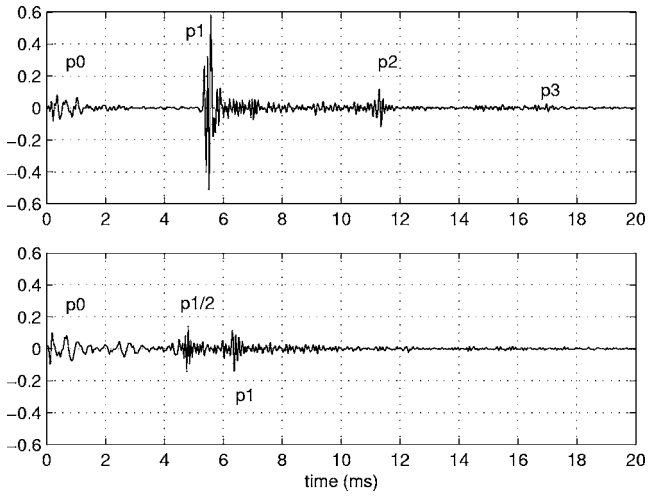


FIG. 2. Both clicks were emitted by the same sperm whale. Sperm whale clicks are often described as a series of regularly spaced pulses $\{p0, p1, p2, \ldots\}$ (top). Clicks recorded in the farfield rarely fit to this model and are composed of a more complicate set of pulses $\{p0, p1/2, p1, \ldots\}$ (bottom).

searchers [see Bahl *et al.* (2002); Goold (1996); Møhl *et al.* (2003) for instance] without being elucidated.

The standard multipulse series $\{p0,p1,p2,\ldots\}$ does not actually fully represent the waveform of sperm whale clicks. Clicks, when recorded off axis, deviate from this standard (Fig. 2, bottom). The authors propose a new and more complete description of the multipulse structure of sperm whale clicks, however, maintaining the standard bent-horn model, and describe what one could call the *leaky bent-horn model*. This description clarifies the discrepancies between the multipulse structure of clicks predicted by the standard model ($\{p0,p1,p2,\ldots\}$) and what is actually observed in practice. Different researchers have reached similar conclusions (Zimmer *et al.*, 2005a). The authors of the present work then do not reillustrate the soundness of the new model, as it has been convincingly done by Zimmer *et al.* (2005a). However, they briefly redefine it as a basis for the application described below. They also supplement the description of the sperm whale click multipulse structure proposed by Zimmer *et al.* (2005a), as it is not complete. For the sake of clarity, the authors of the present work use a notation similar to the notation that Zimmer *et al.* (2005a) introduced.

The standard use of the bent-horn model stipulates that only the initial and the junk pulses $\{p0,p1,p2,p3,\ldots\}$ exit from the sperm whale's snout. The leaky bent-horn model stipulates that (1) the above mentioned frontal $\{p1/2,p3/2,p5/2,\ldots\}$ and distal $\{p1',p2',p3',\ldots\}$ reflections also leak, and that (2) $\{p0,p1,p2,p3,\ldots\}$, $\{p1/2,p3/2,p5/2,\ldots\}$, and $\{p1',p2',p3',\ldots\}$ form distinct sets of pulses that can be separated using an off-axis hydrophone in the farfield. Zimmer *et al.* (2005a) demonstrated that the frontal $p1/2$ pulse indeed leaks from the sperm whale's snout and that the $p1$ pulse exits at the flat anterior surface of its junk. Zimmer *et al.* (2005a) also demonstrated that the travel time differences of the $p1/2$ and $p1$ pulses to the $p0$ pulse depend on the orientation of the whale to the hydrophone. The authors of the present work do not reillustrate the soundness of such statements. They supplement the description of the waveform of a sperm whale click proposed by Zimmer *et al.* (2005a), asserting that it is also composed of the distal reflection $p1'$ and the additional sets of pulses $\{p3/2,p5/2,\ldots\}$, $\{p2,p3\ldots\}$, and $\{p2',p3',\ldots\}$, as described above. Such a description is discussed below.

B. Definition of pulse delays and angles

The delay of the py pulse on the px pulse is noted τ_{xy}, simplified into τ_y if $px=p0$. The delays $\{\tau_{1'},\tau_{1'2'},\tau_{2'3'},\ldots\}$ are all equal to the travel time required by a pulse to travel twice between the distal and frontal sacs through the spermaceti case. Such a travel time is still denoted IPI by the authors, being the *interpulse interval* used in the literature regarding the sperm whale length estimation process mentioned above.

The delays $\{\tau_{1/2},\tau_{3/2},\ldots\}$ of the frontal pulses $\{p1/2,p3/2,\ldots\}$ to the initial $p0$ pulse, and the delays $\{\tau_{1'1},\tau_{2'2},\ldots\}$ of the junk pulses $\{p1,p2,p3,\ldots\}$ to the distal pulses $\{p1',p2',p3',\ldots\}$ depend on the orientation of the whale to the hydrophone, as proven by Zimmer *et al.* (2005a) regarding $\tau_{1/2}$ and $\tau_{1'1}$. Using a hydrophone at a fixed location in the farfield, $\{\tau_{1/2},\tau_{3/2},\ldots\}$ and $\{\tau_{1'1},\tau_{2'2},\ldots\}$, delays will change while the sperm whale moves underwater. Our aim in this work is to determine the nature of the whale's movements using the measurement of these delays. As described below, in the recordings used to illustrate this work, most of the time only pulses from the set $\{p0,p1/2,p1,p1'\}$ are clearly detected. In the study of the whale's movements that follows, the authors have considered such pulses and used the delays $\{\tau_{1/2},\mathrm{IPI},\tau_{1'1}\}$ only.

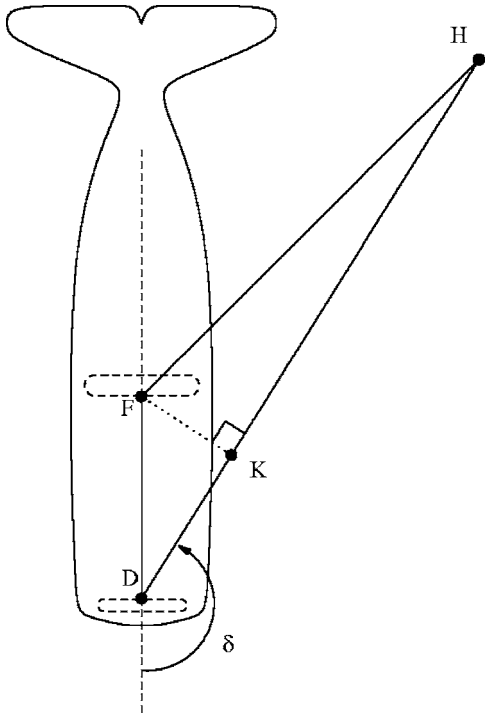


FIG. 3. The off-axis angle of the whale is defined as the angle between the whale dorsorostral axis and the line joining the whale to the hydrophone H. The point D is at the center of the phonic lips (point of emission of the $p0$ pulse). F is the point of frontal reflection $p0\rightarrow p1/2$. K is the orthogonal projection of F on the line (DH).

The authors consider in the following the reference frame of the whale. They define the pitch, yaw, and roll movements of the whale as rotations on its left-right, up-down, and dorsorostral axes. As a difference, Zimmer *et al.* (2005b) and Laplanche *et al.* (2005) have both considered the sperm whale movements in the terrestrial reference frame. The off-axis angle of the whale (labeled $\delta\in[0,\pi]$, and estimated when the whale is emitting a click) is defined as the angle separating the whale dorso-rostral axis to the line joining the whale to the hydrophone (Fig. 3).

C. Estimation of the off-axis angle of a click

The off-axis angle of a whale represents its orientation to a reference point (e.g., the hydrophone). This angle varies with the whale's underwater movements, and, inversely, the variations of the off-axis angle of a whale provide information on the whale's underwater movements. The knowledge of this angle is critical in the following study, and the authors propose a simple method to carry out the estimation of the off-axis angle of a sperm whale click.

The following demonstration applies for off-axis angles verifying $\delta \in [\pi/2, \pi]$ (Fig. 3). The demonstration is slightly different if $\delta \in [0, \pi/2]$, but leads to the same result, i.e., Eq. (5). The leaky bent-horn model predicts, as illustrated in Fig. 3, that the $p0$ pulse follows the path $[DKH]$, and the $p1/2$ pulse follows the path $[DFH]$. The delay $\tau_{1/2}$ of the pulse $p1/2$ on the pulse $p0$ is then equal to (τ_{XY} denoting the travel time of a pulse on the path $[XY]$)

$$\tau_{1/2} = [\tau_{DF} + \tau_{FH}] - [\tau_{DK} + \tau_{KH}]. \quad (1)$$

The $p0$ and $p1/2$ pulses travel through the whale's body and through seawater to reach the hydrophone H. Refraction takes place during this transmission (Flewellen and Morris, 1978; Urick, 1983). Refraction in seawater leads to negligible variations of the quantity $\tau_{FH} - \tau_{KH}$, assuming a hydrophone in the farfield, the path $[KH]$ and $[FH]$ being close to each other. In this case,

$$\tau_{FH} - \tau_{KH} \simeq 0. \quad (2)$$

In view of the previous statement IPI$=\tau_{1'}$, then $\tau_{DF}=$IPI$/2$ leads to

$$\tau_{1/2} \simeq \frac{\text{IPI}}{2} - \tau_{DK}. \quad (3)$$

The off-axis angle $\delta=(\vec{FD},\vec{DK})$ can be found from $\cos\delta \simeq -DK/FD$. One then assumes, given the dimension of the whale's body, that the curvature of the rays due to refraction in the whale's body lead to negligible variations on the delay τ_{DK}. Then, assuming that the $p0$ pulse travels along the path $[DK]$ in a straight line at the mean speed of sound $c_{sw}=c_{DK}$, and along the path $[DF]$ at the mean speed of sound c_s, then

$$\cos\delta \simeq -\frac{DK}{FD} \simeq -\frac{c_{sw}\tau_{DK}}{c_s \text{IPI}/2}. \quad (4)$$

Equations (3) and (4) lead to the following estimator of the off-axis angle of a click, as a function of the standard interpulse interval IPI and the delay $\tau_{1/2}$

$$\delta \simeq \arccos\left(\frac{c_{sw}}{c_s}\frac{2\tau_{1/2} - \text{IPI}}{\text{IPI}}\right). \quad (5)$$

This formula also applies for $\delta \in [0, \pi/2]$ (K being in that case the orthogonal projection of D on the line (FH) and choosing $c_{sw}=c_{FK}$ as the mean speed of sound on the path $[FK]$). The mean speed of sound c_{sw} is a function of the off-axis angle, $c_{sw}(\delta)$. The authors choose $c_{sw}=c_s$ for on-axis angles ($\delta \in [0, \pi/4]$ or $\delta \in [3\pi/4, \pi]$) and $c_{sw}=c_w$ (naming c_w the speed of sound in seawater) for off-axis angles ($\delta \in [\pi/4, 3\pi/4]$). Such an approximation is discussed below.

Equation (5) cannot be properly used to estimate the whale's off-axis angle if it is on-axis ($d\delta/d\tau_{1/2} \to +\infty$ if $\tau_{1/2} \to 0$ or $\tau_{1/2} \to$ IPI). A click with a normalized delay $[1 + (c_s/c_w)/\sqrt{2}]/2 \leqslant \tau_{1/2}/\text{IPI} \leqslant 1$ is on axis and the hydrophone is in front of the whale ($\delta \in [0, \pi/4]$). A click with a normalized delay $0 \leqslant \tau_{1/2}/\text{IPI} \leqslant [1 - (c_s/c_w)/\sqrt{2}]/2$ is on axis and the hydrophone is behind the whale ($\delta \in [3\pi/4, \pi]$). A click with a normalized delay $[1 - (c_s/c_w)/\sqrt{2}]/2 \leqslant \tau_{1/2}/\text{IPI} < [1 + (c_s/c_w)/\sqrt{2}]/2$ is off axis ($\delta \in [\pi/4, 3\pi/4]$), and its off-axis angle can be estimated using

$$\delta \simeq \arccos\left(\frac{c_w}{c_s}\frac{2\tau_{1/2} - \text{IPI}}{\text{IPI}}\right). \quad (6)$$

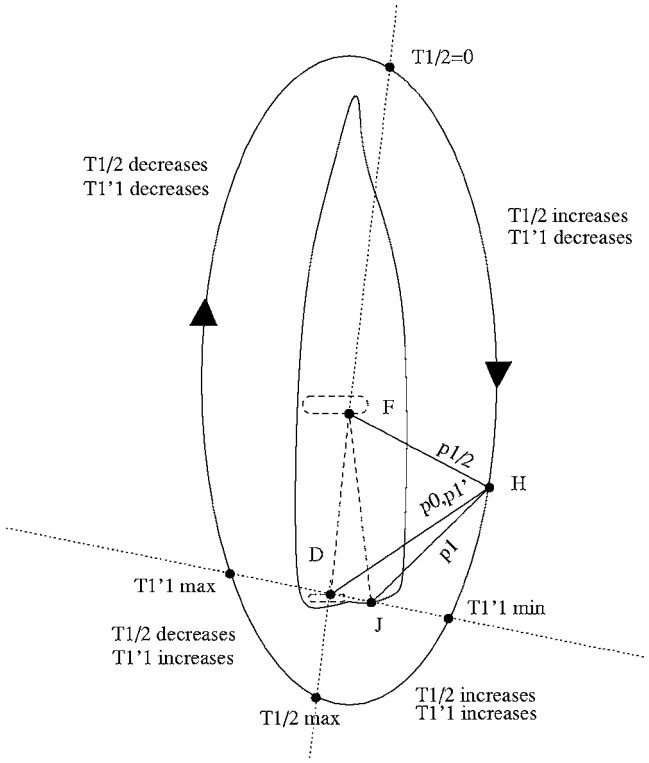


FIG. 4. $\tau_{1/2} \leqslant$ IPI$/2$ if the hydrophone is behind the whale, $\tau_{1/2} \geqslant$ IPI$/2$ if it is in front of the whale. $\tau_{1'1} \leqslant 0$ if the hydrophone is below the whale, $\tau_{1'1} \geqslant 0$ if it is above the whale. The sign of the variations of $\tau_{1/2}$ and $\tau_{1'1}$ during a pitch movement depend on the location of the hydrophone to the whale.

D. Travel time difference between $p1$ and $p1'$

The leaky bent-horn model also predicts, as illustrated in Fig. 4, that the $p1$ pulse follows the path $[DFJH]$ and the $p1'$ pulse follows the path $[DFDH]$ to reach the hydrophone H. The algebraic delay $\tau_{1'1}$ of the echolocation pulse $p1$ on the distal pulse $p1'$ also depends on the orientation of the whale to the hydrophone. As illustrated in Fig. 4 and demonstrated by Zimmer *et al.* (2005a),

$$\tau_{1'1} \in [-DJ/c_w, DJ/c_w], \quad (7)$$

where J is at the center of the flat anterior surface of the junk. By writing the length DJ and the length L_{so} of the spermaceti organ as functions of the whale's full length (noted L), $DJ=\alpha(L)L$, and $L_{so}=\beta(L)L$, approximating the length of the spermaceti organ as the length DF, and using $DF=c_s$IPI$/2$, then

$$\frac{DJ}{c_w} = \frac{\alpha(L)}{\beta(L)}\frac{c_s}{c_w}\frac{\text{IPI}}{2}. \quad (8)$$

The authors then assume that the height to length ratio of the whale's spermaceti organ is equal to the height (L_{hs}) to

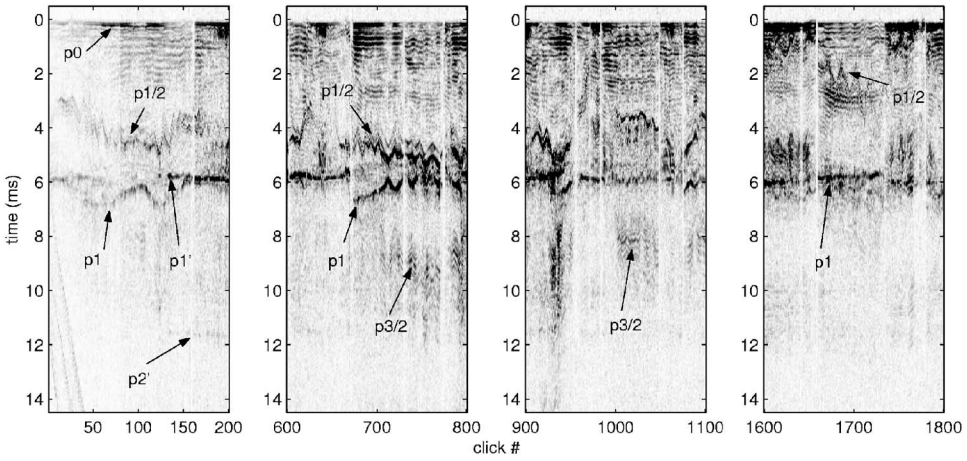


FIG. 5. The normalized, apparent level of 4×200 usual clicks emitted by a sperm whale during a single dive. Sea surface reflections are here visible from clicks #0→#40. A sperm whale click is composed of distal $\{p0,p1',p2',\ldots\}$, frontal $\{p1/2, p3/2,\ldots\}$, and junk $\{p1,p2,\ldots\}$ pulses.

length (L_{ls}) ratio of the whale's skull, such a hypothesis being discussed below. Using empirical relationships providing the height and length of the skull as functions of the whale's full length, $L_{hs}=\gamma(L)L$ and $L_{ls}=\zeta(L)L$, they find $DJ/L_{so}=\alpha(L)/\beta(L)=\gamma(L)/\zeta(L)$. Linear regressions of the results provided for large whales ($11\leqslant L\leqslant 17$ m) by Nishiwaki *et al.* (1963) lead to $\gamma(L)\simeq[12-L/15]/100$ and $\zeta(L)\simeq[16+10L/11]/100$, providing the value of the ratio

$$\frac{\alpha(L)}{\beta(L)}\simeq\frac{12-L/15}{16+10L/11}. \tag{9}$$

The delay $\tau_{1'1}$ indicates how much above or below the whale is situated to the hydrophone. It is positive if the whale swims its back toward the hydrophone and is negative if the whale swims its belly toward the hydrophone. It is equal to zero if the hydrophone is in the lateral plane of the whale.

E. Pitch, yaw, and roll movements

The off-axis angle indicates where the hydrophone is situated in relation to the whale. It is calculated from Eq. (5) using the measurement of the delay $\tau_{1/2}$. Assuming that the line (DF) is close to the dorsorostral axis of the whale, the delay $\tau_{1/2}$, and, consequently, the off-axis angle are constant while the whale makes a *roll movement*. On the other hand, $\tau_{1'1}$ oscillates during such a movement around its mean value 0 and between the extrema given in Eq. (7). Variations of $\tau_{1'1}$ with $\tau_{1/2}$ being constant would then indicate a pure roll movement.

However, $\tau_{1/2}$ and δ are expected to vary during a pitch, a yaw, or a combination of pitch and yaw movements. Assuming $(JD)\perp(DF)$, the delay $\tau_{1'1}$ is constant during a *yaw movement*, whereas $\tau_{1/2}$ oscillates around its mean value, IPI/2. Variations of $\tau_{1/2}$ with $\tau_{1'1}$ being constant would then indicate a pure yaw movement.

Simultaneous variations of $\tau_{1/2}$ and $\tau_{1'1}$ indicate either a combination of roll and yaw movements, or a pitch movement, or a combination of roll, yaw, and pitch movements. Whereas synchronized variations of $\tau_{1/2}$ and $\tau_{1'1}$ would be the result of a pure *pitch movement* (Fig. 4). Such variations are similar (both increasing or both decreasing) if the hydrophone is located in front of and below the whale or behind and above the whale. Such variations are opposite (one increasing, one decreasing) if the hydrophone is located in front of and above the whale or behind and below the whale. By studying the variations of the delays $\tau_{1/2}$ and $\tau_{1'1}$ of a click sequence, one can qualitatively determine the nature of the whale's movements while it emits this click sequence.

III. RESULTS

A. Multipulse structure

A diagram illustrating the variations in the multipulse structure of the clicks emitted by a sperm whale during a single dive is provided in Fig. 5. Four sequences of 200 successive clicks are plotted, using the 2004 dataset presented in Laplanche *et al.* (2005). Only usual clicks (with ICI$>$0.45 s) are considered.

The $p0$ pulse is clearly detected through the dive, given its low directionality (Zimmer *et al.*, 2005b). The $p1$, $p1/2$, and $p1'$ pulses are also well detected. As predicted by the leaky bent-horn model, the $p1/2$ and $p1$ pulses are best detected if the hydrophone is in front of the whale ($\delta\in[0,\pi/2]\Leftrightarrow\tau_{1/2}\in[\mathrm{IPI}/2,\mathrm{IPI}]$), the $p1'$ pulse is best detected if the hydrophone is behind the whale ($\delta\in[\pi/2,\pi]$). The time of arrival of the $p1$ and $p1/2$ pulses fits with the model, as shown later using Eqs. (6) and (7), and clearly demonstrated by Zimmer *et al.* (2005a). The $p1'$ pulse is distinct with the $p1$ pulse, as unequivocally illustrated in the first sequence of Fig. 5. As predicted by the bent-horn model, the $p2'$ pulse is noticeable only if $p1'$ is detected (e.g., in the first sequence of Fig. 5). The $p3/2$ pulse is also detectable (e.g., in the third sequence of Fig. 5). Its travel time does not fit, however, with the model, as discussed later.

B. Pulse delays and off-axis angle

The travel time differences between the $\{p0,p1',p2'\}$ pulses are constant, as predicted by the model. The interpulse interval of the whale that clicks, plotted in Fig. 2 and Fig. 5, is IPI=5.8 ms.

The application of Eqs. (6) and (7) requires the measurement of the whale's length L, the speed of sound c_s in the spermaceti case, and the speed of sound c_w in sea water at the recording location. The whale is of length $L=12.8$ m, using the empirical formula proposed by Rhinelander and Dawson (2004) with IPI=5.8 ms. The authors choose in the following a speed of sound $c_s=1370$ m s^{-1} in the spermaceti

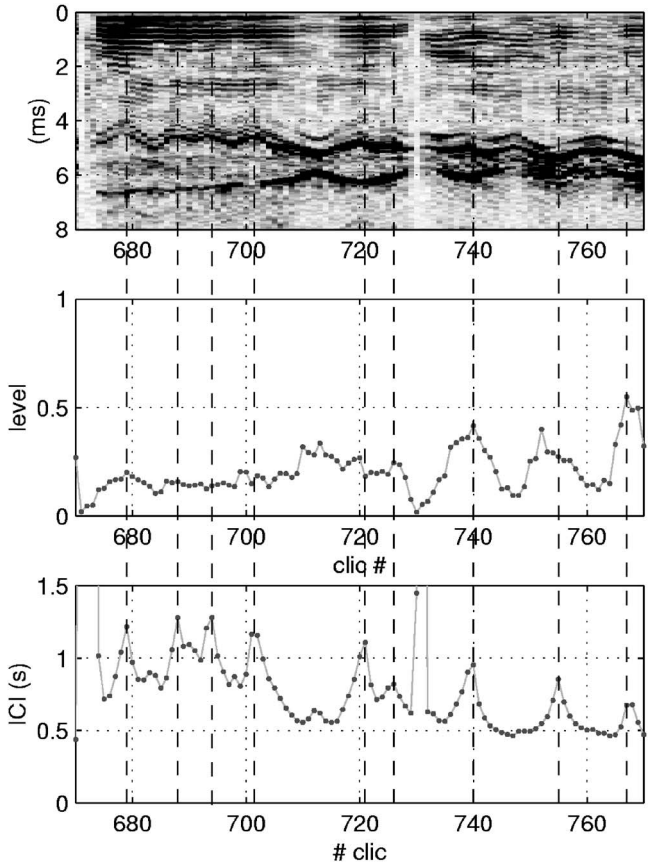


FIG. 6. The whale creates a fast yaw movement during the click sequence #670→#730 and a fast pitch movement during the click sequence #730 →#770. The ICI is in sync with the movement during the first sequence and in sync with the click apparent level during the second sequence.

organ, using the results presented by Flewellen and Morris (1978) for a temperature of 33 °C and a pressure of 1 atm. Such a choice is discussed below. The speed of sound in seawater at the recording location was $c_w \simeq 1500$ m s^{-1}.

When applying Eq. (6), clicks are on axis and the hydrophone is in front of the whale ($\delta \in [0, \pi/4]$) if 0.82 IPI $\leqslant \tau_{1/2} \leqslant$ IPI $\Leftrightarrow 4.8 \leqslant \tau_{1/2} \leqslant 5.8$ ms and behind the whale ($\delta \in [3\pi/4, \pi]$) if $0 \leqslant \tau_{1/2} \leqslant 0.18$ IPI $\Leftrightarrow 0 \leqslant \tau_{1/2} \leqslant 1$ ms. Most of the clicks of the sequences of Fig. 5 are off axis ($\delta \in [\pi/4, 3\pi/4]$, $1 \leqslant \tau_{1/2} \leqslant 4.8$ ms). Some clicks are almost on axis, for instance, click #750, which is $\pi/6$ off axis ($\tau_{1/2} = 5.5$ ms), using (5) with $c_{sw} = c_s$.

Using Eqs. (8) and (9), the delay of $p1$ on $p1'$ should verify $|\tau_{1'1}| \leqslant 0.18$ IPI $\simeq 1.05$ ms. Cranford (1999) suggested that the distance between the blow hole and the eye is a good estimate for the length of the spermaceti organ. Linear regressions of the results provided by Nishiwaki *et al.* (1963) then lead to a tip of snout to center eye distance close to $[15+26L/15]L/100$ m, a tip of snout to blowhole distance close to $[3.5+4.5L/15]L/100$ m, hence a blowhole to center eye distance $\beta(L)L \simeq [11.5+21.5L/15]L/100$ m. This leads to the estimate of the distance between the phonic lips and the location of point J on the anterior surface of the junk $DJ = \alpha(L)L = 0.12L = 1.54$ m. Zimmer *et al.* (2005a) has found $DJ = 1.2$ m for a sperm whale of length $L = 12$ m. Given the similar lengths of the whales observed in both works, one should expect $\alpha(L) = 0.10$ for the 12.8 m whale observed in the present work. This discrepancy comes from the fact that the height to length ratio of the whale's spermaceti organ is actually inferior to the height to length ratio of the whale's skull. This should be expected, in view of the relative importance of the maxillonasalis muscles in the constitution of the whale's snout (Clarke, 1978b). One should correct previous statements by writing $\alpha(L)/\beta(L) = [\gamma(L)/\zeta(L)]/\eta$ with $\eta = 1.2$. In that case, the delay of $p1$ on $p1'$ should verify $|\tau_{1'1}| \leqslant 0.15$ IPI $\simeq 0.9$ ms.

C. Movements

In the example presented in Fig. 5, the whale vertically dives to 400 m from click #1 to click #420 (Laplanche *et al.*, 2005) (only clicks #1 to #200 of this descent are represented in Fig. 5). During this descent, both $\tau_{1/2}$ and $\tau_{1'1}$ delays oscillate, as the consequence of a physical movement of the whale. The whale barely rolls during the descent, the delay $\tau_{1'1}$ being always positive. The whale creates a slow pitch movement, resulting in the synchronous, slow variations of $\tau_{1/2}$ and $\tau_{1'1}$. Variations of $\tau_{1'1}$ during this pitch movement are in sync with the vertical movement described in Laplanche *et al.* (2005). Maxima of the delay $\tau_{1'1}$ (at $t \in \{282, 415, 560\}$ s, clicks #{125, 245, 380}) are detected during the inflections of the vertical movement of the whale [at $t \in \{280, 400, 570\}$ s, Fig. 5 of Laplanche *et al.* (2005)]. The slow pitch movement found using the present method and the vertical movement described in Laplanche *et al.* (2005) must represent the same slow, vertical, pitch movement. The whale also creates a repetitive and faster yaw movement (10-click period variations of $\tau_{1/2}$ with $\tau_{1'1}$ constant) during the descent.

Sperm whales also move actively when emitting usual clicks at depth. The variations of $\tau_{1/2}$ and $\tau_{1'1}$ being constant during the click sequence #670→#730 would result of a periodic yaw movement. Synchronous variations of $\tau_{1/2}$ and $\tau_{1'1}$ during the click sequence #730→#770 would result of a periodic pitch movement. The former movement is in sync with the click rhythm, as illustrated in Fig. 6. The authors suggest that the whale methodically scans the content of expanses of water during the descent and at depth, as discussed later.

IV. DISCUSSION

A. Multipulse structure

The description of the multipulse structure proposed by the authors and Zimmer *et al.* (2005a) using the bent-horn model correctly predicts the existence, the time of arrival differences, and the relative levels of the pulses $\{p0, p1/2, p1, p1', p2'\}$. The model also predicts the existence of the $p3/2$ pulse. The travel time difference of $p3/2$ to $p0$ is, however, shorter than expected. During the click sequence #1000→#1025, one can measure $\tau_{3/2} \simeq 8.1$ ms (Fig. 5), though one should expect $\tau_{3/2} = \text{IPI} + \tau_{1/2} \simeq 9.4$ ms. Whereas the delays $\tau_{1/2}$ and $\tau_{3/2}$ slightly change during this click sequence, the travel time difference $\tau_{3/2} - \tau_{1/2}$ of $p3/2$ to $p1/2$ remains constant. This suggest that the $p3/2$ pulse is indeed the result of a multipath propagation inside the head of the whale and not the reflection by objects surrounding the

whale. The travel time difference $\tau_{3/2}$ is, however, 1.3 ms lower than predicted by the model. According to the benthorn model, the $p0$ pulse travels through the spermaceti case and reflects on the frontal sac ($p1/2$ pulse). The $p1/2$ pulse leaks into the spermaceti junk (leading to the $p1$ pulse) and case (leading to the $p1'$ pulse). This statement has lead the authors to speculate that the $p1/2$ pulse might also leak in sea water. The distal reflection $p1'$, when clearly detected, is followed by an IPI-delayed pulse (the $p2'$ pulse). This has lead to the assumption that $p1'$ might follow an acoustic path inside the whale's snout similar to the one followed by $p0$. The $p1'$ pulse would then reflect on the frontal sac, and, as $p0$ did, lead to a frontal reflection leaking in seawater: the $p3/2$ pulse. The discrepancy underlined above regarding the value of the delay $\tau_{3/2}$ might be the consequence of the inexactitude of this model. The authors did not though manage to explain the reason of such discrepancy.

The $p1/2$ pulse has been defined as the reflection of $p0$ on the frontal sac. The $p1/2$ pulse can, however, be detected using a hydrophone placed behind the whale (see, for instance, the fourth click sequence of Fig. 5). In such cases, the transmission of the $p1/2$ pulse would reach and cross the frontal sac, before leaking in the sea water through the whale's thorax. The $p1/2$ pulse has been described by Zimmer *et al.* (2005a) as arriving as scattered subpulses. Zimmer *et al.* (2005a) have suggested that this could be the result of multiple reflections on the frontal sac. The $p1/2$ pulse is also, regarding the dataset used through this work, composed of subpulses when recorded in front of the whale (Fig. 6). It is, however, not time spread when detected from behind the whale. This would rather suggest that the $p0$ pulse reaches the frontal sac at a single point. This does not forbid multipath propagation of $p0$ within the spermaceti case, as it is expected to occur as a result of the heterogeneous temperature and chemical composition of the case (Flewellen and Morris, 1978). Leakage of the $p0$ pulse through the frontal sac might reflect on the whale's lungs. Such reflections could result, using a hydrophone in front of the whale, of $p1/2$ arriving as scattered subpulses. One should also expect, assuming multiple reflections on the frontal sac resulting from multipath propagation within the spermaceti case, $p1/2$ to form a single broad spread pulse. Subpulses are, however, clearly identifiable (Fig. 6), bearing out the hypothesis describing $p1/2$ as the result of multiple reflections on distinct surfaces. Such a hypothesis, however, does not take into account lung collapsing with pressure at depth, and does not explain why the sperm whale's frontal sac is large and parabolic. Such size and geometry might though indicate a double function of the frontal sac (e.g., an acoustic receptor, as suggested by Kozak *et al.*).

B. Off-axis angle

The off-axis angle is estimated using Eq. (6). The authors have assumed that (1) the speed of sound along the path $[DF]$ is constant through the dive and of known value c_s, (2) the path $[DF]$ is rectilinear and parallel to the click axis, (3) the speed c_{sw} along the path $[DK]$ (if $\delta \in [\pi/2, \pi]$) or $[FK]$ (if $\delta \in [0, \pi/2]$) is equal to the speed of sound in seawater c_w. The speed of sound in the core of the spermaceti case is expected to change through the dive, as a consequence of variations in temperature and pressure, $c_s(T,P)$ (Goold, 1996). Such variations could be the result of depth-dependent variations of the thermodynamical properties of the sea water or the result of whale-controlled changes within the snout (thermoregulation or the use of the maxillonasalis muscles). The sperm whale is expected to control the temperature of its spermaceti case if it uses it to efficiently focus sounds (Flewellen and Morris, 1978), though it does not seem to have the required extensive vascularization system (Cranford, 1999). Anyone knows the precise variations of the temperature of the spermaceti case and junk through the dive. Gordon (1991) have considered $c_s = 1350\ \mathrm{m\ s^{-1}}$, Goold (1996) $c_s = 1430\ \mathrm{m\ s^{-1}}$. The author's choice of $c_s = 1370\ \mathrm{m\ s^{-1}}$ is here arbitrary, and has been chosen as the speed of sound in spermaceti at a temperature of 33 °C (Clarke, 1978a) and at the sea surface level.

In light of such spatial and temporal variations $c_s(x,y,z,t)$ within the spermaceti case, the path $[DF]$ is likely not rectilinear. The speed of sound c_s used in Eq. (6) is not the mean speed of sound c_{sc} in the core of the spermaceti case along the curved path $[DF]$, but the speed required for an acoustic wave to travel rectilinearly between the points D and F with the travel time IPI/2. Assuming that the $p0$ pulse reflects into $p1/2$ at the center of the frontal sac, the line (FD) is inclined relatively to the axis of the rostrum. The latter is, however, close to the axis of the junk. The speed of sound c_{sw} along the path $[DK]$ or $[FK]$ is actually a function of the off-axis angle, $c_{\mathrm{sw}}(\delta)$. The authors have assumed $c_{\mathrm{sw}}(\delta) = c_s$ if $|\delta - \pi/2| \geqslant \pi/4$ and $c_{\mathrm{sw}}(\delta) = c_w$ if $\pi/4 \leqslant \delta \leqslant 3\pi/4$. The knowledge of the spatial and temporal variations of the temperature within the spermaceti organ could resolve all the above approximations, leading to a precise measurement of (1) the speed of sound c_s, (2) the location of the reflection point F, (3) the empirical law $c_{\mathrm{sw}}(\delta)$.

The off-axis angle estimator proposed by the authors in Eq. (6) is then not accurate. Nevertheless, it is a beginning, and this work is worth being pursued. The applications of an off-axis angle estimator are various. One of them could allow to reconstruct the beam pattern of sperm whale clicks using a single hydrophone. The simultaneous measurements from different locations of the off-axis angle of a sperm whale could also be used to find the three-dimensional orientation (in the terrestrial reference frame) of this whale.

C. Rotational movements

Laplanche *et al.* (2005) have presented a model describing how sperm whales might move and echolocate while searching for prey. This model is partially correct. It is grounded on the hypothesis that sperm whales control click level and rhythm during echolocation. Such a hypothesis has been made in light of the apparent synchronization of click level and ICI (for instance, in Fig. 7 of this reference). Nevertheless, such synchronization could result when the whale synchronizes its click rhythm to an oscillatory movement, and when the observer's position is *less than ideal.* Laplanche *et al.* (2005) rejected such a hypothesis, thinking

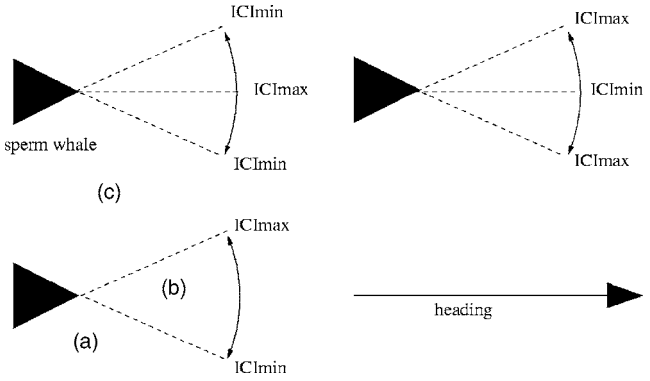


FIG. 7. Three models describing the fast, periodic movement of the whale in sync with its ICI. Only the third model could lead to the synchronization $ICI_{max} \Leftrightarrow \delta_{max}$ using a hydrophone located either in front of or behind the whale [it is located in (a)]. The click apparent level and the ICI are in sync ($ICI_{max} \Leftrightarrow AL_{max}$) if the hydrophone is located in (c). There is no apparent synchronization between the ICI, the off-axis angle, and the click AL if the hydrophone is in sector (b).

that an ICI-movement synchronization would lead to an ICI-level synchronization only for on-axis recordings. Actually, the lateral asymmetry in the acoustic behavior of sperm whales while searching for prey described later could explain why click level and rhythm can so often be in apparent sync.

As illustrated regarding the click sequence #670 $\rightarrow$#730 (Fig. 6), the sperm whale clicking activity can be in sync with the fast, periodic movement it creates (in this case a yaw movement). In this example, the hydrophone is in front of the whale ($\delta \leqslant \pi/2$) and the maxima of the ICI correspond to the maxima of the off-axis angle. A similar relationship (i.e., $ICI_{max} \Leftrightarrow \delta_{max}$) can be observed using a hydrophone behind the whale ($\delta \geqslant \pi/2$).

The authors think of three models describing the fast, periodic, rotational movement of the whale in sync with its ICI (Fig. 7). Such a rotational movement is assumed to be planar in the following. The first model stipulates that the whale echolocates at slow rate (ICI_{max}) in front of it and faster (ICI_{min}) on its sides. The second model stipulates that the whale echolocates at high rate (ICI_{min}) in front of it and slower (ICI_{max}) on its sides. The third model stipulates that the whale echolocates at slow rate (ICI_{max}) on one side and faster (ICI_{min}) on the other.

One can show that during a periodic, planar, rotational movement, the off-axis angle reaches its extreme, either when the hydrophone is in the plane containing the whale dorsorostral axis and the rotational axis or during the inflections of the movement. Whereas the synchronization of the ICI and click level may appear approximate (e.g., during the click sequence #730$\rightarrow$#770 of Fig. 6), the synchronization of the ICI and the off-axis angle is unequivocal (e.g., during the click sequence #670$\rightarrow$#730). The authors, whereas they should have not so easily rejected the hypothesis that the ICI-apparent level synchronization being the result of the hydrophone being in a *less than ideal* location [Laplanche *et al.* (2005)], here they reject the hypothesis that the ICI-δ synchronization mentioned above is apparent only. They assume in the following that the extreme of the off-axis angle (as those illustrated in Fig. 6) are the consequence of inflections in the rotational movement of the whale.

One can then show that only the third model could lead to an $ICI_{max} \Leftrightarrow \delta_{max}$ synchronization using a hydrophone located either in front of or behind the whale. Such an important result, an asymmetry in the use of the sperm whale's biosonar, need to be borne out by analyzing more data before being considered as representative of the sperm whale acoustic behavior. It could be, for instance, the result of the asymmetry in the anatomy of the sperm whale sould generator, or part of a metabolic saving foraging strategy (Laplanche, 2005).

Such an asymmetry in the sperm whale acoustic behavior also explains the often observed synchronization of the ICI and apparent click level (AL). As illustrated in Fig. 7, the relationship $ICI_{max} \Leftrightarrow AL_{max}$ is detectable using a hydrophone located in a large portion of space [on side (C)]. The ICI-δ-AL appear to be decorrelated when using a hydrophone in front of the whale [in sector (B)].

The technique presented in Laplanche *et al.* (2005) has been used to measure the vertical movements of diving sperm whales. It does not detect horizontal movements, hence the inexactitude of the model previously presented by the authors. However, the existence of horizontal movements does not alter the measurement of vertical movements, and the technique used by Laplanche *et al.* (2005) remains valid. The description of the vertical movements of sperm whales while descending and foraging at depth is then to be supplemented with the description of their horizontal movements.

The sperm whale, regarding the dive chosen as an example, makes a slow vertical pitch movement and a faster yaw movement with no rolling during the descent. Data presented by Zimmer *et al.* (2005a) would rather suggest that sperm whales create a fast pitch movement during the descent as a result of fluking. The fast movement, has, however, not been detected by Laplanche *et al.* (2005), suggesting that it is horizontal. The authors' present results suggest that, in the example illustrating this work, the whale would analyze the content of a horizontal water expanse during this descent, by making a slow, vertical, pitch movement and create a faster, horizontal, yaw movement.

At depth, regarding the sequence #670$\rightarrow$#730 (Fig. 6), the whale makes a slow vertical pitch movement, barely rolls, and creates a faster yaw movement. The latter movement would be horizontal, for reasons similar to those mentioned above, though this statement is highly speculative, because Laplanche *et al.* (2005) have used a different dataset to show that sperm whales would not create fast, vertical movements during the searching phases. The whale would, in this case, analyze the content of a vertical water expanse. Concerning the sequence #730$\rightarrow$#770 (Fig. 6), the fast movement would be a pitch movement.

The authors have not analyzed enough data to conclude on how sperm whales precisely move when searching for prey. The method itself is qualitative, as a result of making 2 measurements—$\tau_{1/2}$ and $\tau_{1'1}$—when trying to estimate three angles-pitch, yaw, and roll. The authors could not use it to precisely determine the geometry of the water masses analyzed by sperm whales when foraging. Nevertheless, the method presented is promising, and the first results demon-

strate that sperm whales actively search for their prey, by creating a fast, periodic movement in sync with their acoustic activity.

V. SUMMARY AND CONCLUSIONS

The traditional view of the sperm whale click multipulse structure describes a click as a regularly spaced series of pulses $\{p0,p1,p2,p3,\ldots\}$. Nevertheless, additional pulses are often detected within this structure. Their presence has been explained using the standard bent-horn model of the sperm whale sound generator. Frontal reflections $\{p1/2,p3/2,p5/2,\ldots\}$ and distal reflections $\{p1',p2',p3',\ldots\}$ would also leak from the sperm whale snout, leading to a more complex sperm whale click multipulse structure when recorded by a farfield, off-axis hydrophone.

The delays $\tau_{1/2}$ (of $p1/2$ on $p0$) and $\tau_{1'1}$ (of $p1$ on $p1'$) are variable and depend on the whale's orientation to the hydrophone. The off-axis angle of the whale can be calculated using the measurements of the standard IPI and the delay $\tau_{1/2}$. The nature of the movements of the whale (if it is pitch, yaw, or roll) can be measured by analyzing the variations of the delays $\tau_{1/2}$ and $\tau_{1'1}$. Variations of $\tau_{1/2}$ alone (with $\tau_{1'1}$ constant) indicate pure yaw movements, variations of $\tau_{1'1}$ alone (with $\tau_{1/2}$ constant) indicate pure roll movements, and synchronous variations of $\tau_{1/2}$ and $\tau_{1'1}$ indicate pure pitch movements. The analysis of the waveform of usual click sequences then leads to some insight on the sperm whale movements when foraging.

Sperm whales would barely roll when descending and when searching for prey at depth. In the example illustrating this work, the authors have shown that the whale made a slow vertical pitch movement and at the same time created a faster, periodic yaw movement when descending. When searching for prey at depth, the sperm whale made a slow pitch movement and created a faster pitch or yaw movement in sync with the clicking activity. The authors have suggested that sperm whales would, at depth, make an asymmetric scan of the surrounding water. The method presented is easily implemented, is repeatable, and could lead to new insights on the foraging behavior of sperm whales or other toothed whales using a very simple apparatus consisting of a single hydrophone.

ACKNOWLEDGMENTS

Thanks to L. Conteville (Université Paris XII, France) for helping in analyzing the data. Recordings could not have been made without the help of the Association Breach (http://www.breach-asso.org, France), which provided boats, skippers, and ecovolunteers. Thanks to the Association KetoΣ (France) for undertaking the airborne survey. Many thanks to S. Wiener for reviewing an earlier version of the manuscript and to L. Adam, M. Adam, and C. Bédard for refining the English. The authors also thank M. Macé for his help on the bibliography and W. Zimmer for his useful and encouraging comments on the accepted version of the manuscript.

Backus, R., and Schevill, W. (**1966**). "*Physeter* Clicks," in *Whales, Dolphins and Porpoises*, edited by K. Norris (University of California Press, Berkeley), pp. 510–527.

Bahl, R., Ura, T., and Fukuchi, T. (**2002**). "Towards identification of sperm whales from their vocalizations," *Seisan-Kenkyu bimonthly J. of IIS*, University of Tokyo, Vol. **54-6**, pp. 51–55.

Clarke, M. (**1978a**). "Buoyancy control as a function of the spermaceti organ in the sperm whale," J. Mar. Biol. Assoc. U.K. **58**, 27–71.

Clarke, M. (**1978b**). "Structure and proportions of the spermaceti organ in the sperm whale," J. Mar. Biol. Assoc. U.K. **58**, 1–17.

Cranford, T. (**1999**). "The sperm whale's nose: Sexual selection on a grand scale?," Marine Mammal Sci. **15**, 1133–1157.

Flewellen, C., and Morris, R. (**1978**). "Sound velocity measurements on samples from the spermaceti organ of the sperm whale (*Physeter catodon*)," Deep-Sea Res. **25**, 269–277.

Goold, J. (**1996**). "Signal processing techniques for acoustic measurement of sperm whale body lengths," J. Acoust. Soc. Am. **100**, 3431–3441.

Gordon, J. (**1991**). "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations," J. Zool. **224**, 301–314.

Jaquet, N., Dawson, S., and Douglas, L. (**2001**). "Vocal behavior of male sperm whales: Why do they click?" J. Acoust. Soc. Am. **109**, 2254–2259.

Kawakami, T. (**1980**). "A review of sperm whale food," Sci. Rep. Whales Res. Inst. (Tokyo) **32**, 199–218.

Laplanche, C. (**2005**). "Etude par acoustique passive du comportement de chasse du grand cachalot," Ph.D. thesis, Université Paris 12, Créteil.

Laplanche, C., Adam, O., Lopatka, M., and Motsch, J. (**2005**). "Male sperm whale acoustic behavior observed from multipaths at a single hydrophone," J. Acoust. Soc. Am. **118**, 2677–2687.

Lockyer, C. (1981). "Estimates of growth and energy budget for the sperm whale," *Physeter* catodon, in Mammals in the Seas, vol. **3**, FAO series No. 5, 489–504.

Madsen, P., Johnson, M., Aguilar de Soto, N., Zimmer, W., and Tyack, P. (**2005**). "Biosonar performance of foraging beaked whales (*Mesoplodon densirostris*)," J. Exp. Biol. **208**, 181–194.

Madsen, P., Payne, R., Kristiansen, N., Wahlberg, M., Kerr, I., and Møhl, B. (**2002**). "Sperm whale sound production studied with ultrasound time/depth-recording tags," J. Exp. Biol. **205**, 1899–1906.

Miller, P., Johnson, M., and Tyack, P. (**2004**). "Sperm whale behavior indicates the use of echolocation click buzzes 'creaks' in prey capture," Proc. R. Soc. London, Ser. B **271**, 2239–2247.

Møhl, B., Wahlberg, M., Madsen, P., Heerfordt, A., and Lund, A. (**2003**). "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154.

Møhl, B., Wahlberg, M., Madsen, P., Miller, L., and Surlykke, A. (**2000**). "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648.

Mullins, J., Whitehead, H., and Weilgart, L. (**1988**). "Behavior and vocalizations of two single sperm whales, *Physeter macrocephalus*, off nova scotia," Can. J. Fish. Aquat. Sci. **45**, 1736–1743.

Nishiwaki, N., Ohsumi, S., and Maeda, Y. (**1963**). "Change of form in the sperm whale accompanied with growth," Sci. Rep. Whales Res. Inst. (Tokyo) **17**, 1–14.

Norris, K., and Harvey, G. (**1972**). "A theory for the function of the spermaceti organ of the sperm whale (*Physeter catodon L.*)," in *Animal Orientation and Navigation*, edited by S. Galler, K. Schmidt-Koenig, G. Jacobs, and R. Belleville (NASA, Washington, D.C.), pp. 397–417.

Rhinelander, M., and Dawson, S. (**2004**). "Measuring sperm whales from their clicks: Stability of interpulse intervals and validation that they indicate whale length," J. Acoust. Soc. Am. **115**, 1826–1831.

Urick, R. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York), 3rd ed..

Watkins, W., Dahar, M., DiMarzio, N., Samuels, A., Wartzok, D., Fristrup, K., Gannon, D., Howey, P, and Maiefski, R. (**1999**). "Sperm whale surface activity from tracking by radio and satellite tags," Marine Mammal Sci. **15**, 1158–1180.

Whitehead, H. (**2003**). *Sperm Whales: Social Evolution in the Ocean* (The University of Chicago Press, Chicago).

Whitehead, H., and Weilgart, L. (**1990**). "Click rates from sperm whales," J. Acoust. Soc. Am. **87**, 1798–1806.

Zimmer, W., Madsen, P., Teloni, V., Johnson, M., and Tyack, P. (**2005a**). "Off-axis effects on the multipulse structure of sperm whale usual clicks with implications for sound production," J. Acoust. Soc. Am. **118**, 3337–3345.

Zimmer, W., Tyack, P., Johnson, M., and Madsen, P. (**2005b**). "Three-dimensional beam pattern of regular sperm whale clicks confirms bent-horn hypothesis," J. Acoust. Soc. Am. **117**, 1473–1485.

A helical biosonar scanning pattern in the Chinese Noctule, *Nyctalus plancyi*

Rolf Müller[a] and Hongwang Lu
School of Physics & Microelectronics, Shandong University, 5 Hongjia Lou, 250100 Jinan, China

Shuyi Zhang
Institute of Zoology, Chinese Academy of Sciences, 25 Beisihuan Xilu, 100080 Beijing, China

Herbert Peremans
Department of Environment & Technology Management, University of Antwerp, 13 Prinsstraat, 2000 Antwerp, Belgium

(Received 6 January 2006; revised 6 April 2006; accepted 6 April 2006)

Directivity and sound diffraction of the pinna of the Chinese Noctule (*Nyctalus plancyi*) have been studied numerically. The pinna was found capable of generating a periodic helical scanning pattern over frequency, if the tragus and the thickened lower ledge of the pinna rim were in an appropriate position. During the helical scan, a directivity pattern with a strong mainlobe alternated with a pattern dominated by a conical sleeve of sidelobes. This alternation was present, even when an unfavorable arrangement of the pinna disrupted the overall helical scanning pattern. In the fully formed helical scan, the orientation of main and sidelobes for different frequencies revealed a spatial ordering which extends volume coverage. Five different pinna parts have been removed from the digital pinna-shape representations in turn to assess their influence on the directivity. Of these parts, the tragus stem and the thickened lower ledge of the pinna rim were found to have the largest overall impact. The anatomical prominence of these structures was hence in agreement with their acoustic functionality. In the near-field, tragus stem and lower ledge were seen to act primarily through large shifts in the wavefield phase in both directions.
[DOI: 10.1121/1.2200151]

PACS number(s): 43.80.Ka, 43.64.Ha [JAS] Pages: 4083–4092

I. INTRODUCTION

Among the more than 1000 different species of biosonar-utilizing microbats, a highly diverse set of outer ear (pinna) shapes has evolved. Many of these pinnas feature prominent structural elements like an enlarged tragus or antitragus, thickened ledges along the pinna rim, as well as grooves and ripples on the sound-exposed pinna surface. It may be hypothesized that a large portion of these structural features serve an acoustic function and hence may give rise to an equally diverse set of acoustic properties to support the various biosonar tasks that these animals are known to solve. The possible effects of these adaptations on the acoustic function of the pinna could range from minor modifications to the common mammalian pattern to the introduction of novel functional principles.

In order to characterize the acoustic function of a bat pinna fully, it is necessary to obtain a high-resolution estimate of its directivity pattern, i.e., the sensitivity of the ear as a function of azimuth, elevation, and frequency. Coarse measures of the directivity, for example, its overall width, which have been measured on fairly large samples of pinnas from different bat species before (Obrist *et al.*, 1993), cannot capture accurately the part of the functional diversity that is likely to manifest itself in the presence, shape, and orientation of the main lobe and the sidelobes. Yet, all these features are of potential functional relevance, because they determine how the ear's sensitivity is distributed in space as a function of frequency. This distribution will impact the performance in many—if not all—sonar tasks.

Establishing which structural elements have a significant influence on the directivity pattern requires subjecting the features in question to experimental manipulations. This can be done either in real physical experiments (Aytekin *et al.*, 2004; Wotton *et al.*, 1996, 1995) or in numerical experiments, where the digital representation of a shape is altered (Müller, 2004). Digital shape manipulations have the advantage that each structural feature can be removed or altered without any side effects on the others and that the alterations are completely reversible. The latter is important when possible interactions between several structural properties have to be taken into account. Reversals are needed in such cases to generate all possible combinations of the individual features being present or absent in a single pinna.

In the work presented here, high-resolution numerical predictions of the directivity pattern are derived for the ear of the Chinese Noctule (*Nyctalus plancyi*) in order to provide a detailed picture of how the spatial sensitivity of this ear behaves as a function of frequency. This is combined with digital shape manipulations in which a set of derived ear shapes is created. This shape set comprises all combinations of five selected conspicuous morphological features being present or having been removed. By comparing the directivity pat-

[a]Electronic mail: mueller@sdu.edu.cn

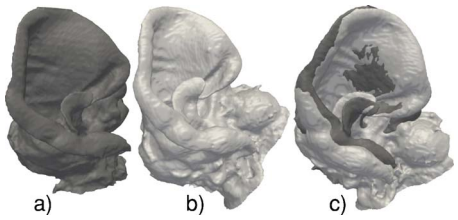


FIG. 1. Renderings of the investigated pinna shapes using triangular surface meshes: (a) right pinna, (b) left pinna (mirrored), (c) manually aligned superposition of right and left pinna showing the differences in the mounting positions.

terns across the entire set of shapes, it is possible to determine which of the studied features are major contributors to the directivity pattern and to check for possible interactions between them.

The general goal of this research is to contribute a case study, which—together with other case studies of this type—can ultimately lead to a better understanding of the functional diversity in the directivity patterns that bat pinnas are capable of generating. It is also meant as a case study on how this—largely unexplored—diversity in function relates to the known diversity in pinna morphology. In addition to the results on the directivity patterns and their relationship to morphology, the present work also introduces two methods to the numerical study of bat pinna. In particular, it demonstrates how a digital representation of a pinna shape can be augmented with an artificial ear canal and how a complete set of manipulated pinna shapes can be generated through a Boolean combination of binary voxel representations.

II. METHODS

A. Biological samples

An adult female specimen of the Chinese Noctule (*Nyctalus plancyi*) was collected on November 10, 2004 from the attic of a house in western Hubei Province, China (geographical coordinates: 111°18.2′E, 32°23.8′N). Biosonar pulses recorded from the specimen in a flight room fell into a frequency band extending from 23.1 to 73.8 kHz; for a larger sample of Chinese Noctules from Hubei Province (9 individuals), a frequency range from 24.9±2 to 79.4±9.2 kHz (mean±standard deviation) has been reported (Wang *et al.*, in preparation).

After the animal was sacrificed for a virological study, the pinnae were surgically removed from the carcass and stored in a weak alcohol solution for a few days until they were subjected to micro-computer tomographic scanning at a resolution of 30 μm (isotropic voxel size) using a Scanco μCT40 scanner. A rendering of the obtained digital shape representations is shown in Fig. 1. The two pinnae (right and left) were mounted with the tragus and the thickened lower ledge in slightly different positions: In the right pinna, the tragus was positioned more outward and the lower ledge was shifted upward and inward when compared to the mounting position of the left pinna (see Fig. 1). In its mounting position, the right pinna was measured to be 15 mm wide, 8 mm deep, and 19 mm high. The respective measurements for the left pinna were 17, 8, and 19 mm. The left pinna was mirrored to facilitate the comparison with the right pinna.

FIG. 2. Artificial ear canal (dark shading, arrow) attached to the right pinna in a rendering of a smoothed triangulated surface mesh.

B. Image processing

The l6-bit gray-scale images representing the tomographical cross sections of the pinna shapes were processed to yield an estimate of the boundary between pinna tissue and the surrounding air. For this purpose, the gray values of the voxels in the stack of cross-section images were first smoothed with a three-dimensional Gaussian low pass filter (the standard deviation in each dimension: two voxels) and then thresholded. The number of misclassified voxels inside the tissue and air volumes was reduced by applying a three-dimensional morphological open-close filtering operation with a kernel size of three voxels. Finally, small unconnected regions of pixels classified as "tissue" were removed if their area in a given image was less than 16 pixels. The binary images were then subsampled to a voxel size of 120 μm, thereby setting the resolution used in generating the finite element meshes (the ear boundary surface of the finite-element mesh is shown for the right pinna in Fig. 3). The results of these automated image processing steps were screened and a few remaining obvious artifacts were manually corrected using a bitmap-image processing program.

C. Artificial ear canal

It is difficult to preserve the shape of the ear canal during preparation of the biological samples, because it is supported by the surrounding bone and all bone has to be removed from the sample in order to facilitate high-quality imaging of the air-tissue boundary. Since the shape of the original ear canal could not be preserved for this reason, an artificial ear canal was attached to the digital shape representations of both ears (see Fig. 2). The presence of an ear canal

ensures that a planar wavefront exists at the ear-canal aperture opposite to the source whenever the studied wavelength is larger than the diameter of the ear canal. This is the case in the present study, where the ear canal opening had a maximum diameter of less than 2.5 mm and the smallest studied wavelength (see Sec. II D) was 4.3 mm. Apart from the possible effect of shape of the wavefront in the aperture, the precise shape of the ear canal does not contribute to the normalized directivity pattern and can hence be ignored for the purpose of this study (Müller, 2004). The artificial ear canal was constructed on a set of hand-marked voxels selected along the rim of the ear canal aperture in the concha. An eigenvector/eigenvalue decomposition was performed on the position of these rim voxels in three-dimensional Cartesian space. The eigenvector with the smallest associated eigenvalue was taken as an estimate for the aperture normal (the two eigenvectors associated with the two largest eigenvalues are the principle axes of the aperture). New tissue voxels were then added to rim voxels along the direction of the aperture normal until the boundary of the computational volume was reached. The artificial ear canal was closed at its proximal end with a perfectly reflecting boundary.

D. Numerical methods

Each voxel classified as "air" in the subsampled stack of binary cross-section images was converted into a linear (first-order) cubic finite element with an edge length of 120 μm and eight nodes. For the studied wavelengths (see later), this corresponds to a spatial resolution of 36 to 117 finite elements per wavelength. The entries in the element stiffness matrices $K_{ij}^{(\mathrm{el})}$ are real numbers that were determined by analytically evaluating the expression

$$K_{ij}^{(\mathrm{el})} = \Delta K_{ij}^{(\mathrm{el})} - k^2\,\Delta M_{ij}^{(\mathrm{el})} = \int_{V^{(\mathrm{el})}} \nabla N_i^T\,\nabla N_j - k^2 N_i N_j\,dV, \tag{1}$$

where $\Delta K_{ij}^{(\mathrm{el})}$ is the stiffness term, $\Delta M_{ij}^{(\mathrm{el})}$ is the mass term, N_i is the ith shape function, $k=2\pi f/c$ the wavenumber, and $V^{(\mathrm{el})}$ the volume of the element. Stiffness and mass terms are both independent of frequency and the element stiffness matrix is the same for all elements in the regular cubic finite-element mesh.

The linear shape function for the node $(-1,-1,-1)$ given in normalized local element coordinates (ξ,η,ζ) is

$$N_1(\xi,\eta,\zeta) = \frac{1}{8}\times(1-\xi)\times(1-\eta)\times(1-\zeta). \tag{2}$$

The shape functions for the other nodes are found by changing the sign of the local coordinates to ensure that the shape function has a value of one at its respective node and zero at all others.

The cuboidal computational volume of the finite element calculations was chosen to enclose the ear shape generously, leaving several elements representing air in each direction, but at the bottom of the ear and at the proximal end of the artificial ear canal. On the boundary of the computational volume, reflection-free sound propagation was modeled by a layer of three-dimensional mapped wave-envelope infinite elements (Astley and Macaulay, 1998). The entries in the element stiffness matrix of the infinite elements are given by

$$K_{ij}^{(\mathrm{el})} = \Delta K_{ij}^{(\mathrm{el})} + jk\,\Delta C_{ij}^{(\mathrm{el})} - k^2\,\Delta M_{ij}^{(\mathrm{el})}, \tag{3}$$

where j is the imaginary unit and $\Delta C_{ij}^{(\mathrm{el})}$ the damping term. In the radial direction, Jacobi polynomials of order three with parameters $\alpha=2$ and $\beta=0$ were used as a basis function following a suggestion by Dreyer and von Estoff (2003). The shape functions for the infinite elements were integrated numerically using Gauss-Legendre quadrature.

A point source was placed at a single element node toward the proximal end of the artificial ear canal and near the center of its cross section. The resulting well-posed linear problem was solved iteratively for the complex wavefield amplitudes at the element nodes using a biconjugate gradient stabilized method (Barret *et al.*, 1994; van der Vorst, 2003) with a successive over-relaxation preconditioner. The preconditioner's splitting matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \frac{1}{\omega}\mathbf{D} - \mathbf{C}_l, \tag{4}$$

where $\mathbf{D}$ is the diagonal matrix of the system matrix (master stiffness matrix), $\mathbf{C}_l$ its strictly lower triangular portion, and $0<\omega<2$ the relaxation factor (set to a value of 1.2). The implementation of the linear system solving routines was taken from the Portable, Extensible Toolkit for Scientific Computation (PETSc) software library (Balay *et al.*, 1997).

The wavefield estimates on the outer layer of finite-element nodes were projected to a set of points on the surface of a sphere in the farfield in order to obtain the normalized directivity pattern $0\leqslant D(\phi,\theta,f)\leqslant 1$, i.e., the relative gain of the pinna as a function of azimuth ϕ, elevation θ, and frequency f, which is valid for all distances in the farfield (Urick, 1983). The angular resolution was set to 1° for azimuth and elevation, resulting in 65 160 gain estimates being made for each ear shape and frequency. The farfield projection was carried out using the Kirchhoff integral (Jackson, 1999; Ramahi, 1997),

$$\Psi(\vec{\mathbf{x}}) = -\frac{1}{4\pi}\int_S \frac{e^{jkR}}{R}\vec{\mathbf{n}}\cdot\left[\nabla\Psi + jk\left(1+\frac{j}{kR}\right)\frac{\vec{\mathbf{R}}}{R}\Psi\right]ds, \tag{5}$$

where $\vec{\mathbf{R}}$ is the vector between the surface element ds and the position $\vec{\mathbf{x}}$, $\vec{\mathbf{n}}$ is the outward pointing surface normal, Ψ is the field value on S, and k is the wavenumber. The product $\vec{\mathbf{n}}\cdot\nabla\Psi = \partial\Psi/\partial\vec{\mathbf{n}}$ is the derivative of the field Ψ with respect to the surface normal $\vec{\mathbf{n}}$.

Wavefield solutions were computed for 111 frequencies linearly spaced between 24.5 to 79.5 kHz in steps of 500 Hz. The frequency values were converted into wavenumbers, assuming a sound speed of 343.322 m/s corresponding, for example, to air with a temperature of 20 °C, an adiabatic constant (γ) of 1.4, and a molecular weight of 0.02895 kg/mol.

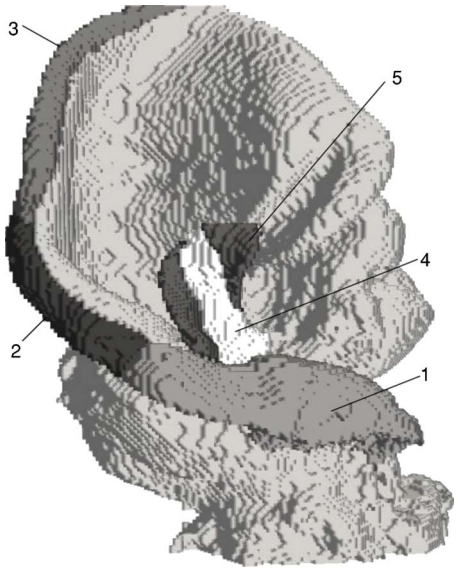


FIG. 3. Definition of the investigated ear-shape components in the right ear: (l) lower ledge, (2) central ledge, (3) upper ledge, (4) tragus stem, (5) tragus flaps (on both sides of the tragus stem). The rendered surface is the pinna boundary of the finite-element mesh.

E. Digital surgery

Five pinna parts were selected for digital removal (see Fig. 3), three parts in the posterior rim of the pinna and two parts pertaining to the tragus. In the posterior rim of the pinna, three parts were studied: the lower ledge, the central ledge, and the upper ledge (3). The fiducial points demarcating these parts were a thickening in the lower ledge and a bend in the pinna rim between the central and the upper ledge. The tragus was divided into the tragus stem and two lateral flaps, with the flaps being defined as the tissue that laterally extends beyond the width of the tragus in the center portion of its stem.

Each part was removed only once from the digital shape representation. This was done by flipping the gray value of individual voxels or groups of voxels where appropriate by hand. Using the resulting five different shapes as input, all 32 possible combinations of the individual parts being present or absent were generated using the appropriate Boolean combination of the representations, where one part was missing. Each combination was labeled by a bit code in which each part is represented by a binary digit and its removal is marked by setting the respective digit to "l."

To measure the distance between directivity patterns for a given frequency f, a correlation coefficient $\rho_{ij}(f)$ between the directivity functions $D_i(\phi,\theta,f)$ and $D_j(\phi,\theta,f)$ was computed over the two angular dimensions ϕ and θ. The correlation coefficient is given by

$$\rho_{ij}(f) = \frac{C_{ij}(f)}{\sigma_i(f)\sigma_j(f)}, \tag{6}$$

where $C_{ij}(f)$ is the covariance and $\sigma_i(f), \sigma_j(f)$ are the standard deviations of the directivity functions for frequency f. Scaled versions of these terms were estimated from the samples of the directivity functions as

$$C_{ij}(f) = \sum_k \sum_l [D_i(\phi_k,\theta_l,f) - \bar{D}_i(f)] \times [D_j(\phi_k,\theta_l,f) - \bar{D}_j(f)], \tag{7}$$

$$\sigma_i(f) = \sqrt{\sum_k \sum_l [D_i(\phi_k,\theta_l,f) - \bar{D}_i(f)]^2},$$

$$\sigma_j(f) = \sqrt{\sum_k \sum_l [D_j(\phi_k,\theta_l,f) - \bar{D}_j(f)]^2}, \tag{8}$$

where $\bar{D}_i(f), \bar{D}_j(f)$ are the function averages over all directions. In Eqs. (7) and (8), scaling of the estimates by the inverse of the number of samples can be omitted, since these scaling constants cancel in Eq. (6).

F. Visualization of results

Directivity patterns for a given frequency were mapped into a two-dimensional plane using the Lambert Azimuthal Equal-Area projection in which the coordinates x, y in the projection plane are given in terms of azimuth ϕ and elevation θ by (Weisstein, 2002)

$$x = A \cos\theta \sin(\phi - \phi_0)$$

$$y = A[\cos\theta_0 \sin\theta - \sin\theta_0 \cos\theta \cos(\phi - \phi_0)]$$

$$A = \sqrt{\frac{2}{1 + \sin\theta_0 \sin\theta + \cos\theta_0 \cos\theta \cos(\phi - \phi_0)}}, \tag{9}$$

where ϕ_0, θ_0 is the center of the projection. This projection was chosen to display the directivity pattern because its equivalence (i.e., equal-area) property makes it well suited for representing how sensitivity is distributed with angle.

Three-dimensional isosurfaces for the farfield wave amplitudes were rendered to visualize the "sensitivity beam" of the pinna at a given frequency. The wavefield amplitudes were obtained in spherical coordinates $(\phi,\theta,r$, where r is the radius) by multiplying the directivity gain $D(\phi,\theta,f)$ with the amplitude of a point source $1/r$, i.e., the gain g is

$$g(\phi,\theta,r,f) = D_i(\phi,\theta,f) \times \frac{1}{r}. \tag{10}$$

Hence, the position of an isosurface in a given direction ϕ,θ can be found directly by setting Eq. (10) equal to a constant and solving for r. This facilitates the determination of high-resolution isosurfaces because only wavefield amplitudes right on the surface need to be calculated. Because the radial distance r of the isosurface and the selected directivity gain threshold are inversely proportional [see Eq. (10)], different choices of the latter only result in a scaling of the isosurfaces and isosurface shape can be judged without paying attention to the particular threshold value chosen or the scale of the rendering.

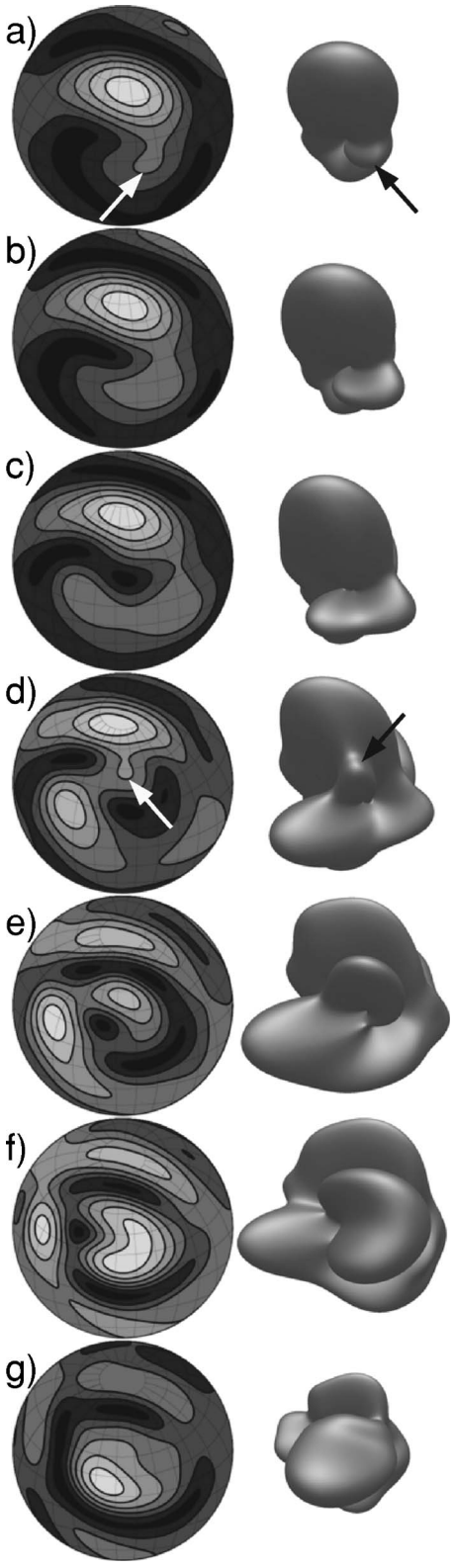


FIG. 4. Helical scanning cycle in the right pinna illustrated by normalized directivity maps (left column, Lambert azimuthal equal-area projection, projection center: azimuth $\phi_0=80°$, elevation $\theta_0=35°$, see Eq. (9), normalized directivity amplitude is mapped to gray value using seven linearly spaced intervals between 0 and 1) and corresponding three-dimensional isosensitivity surfaces (right column): (a) 38 kHz: a new sidelobe (arrow) is spun off the main lobe; (b) 40 kHz: sidelobe growth; (c) 42 kHz: further sidelobe growth; (d) 44 kHz: new main lobe (arrow) formed; (e) 46 kHz: main lobe growth; (f) 48 kHz: further main lobe growth; (g) 50 kHz: sidelobe retraction.

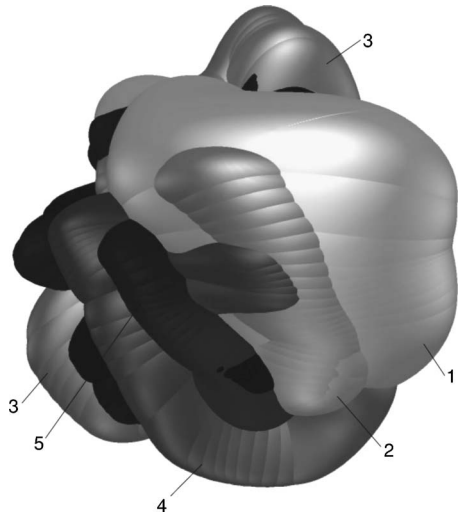


FIG. 5. Spatial ordering of main and sidelobes in a rendering of isosensitivity surfaces for the right pinna. Numbers denote directivity pattern periods over frequency: (l) main lobe of the first observed period from the low frequency band edge; (2) main lobe of the second observed period; (3) sidelobes of the second observed period; (4) main lobe of the third period; (5) main lobe fourth period.

III. RESULTS

A. Directivities of the natural ear shapes

The directivity pattern of the right ear was found to display an approximately periodical pattern over frequency, in which similar kinds of directivity pattern follow each other in a repeatable sequence: With rising frequency, a new sidelobe is spun off the mainlobe and then grows in size [Figs. 4(a)–4(c)] to form a conical sleeve together with the remnants of the former main lobe [Figs. 4(d)–4(f)]. Close to the center of this side-lobe sleeve, a new main lobe is created [Fig. 4(d)], which then concentrates most of the sensitivity, while the sidelobe sleeve recedes [Fig. 4(g)]. In total, approximately four cycles of this kind were found inside the investigated frequency band, with a strong main lobe being found present around 28, 35.5, 50.5, and 65 kHz and absent around 30.5, 43.5, and 56.5 kHz. The periodicity is only approximate, because neither the length of the period nor the shape of the directivity pattern were exactly repeated. The orientation of the main and sidelobes changed as a function of frequency in a systematic fashion, both within a period and between different periods in the succession of directivity patterns. Within periods, the largest sensitivity of the main lobe or in the sidelobe sleeve shifts gradually, performing a continuous smooth scan (Fig. 5). Between periods, the location of the lobes shifts so that the continuous scans performed in each period exhibit little overlap and gaps in the sensitivity are filled in (Fig. 5). Taken together, changes in the lobe orientation within and between periods result in a helical scanning pattern, in which spatial sensitivity is shifted between the main lobe and the sidelobe sleeve as it probes space in a regular spiral pattern. In the left ear, the regular periodical scanning pattern seen in the right ear was only replicated to a very limited extent. Only one period was found between approximately 40 and 52 kHz (Fig. 6), where a main lobe was replaced by a sleeve of sidelobes. Below

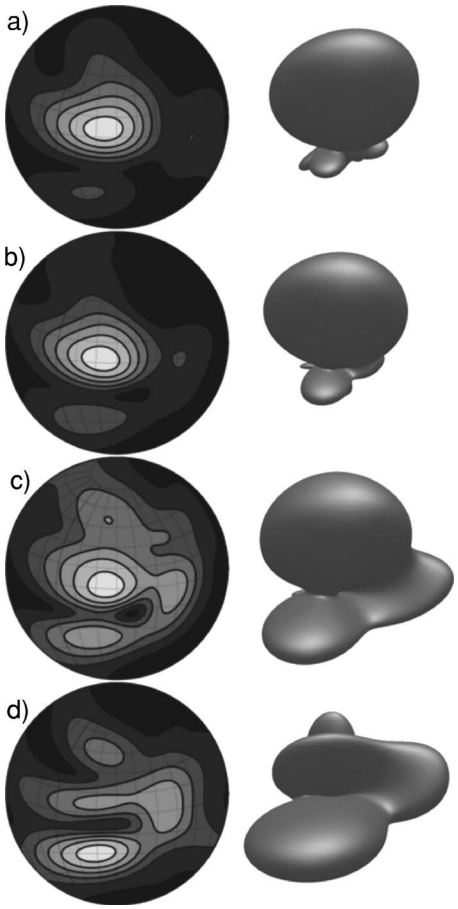


FIG. 6. Helical scanning cycle in the left pinna illustrated by normalized directivity maps [left column, Lambert azimuthal equal-area projection, projection center: azimuth $\phi_0=45°$, elevation $\theta_0=15°$, see Eq. 9, mapping of directivity amplitudes to gray values as in Fig. 4] and corresponding three-dimensional isosensitivity surfaces (right column): (a) 40 kHz: sidelobe is spun off the main lobe; (b) 44 kHz: sidelobe growth; (c) 48 kHz: sidelobe growth; (d) 52 kHz: sidelobes overtake the main lobe.

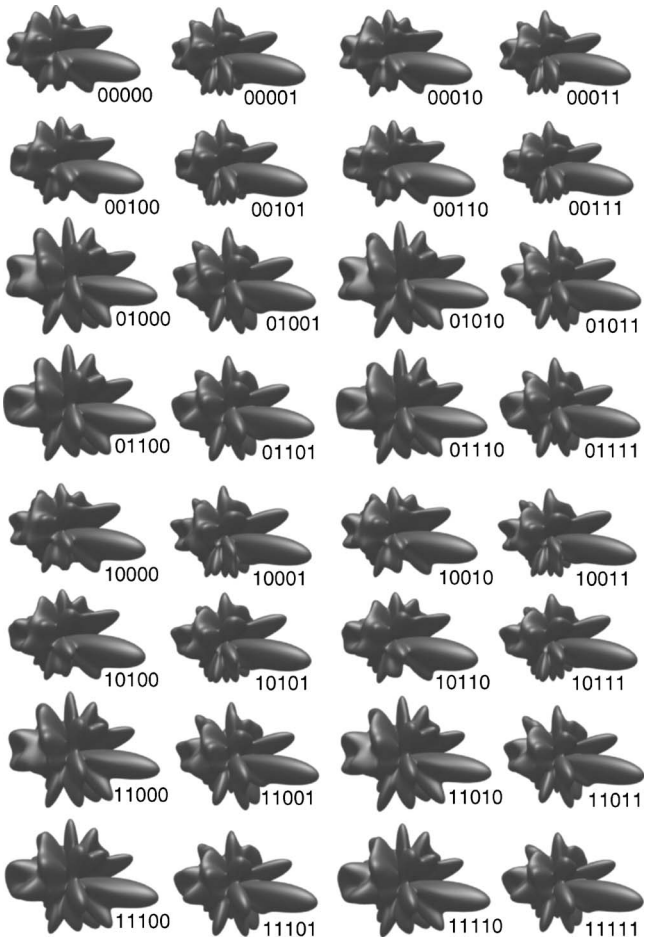


FIG. 7. Beam pattern at 65 kHz for all possible combinations of presence (marked with a "0") and absence (marked with a "1") of the shape components shown in Fig. 3. Upper left: all components present; lower right: all components removed. Components are coded according to the sequence given in Fig. 3, with the tragus flaps represented by the leftmost (most significant) digit and the lower ledge of the pinna rim by the rightmost (least significant) digit.

this frequency band, the directivity pattern tended to emphasize a single main lobe with small, irregular sidelobes, whereas above the frequency band, it was split into several lobes.

B. Directivities of manipulated shapes

For all frequencies at which the manipulated ear shapes were studied (25, 35, 45, 55, and 65 kHz), but the highest (75 kHz), removal of the ear edge structures manifested itself primarily in the sidelobes (see Fig. 7). The directivity pattern at 75 kHz was found to be broken up into multiple lobes, thus no meaningful distinction between main and sidelobes could be made.

The removal of different pinna parts altered directivity patterns to a different extent depending on the parts removed and on the frequency considered. Low-frequency directivity patterns obtained for the different ear manipulations showed a fairly strong correlation, irrespective of the shape manipulation performed (see Fig. 8). This means that while directivity patterns did undergo limited changes in response to the manipulations carried out, they retained a rather high overall similarity. In contrast to this, high and low correlation coefficients were found when comparing the directivity patterns generated by the manipulated shapes at higher frequencies. For these frequencies, the shape manipulations did thoroughly alter the directivity pattern in many cases, but not all the time (see Fig. 8). The correlation coefficient matrices for all frequencies were dominated by two approximately additive checkerboard patterns with different spatial frequency. Because of the larger range in the correlation coefficients, both patterns became more pronounced for the higher frequencies (see Fig. 8). These two prominent checkerboard patterns were found to correspond to the lower-ledge portion of the pinna rim (represented by the least significant bit in the bit patterns in Figs. 7 and 8) and the tragus stem (represented by the second most significant bit in the bit patterns in Figs. 7 and 8). Their effect on the similarity of the directivity patterns as measured by the correlation coefficient was found to be approximately additive and dominating the overall similarity of the directivity patterns as measured by the correlation coefficient.

C. Wavefield properties

For both pinnae (right and left) and over the entire frequency band studied, the isosurfaces for the stationary wave-

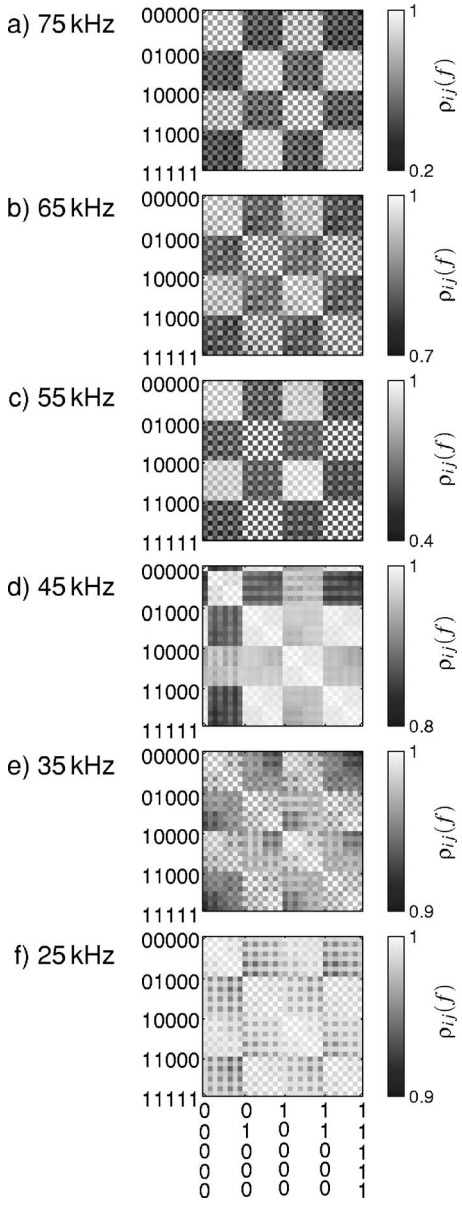


FIG. 8. Correlation [correlation coefficient $\rho_{ij}(f)$, see Eq. (6)] between directivity gains: (a) 75 kHz, (b) 65 kHz, (c) 55 kHz, (d) 45 kHz, (e) 35 kHz, and (f) 25 kHz. Components are coded according to the sequence given in Fig. 3 with the tragus flaps represented by the leftmost digit and the lower ledge by the rightmost digit.

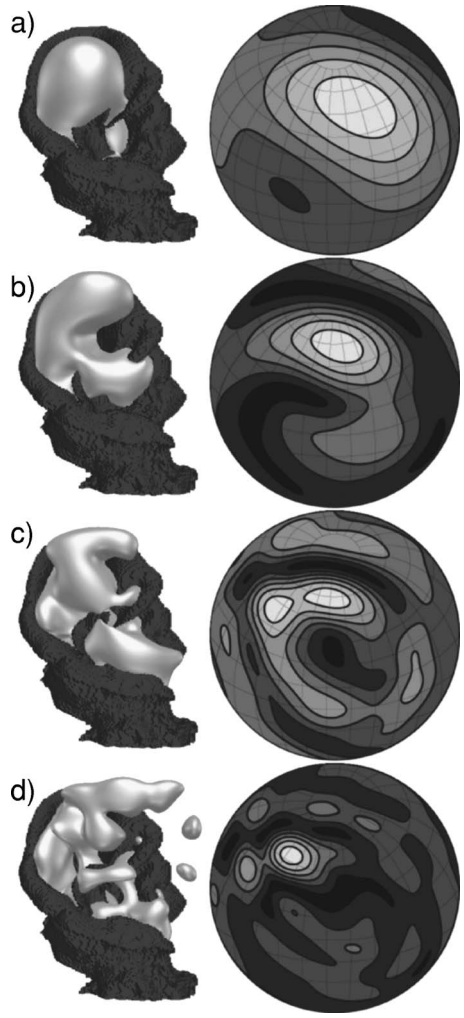


FIG. 9. Isosurfaces for wavefield amplitude compared to normalized directivity patterns [Lambert azimuthal equal-area projection, projection center: azimuth $\phi_0=80°$, elevation $\theta_0=30°$, see Eq. (9), mapping of directivity amplitudes to gray values as in Fig. 4] at (a) 25 kHz, (b) 40 kHz, (c) 55 kHz, and (d) 70 kHz.

field amplitudes in the computational domain of the finite-element model were found to be crescent shaped to some extent. The concave side of the crescent was aligned with the posterior portion of the pinna (see Fig. 9). The direction of the crescent's upper end was aligned roughly with the main lobe of the directivity pattern where present or with the upper portion of the sidelobe sleeve in directivity patterns without a distinct main lobe, whereas the lower end of the crescent was roughly aligned with the lower sidelobes, where present (see Fig. 9). The pinna shape manipulations that were performed, in particular removal of the pinna rim's lower ledge and the tragus stem, did not have a large or otherwise clearly discernible influence on the overall pattern of the wave amplitude in the near field. In contrast to this, the removal of the lower ledge and tragus stem did have a clearly discernible impact on the phase of the wavefield for all frequencies studied. The exact nature of the changes in the spatial pattern of the wavefield phase depended on frequency. Both phase advances and delays were found. Most commonly, the changes affected the volume anterior to the tragus and above the lower ledge (see arrow 1 in Fig. 10). However, effects in regions farther away from the removed structures were also seen, for example, at 35 kHz, where removal of the tragus stem led to a phase advance in front and above of the anterior rim of the pinna. Removal of the lower ledge of the pinna rim and the tragus stem were found to be capable of producing opposite effects on the wavefield phase. At 55 kHz, for instance, the wavefield phase was advanced in the volume anterior of the tragus by values up to approximately 90° when the lower ledge was removed [see Fig. 10(b)]. In contrast to this, removal of the tragus stem shifted the wavefield phase in the opposite direction (phase delay) in a volume that overlapped with the volume affected by removal of the lower ledge to a large extent. In addition, sharp boundaries between the volumes affected by the differential

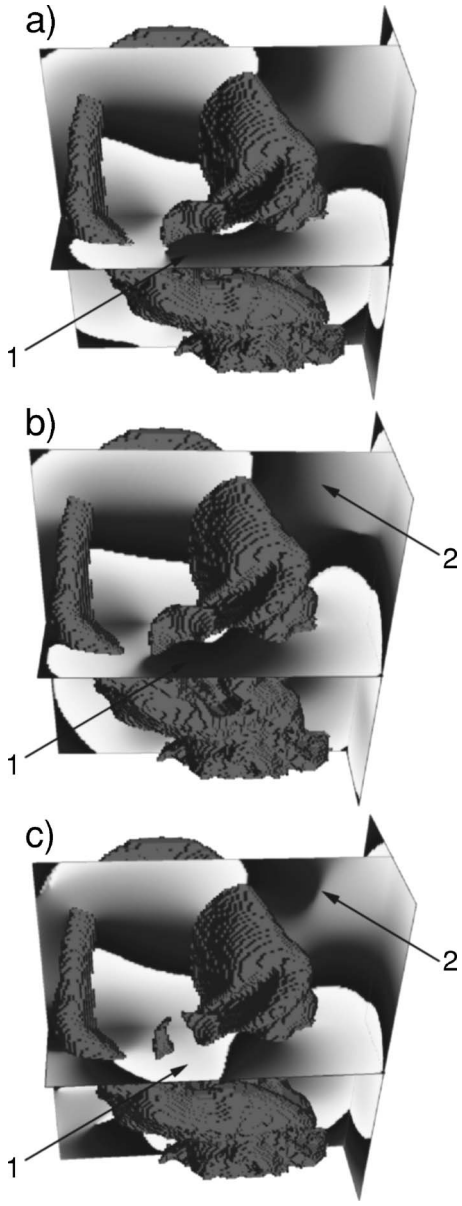


FIG. 10. Planar cuts through the wavefield phase values at 55 kHz: (a) original ear shape; (b) lower ledge of pinna rim removed; (c) tragus stem removed. The phase values are coded by gray levels, ranging from white for −180° to black for 180°. The sharp transitions between black and white are due to the cyclical nature of the phase and do not constitute wavefield discontinuities. Arrows: (l) bidirectional phase change in volume anterior to tragus; (2) example of a sharpening of boundary in phase value.

phase advances and delays and neighboring regions were created (see Fig. 10(c), arrow 2). The latter effect was also observed at other frequencies.

IV. DISCUSSION

The results of the numerical acoustic analysis presented here have demonstrated that the ear of the Chinese Noctule is capable of generating a motion-free helical scanning pattern as a function of frequency. This pattern is noteworthy because of its periodical character (qualitatively similar directivities are repeated) and, in particular, because it alternates between allocating sensitivity to the main lobe and to a conical sleeve of sidelobes. While motion-free scanning patterns have been reported in bats [for example, Firzlaff and Schuller (2003); Müller and Hallam (2005)] as well as in other mammals [for example, the Rhesus monkey, Spezio *et al.* (2000)] before, none of these patterns were shown to have a helical geometry or a periodical alternation in the dominance over sensitivity between a main lobe and a sleeve of sidelobes, which hence appears to be a pattern not described before. Therefore, the pinna of the Chinese Noctule provides another data point to support the hypothesis that the diversity in the shapes of bat pinnae gives rise to a similar diversity in the beam patterns they generate. However, it remains to be established whether this scanning pattern is exclusively associated with the biosonar system of microbats or is a feature shared more widely among mammals.

Motion-free scanning has the general advantage of providing both wide spatial coverage (by pooling information across frequency channels) and spatial resolution (by processing each frequency channel separately) within a very short time duration (the length of the sonar pulse). The regular spatial ordering of successive main and sidelobes (see Fig. 5) in the helical scanning pattern of the Chinese Noctule indicates that both aspects may have played a role in its evolution. However, the functional implications of this particular motion-free scanning pattern has yet to be determined. It may be hypothesized that the alternation between a central main lobe and a sleeve of sidelobes (see Figs. 4 and 6) are used to construct a spatial bandpass filter by combining input from the respective frequency with an opposite sign. Such a mechanism would be similar to the "center-surround antagonisms" known to occur in the mammalian visual system [Westheimer (2004)].

By mounting the left and the right ear from the same animal with slightly different positions of the lower ledge on the pinna rim and the tragus (see Fig. 1), it was demonstrated that the Chinese Noctule could alter its beam pattern considerably by making only small adjustments to the position of these two prominent structures on the lower edge of its pinna. Such small alterations were observed in the animal, although the exact positions assumed by the tragus and the lower ledge of the pinna rim could not be determined in the living animal. The presence of alternations between a central main lobe and a sleeve of sidelobes in the directivity patterns obtained for both mounting configurations (see Figs. 4 and 6) could be taken as an indication that this alternation is a part of the pinna's core functionality, which may be left unchanged by any alteration within the limits of a natural biological function. Nevertheless, the changes in the beam pattern are considerable, and this poses a design challenge for all other components of the animal's biosonar system that have to be integrated with the beam pattern. Thus, if—for instance—behavior and neural signal processing is to be matched to the animal's directivity pattern, these mechanisms must be likewise very flexible to accommodate these changes, and they must be provided with information on the current directivity pattern to make the appropriate adjustments.

Given the complicated nature of the diffraction process that takes place on the surface of the irregularly shaped pinna, the effect of removing pinna parts on the overall cor-

relation of the directivity pattern was found to be surprisingly simple: effects of removing different parts of the pinna on directivity correlation were to a very good approximation additive (see Fig 8). This can be taken as an indication that multipath interactions between the studied pinna parts do not contribute significantly to the directivity patterns.

The ear of the Chinese Noctule—like the ear of other bats in the same genus—has two obvious distinguishing features: a tragus that is broadened by lateral flaps on either side and a prominent sturdy ledge on the lower portion of pinna rim below the tragus. The results of the digital surgery experiments demonstrate that the anatomical prominence of the lower ledge is mirrored by its acoustical importance. The structure was found to be as important as the tragus stem in determining the beam pattern. Removal of the lower ledge did alter the overall directivity pattern much more than removal of the other two studied portions of the posterior pinna rim, which are also much less prominent anatomically. As far as could be established from the published record accessible to the authors, this is the first time that the functional acoustic importance of such a prominent ledge in a bat ear has been demonstrated. The overall effect of the second prominent structure of the noctule pinna, the broad tragus flaps, was found to be small and hence its importance could not be established by the results of the present study. The similarity of the directivity patterns with and without tragus flaps as measured by the correlation coefficient between the patterns failed to highlight the influence of the tragus flap. However, despite these results, it is still possible that the flaps are responsible for subtle changes in the directivity pattern that do not stand out when correlating entire directivity patterns but that may nevertheless have important local functional consequences. Minor changes in the direction of a few spectral notches, for example, may not affect the overall correlation of the beam patterns very much, but could have important consequences for telling the elevation of a sound source (Wotton *et al.*, 1995). It is also conceivable that the effect of the tragus flaps is limited to specific narrow frequency bands, which fall between the frequencies studied here.

The structures found to excise an important influence on the overall beam pattern (the lower ledge of the pinna rim and the tragus stem) were both found to have a much more obvious impact on the near-field phase than on the near-field amplitude. These effects were most pronounced in the region anterior of the tragus (see Fig. 10) and the farfield effects on the directivity pattern were mainly manifest in the side-lobes (see Fig. 7). Both results are similar to findings in the Big Brown Bat (*Eptesicus fuscus*), where the position of the tragus was found to affect the formation of a sidelobe by shifting the phase in the near-field (Müller, 2004). The results in the Chinese Noctule as well as in the Big Brown Bat are both consistent with the argument that, since the structural detail of the bat pinna (tragus, ledges, etc.) is on a similar size scale as the wavelength, shielding of portions of the incoming wave is not a feasible functional mechanism, as these structures are unable to cast a significant acoustic shadow for the wavelengths in question.

V. FUTURE WORK

The work reported here has demonstrated that the ear of the Chinese Noctule is capable of generating an unusual helical scanning pattern. Furthermore, it has shown that prominent structural features of the ear—the tragus and the thick lower ledge of the pinna rim—exert a crucial influence on beam pattern by shifting the phase of near-field. However, the acoustic mechanism behind the periodicity in the beam pattern over frequency still needs to be established. Because the periodicity does only reproduce qualitative features of the directivity pattern (e.g., the presence of a main lobe or a sleeve of sidelobes), but not exactly the same directivity, pattern recognition methods that can pick up correlated properties in the near-field phase and the directivity need to be developed and tested to determine whether they can address this question. Once correspondences between near-field (phase) patterns and directivity patterns have been established, they can be used as a basis for investigating the mechanisms.

The configurations at which the left and right ear were mounted for scanning represent only two samples from the entire pinna shape configuration space, which the animal may have at its disposal when it controls the shape of its pinna through the action of muscles inserting in the area. A more extensive probing of this space is highly desirable in order to obtain a better understanding of which options the animal has, in principle, and how transitions between them can be made. This question could be addressed by mounting and scanning pinna samples in a larger number of different configurations, or—more conveniently—by deforming the digital ear shape models nonrigidly. In the latter approach, actual scanned ear shape configurations could be used as reference points in the pinna shape space from which intermediate configurations could be reached by deforming an elastic shape model. Potentially suitable methods that allow nonrigid shape deformation and concurrent identification of corresponding landmark points to support the generation of intermediate shapes are already available (Sclaroff and Pentland, 1995).

ACKNOWLEDGMENTS

This work has been supported by the European Commission, Future and Emerging Technologies (projects CIRCE and CILIA—Contracts No. IST-2001-35144 and No. 016039) and a grant from Shandong University. Shuyi Zhang is supported by the National Geographic Society (Grant No. 7806-05). Professor Ling Qin and Mr. Benson Yeung from the Department of Orthopaedics and Traumatology at the Chinese University of Hong Kong made their micro-CT facilities available for this work and helped with performing the micro-CT scans.

Astley, R. J. and Macaulay, G. J. (**1998**). "Three-dimensional wave-envelope elements of variable order for acoustic radiation and scattering. Part I. Formulation in the frequency domain," J. Acoust. Soc. Am. **103**, 49–63.

Aytekin, M., Grassi, E., Sahota, M., and Moss, C. F. (**2004**). "The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation," J. Acoust. Soc. Am. **116**, 3594–3605.

Balay, S., Gropp, W. D., McInnes, L. C., and Smith, B. F. (**1997**). "Efficient

management of parallelism in object oriented numerical software libraries," in *Modern Software Tools in Scientific Computing*, edited by E. Arge, A. M. Bruaset, and H. P. Langtangen (Birkhäuser Press, Boston, MA), pp. 163–202.

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., and der Vorst, H. V. (**1994**). *Templates For the Solution of Linear Systems: Building Blocks For Iterative Methods*, 2nd ed. (SIAM, Philadelphia, PA).

Dreyer, U. and von Estorff, O. (**2003**). "Improved conditioning of infinite elements for exterior acoustics," Int. J. Numer. Methods Eng. **58**, 933–953.

Firzlaff, U. and Schuller, G. (**2003**). "Spectral directionality of the external ear of the Lesser Spear-nosed Bat, *Phyllostomus discolor*," Hear. Res. **181**, 27–39.

Jackson, D. E. (**1999**). *Classical Electrodynamics*, 3rd ed. (Wiley, New York).

Muller, R. (**2004**). "A numerical study of the role of the tragus in the big brown bat," J. Acoust. Soc. Am. **116**, 3701–3712.

Müller, R. and Hallam, J. C. T. (**2005**). "Knowledge mining for biomimetic smart antenna shapes," Rob. Auton. Syst. **50**, 131–145.

Obrist, M., Fenton, M., Elger, J., and Schlegel, P. (**1993**). "What ears do for bats: A comparative study of pinna sound pressure transformation in chiroptera," J. Exp. Biol. **180**, 119–152.

Ramahi, O. M. (**1997**). "Near- and far-field calculations in fdtd simulations using Kirchhoff surface integral representation," IEEE Trans. Antennas Propag. **45**, 753–759.

Sclaroff, S. and Pentland, A. (**1995**). "Modal matching for correspondence and recognition," IEEE Trans. Pattern Anal. Mach. Intell. **17**, 545–561.

Spezio, M. L., Keller, C. H., Marrocco, R. T., and Takahashi, T. T. (**2000**). "Head-related transfer functions of the rhesus monkey," Hear. Res. **144**, 73–88.

Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).

van der Vorst, H. A. (**2003**). *Iterative Krylov Methods For Large Linear Systems*, Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge, UK).

Wang, Y., Luo, F., Li, G., Jones, G., Parsons, S., and Zhang, S. (in preparation). "Molecular phylogenetic status, echolocation calls and diet of the Chinese Noctule (*Nyctalus plancyi*)," Acta Chiropterologica.

Weisstein, E. W. (**2002**). *CRC Concise Encyclopedia of Mathematics*, 2nd ed. (CRC Press, Boca Raton, FL).

Westheimer, G. (**2004**). "Center-surround antagonism in spatial vision: Retinal or cortical locus?," Vision Res. **44**, 2457–2465.

Wotton, J. M., Haresign, T., Ferragamo, M. J., and Simmons, J. A. (**1996**). "Sound source elevation and external ear cues influence the discrimination of spectral notches by the Big Brown Bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **100**, 1764–1776.

Wotton, J. M., Haresign, T., and Simmons, J. A. (**1995**). "Spatially dependent acoustic cues generated by the external ear of the Big Brown Bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **98**, 1423–1445.

Correlation analysis of the beam angle dependence for elastography

Min Rao and Tomy Varghese[a)]
Department of Medical Physics, The University of Wisconsin-Madison, 1300 University Avenue, 1530 MSC, Madison, Wisconsin 53706

(Received 10 November 2005; revised 17 March 2006; accepted 20 March 2006)

Signal decorrelation is a major source of error in the displacements estimated using correlation techniques for elastographic imaging. Previous papers have addressed the variation in the correlation coefficient as a function of the applied compression for a finite window size and an insonification angle of zero degrees. The recent use of angular beam-steered radio-frequency echo signals for spatial angular compounding and shear strain estimation have demonstrated the need for understanding signal decorrelation artifacts for data acquired at different beam angles. In this paper, we provide both numerical and closed form theoretical solutions of the correlation between pre- and post-compression radio-frequency echo signals acquired at a specified beam angle. The expression for the correlation coefficient obtained is a function of the beam angle and the applied compression for a finite duration window. Accuracy of the theoretical results is verified using tissue-mimicking phantom experiments on a uniformly elastic phantom using beam-steered data acquisitions on a linear array transducer. The theory predicts a faster decorrelation with changes in the beam or insonification angle for longer radio-frequency echo signal segments and at deeper locations in the medium. Theoretical results provide useful information for improving angular compounding and shear strain estimation techniques for elastography.
 [DOI: 10.1121/1.2195290]

PACS number(s): 43.80.Qf [TDM] Pages: 4093–4101

I. INTRODUCTION

Elastography, a method for imaging the tissue stiffness, has gained interest for the diagnosis of disease in recent years.[1–16] This method estimates local strains typically in the direction of compression by computing the gradient of shifts in echo arrival times after a quasistatic tissue compression. The echo shifts or tissue displacement are computed using a one-dimensional gated cross-correlation analysis of the pre- and post-compression radio-frequency (RF) signals, with overlapping windows to improve resolution. Recent reports have presented approaches that estimate both the normal and shear strains in elastography.[16,17] However, the distortion of the echo signal as a result of the applied compression introduces decorrelation noise, which is a major source of error in strain imaging.[3,4]

Many algorithms have been developed to reduce decorrelation noise and to improve the elastographic signal-to-noise ratio (SNRe), such as temporal stretching,[18–20] multi-compression averaging,[21] and wavelet denoising.[22] Spatial-angular compounding for elastography was recently introduced by our group[23–25] to reduce noise artifacts in the resulting compounded elastograms. This method averages multiple weighted angular strain estimates around the same region-of-interest (ROI) acquired from different beam insonification angles. Elastograms generated from the pre- and post-compression signals acquired at these different beam angles are referred to as angular elastograms. However, as previously described,[24,25] when the ultrasound beam is steered away from the direction of compression, the angular elastograms suffer from larger echo signal decorrelation in the pre- and post-compression RF echo signals, when compared to the nonsteered condition, where the data acquisition and the direction of compression are the same. To investigate and optimize the performance of spatial-angular compounding for elastography, it would be helpful to have a priori knowledge regarding the noise properties of the compounded angular elastograms. The correlation coefficient has been utilized previously to ascertain the degree of signal decorrelation and to quantify improvements in strain imaging performance. In this paper, a theoretical formulation is therefore developed to account for the increased signal decorrelation between the pre- and post-compression RF signals acquired when the ultrasound beam is at a different angle with respect to the direction of compression.

To measure the performance of the cross correlation based strain estimator, a large amount of theoretical work has been reported in the literature to calculate the correlation coefficient between the pre- and post-compression RF signals.[26–30] The correlation coefficient with motion compensation due to the axial deformation of elastic tissue was derived by Meunier and Bertrand[26] using a two-dimensional (2D) Gaussian model. Cross-correlation functions for tissue like media that exhibit deterministic and stochastic strain profiles were derived by Bilgen and Insana.[27] To study the degradation in the elastographic image quality due to the lateral and elevational motion of the scatterers, a 3D model was proposed by Kallel and Ophir[28] for predicting the correlation coefficient using separable point-spread functions in

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

the axial, lateral, and elevational directions, respectively. An expression for the correlation coefficient that depends on both the applied compression and the finite size of the window used for the cross-correlation processing was developed by Varghese *et al.*[29] This formulation accounted for the intrawindow signal decorrelation due to the applied compression. The performance of the standard cross correlation approach used to estimate the complex motions and the resulting deformations in 3D was investigated by Bilgen.[30] However, all these models assume that the compression and ultrasound beam are along the same direction, and therefore are not suitable to compute the correlation coefficient for angular elastograms obtained by beam steering.

Many studies have also been performed to investigate spatial correlations in ultrasound speckle pattern motion in medical ultrasonic images.[31–35] All of these prior studies were directed toward B-mode image compounding, where the theories developed studied correlations between echo signals acquired from the same spatial location but different insonification angles to reduce speckle noise artifacts in the B-mode images. The analysis was recently extended by Chen *et al.*[36] to study the correlation of RF signals that intersect at the same spatial location obtained from different angular insonifications. This formulation is useful in evaluating the improvement in parametric images of the scatterer size or attenuation coefficient with spatial angular compounding, since these parameters are estimated from the angular RF data. Since the contribution of the insonification angle was addressed in these models we have modified this approach to investigate angular compounding for elastography by including the applied compression and the use of a finite sized processing window.

In this paper, initially following the framework used by Chen *et al.*,[36] a closed form expression is derived for the correlation coefficient between pre- and post-compression RF signals acquired at specific beam angles. To corroborate the theoretical development, experimental results using a tissue-mimicking phantom are presented that verify and validate the theoretical expressions. Factors that affect the signal decorrelation versus the beam angle are discussed. Both the rigorous numerical derivation and the closed form expressions for the correlation coefficient can be used to predict the rate of signal decorelation in the angular elastograms and provide a clear understanding of the performance of angular compounding techniques in elastography.

II. THEORY

Ultrasonic RF echo signals before and after compression is modeled by

$$s_1(x,z) = P(x,z) \otimes T(x,z),$$

$$s_2(x,z) = P(x,z) \otimes T(x,az), \tag{1}$$

where the subscripts 1 and 2, specify the pre- and post-compression echo signals obtained from an elastic tissue medium, $T(x,z)$ denotes the tissue scattering function, and $P(x,z)$ is the pulse-echo point-spread function (PSF) of the imaging system. The symbol $\otimes$ denotes the convolution operation. The parameter a is the compression or strain factor that scales the tissue scattering function defined in terms of the actual applied tissue strain ε, $a=1/(1-\varepsilon)\sim 1+\varepsilon$ for $\varepsilon \ll 1$. Here a 2D model is used because the ultrasound beam in currently available transducers is much wider in the elevational direction compared to the lateral direction. Thus, the scatterer movement in the elevational direction does not cause significant loss of coherence in the echoes.

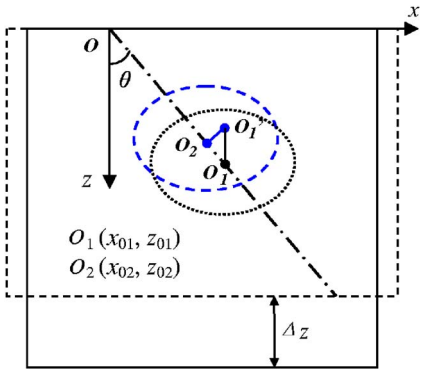


FIG. 1. (Color online) Schematic diagram illustrating the acquisition of pre- and post-compression RF signals with a beam steering angle θ. The precompression RF signals are obtained from scatterers at position O_1, while the post-compression signals are obtained from the corresponding scatterers at location O_2.

There are several models that define the $T(x,z)$ term. To simplify the tissue model, we assume a large number of very small inhomogeneities (Rayleigh scatterers) with respect to the wavelength of the PSF[26]

$$T(x,z) = \sum_i T_i \delta(x - x_i, z - z_i), \tag{2}$$

where $\delta(x,z)$ is the 2D Dirac or impulse function, (x_i,z_i) denotes the randomly distributed centers of each inhomogeneity, and T_i the echogenicity of each scatterer. The scatterer distribution is assumed to be δ correlated, which means the correlation length of scatterers is very short compared to the acoustic wavelength of the transmit pulse.

The cumulative signal amplitude from scatterers at position O_1 (x_{01},z_{01}) in the precompression medium can be written as[31]

$$s_1 = \sum_i |T_i||P_{1,i}|\exp(j\vartheta_i). \tag{3}$$

The subscript i refers to an individual scatterer, $|P|$ is the magnitude of the pulse-echo PSF from the scatterers, and $\exp(j\vartheta)$ represents the combined phase of T and P. The phase factors in ϑ are uniformly distributed over 2π radians with zero mean. An illustration of this concept is shown in Fig. 1. We establish our coordinates by setting the position of transducer as the origin, and the beam steering angle is θ.

A quasistatic compression is then applied to the medium along the $-z$ direction, with the scatterers moving toward the transducer. For simplicity, we assume the position O_1 is at the center of the medium in the lateral direction, which implies that the scatterer movement in the lateral direction is small enough to be considered negligible. Thus, tissue scat-

terers around O_1 move to position O_1' after compression, as shown in Fig. 1. The post-compression signal is then generated from scatterers at position O_2 (x_{02}, z_{02}), which are the projection positions of O_1' on the ultrasound beam. The cumulative signal strength after compression can be written as

$$s_2 = \sum_i |T_i||P_{2,i}|\exp(j\vartheta_i)\exp(j\Delta\phi_i), \tag{4}$$

where $\Delta\phi_i$ is the phase difference between pre- and post-compression signals, which can be written as $4\pi(R_{2,i} - R_{1,i})/\lambda_0$, where λ_0 is the wavelength at the center frequency, and $R_{1,i}$, $R_{2,i}$ are the distances from the ith scatterer to the transducer for the pre- and post-compression situations, respectively,

$$R_{1,i} = \sqrt{x_i^2 + z_i^2},$$

$$R_{2,i} = \sqrt{x_i^2 + (z_i/a)^2}, \tag{5}$$

$$R_{2,i} - R_{1,i} = \frac{1 - a^2}{a^2(R_{1,i} + R_{2,i})} z_i^2.$$

The cross correlation between the signals acquired before and after compression can be written as[31]

$$\langle s_1 s_2^* \rangle = \sum_i |T_i|^2 |P_{1,i}||P_{2,i}|\exp(j\Delta\phi_i). \tag{6}$$

The 2D PSF in the focal zone of each ultrasound beam of a linear array transducer can be expressed in the following form:

$$|P(x,z)| = p_x(x)p_z(z), \tag{7}$$

where p_x represents the lateral beam spread function, and p_z represents the axial spread function. For a rectangular aperture, for example, the lateral PSF for a pulse-receive response at the focus or in the far-field can be written as

$$p_x(x) = \sin^2(\pi f_0 x)/(\pi f_0 x)^2, \tag{8}$$

where $f_0 = D/r\lambda_0$, D is the effective transducer aperture, and r is the focal distance, $r = z_{01,02}/\cos\theta$. If we assume that the ultrasound pulse transmitted by the transducer has a Gaussian envelope with characteristic width σ_z, then p_z can be expressed as

$$p_z(z) = \exp(-z^2/2\sigma_z^2). \tag{9}$$

For the i^{th} scatterer, the lateral distances to the beam axis for the pre- and post-compression cases can be written as

$$l'_{1,i} = (x_i - x_{01})\cos\theta - (z_i - z_{01})\sin\theta,$$

$$l'_{2,i} = (x_i - x_{02})\cos\theta - (z_i/a - z_{02})\sin\theta. \tag{10}$$

Similarly, the axial distances from the i^{th} scatterer to the center point are

$$l''_{1,i} = (x_i - x_{01})\sin\theta + (z_i - z_{01})\cos\theta,$$

$$l''_{2,i} = (x_i - x_{02})\sin\theta + (z_i/a - z_{02})\cos\theta. \tag{11}$$

Hence, the PSF for the i^{th} scatterer can be written as

$$P_{k,i} = p_x(l'_{k,i})p_z(l''_{k,i}), \quad (k = 1,2). \tag{12}$$

Substituting Eq. (12), into Eq. (6), we obtain

$$\langle s_1 s_2^* \rangle = \sum_i |T_i|^2 |p_x(l'_{1,i})||p_z(l''_{1,i})||p_x(l'_{2,i})||p_z(l''_{2,i})| \times \exp(j\Delta\phi_i). \tag{13}$$

Going to a continuous representation, x_i and z_i become x and z, and the summation becomes a 2D integral about x and z. Thus,

$$\langle s_1 s_2^* \rangle = B' \int\int |p_x(l'_1)||p_z(l''_1)||p_x(l'_2)||p_z(l''_2)| \times \exp(j\Delta\phi)dxdz, \tag{14}$$

where B' is a normalization factor. An analytical closed form solution for the above equation is difficult to obtain. Thus, it is necessary to either resort to numerical solutions or apply approximations to simplify the expression in Eq. (14).

To simplify the 2D integral in Eq. (14), we utilize a Gaussian envelope to model the lateral PSF as shown

$$p_x(x) = \exp(-x^2/2\sigma_x^2). \tag{15}$$

For small beam angles ($\theta < 10°$), $\sin\theta \ll \cos\theta$, the $\sin\theta$ term in Eqs. (10) and (11) can be ignored. The phase term $\exp[4\pi j(R_2 - R_1)/\lambda_0]$ in Eq. (14) is a function of x and z. Since only scatterers close to the region of interest contribute to the signal, we can approximate the phase term as $\exp[2\pi jz\cos\theta(1-a^2)/a^2\lambda_0]$. We can then perform the integration in the lateral and axial direction separately as discussed by Wagner *et al.*[31] and obtain

$$\langle s_1 s_2^* \rangle = B'' \int |p_z[(z - z_{01})\cos\theta]||p_z[(z/a - z_{02})\cos\theta]|\exp\left[2\pi jz\cos\theta\frac{1-a^2}{a^2\lambda_0}\right]dz, \tag{16}$$

where B'' is the integral in the lateral direction

$$B'' = \frac{\sigma_x}{\cos\theta}\sqrt{\pi}\exp\left[-\frac{1}{4}\frac{\cos^2\theta}{\sigma_x^2}(x_{01} - x_{02})^2\right]. \tag{17}$$

Equation (16) follows the notation of a Fourier transformation (FT). The cross-correlation function is therefore a FT of the multiplication of the two PSF

$$\langle s_1 s_2^* \rangle = B''[FT\{|p_z[(z - z_{01})\cos\theta]||p_z[(z/a - z_{02})\cos\theta]|\}]_{f=((1-a^2)/a^2\lambda_0)\cos\theta}, \tag{18}$$

where FT denotes the Fourier transformation operator. The correlation coefficient is defined as[27]

$$\rho = \frac{\langle s_1 s_2^* \rangle}{\sqrt{\langle s_1 s_1^* \rangle \langle s_2 s_2^* \rangle}}. \tag{19}$$

Substituting Eq. (18) into Eq. (19) and after the integration procedure we obtain

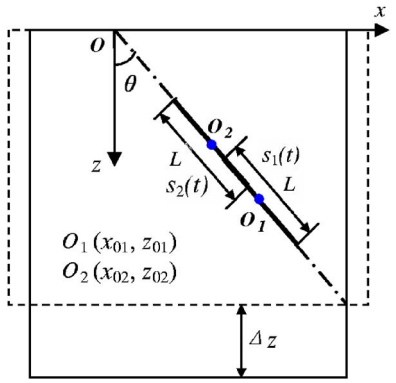


FIG. 2. (Color online) Schematic illustration of the angular RF data segment acquired before and after compression.

$$\rho = \sqrt{\frac{2a}{a^2+1}} \exp\left\{ -\frac{\cos^2\theta}{4}\left[\frac{(x_{01}-x_{02})^2}{\sigma_x^2} + \frac{(z_{01}-az_{02})^2}{\sigma_z^2} \right] \right\} \exp\left(-\frac{\sigma_z^2\pi^2 f^2}{\cos^2\theta} \right)_{f=((1-a^2)/a^2\lambda_0)\cos\theta}. \tag{20}$$

For elastographic processing, finite gated segments of the echo signal are utilized. The displacement at location O_1 is estimated from RF segments illustrated with thick lines, as shown in Fig. 2. A gated data segment is generally selected using a rectangular window. Generally, the same window is used for both pre- and post-compression data segments. Attenuation and focusing effects can be included in the window function $w(t)$, so that the intensity of $s(t)$ can be taken to be constant with depth. Thus, the correlation coefficient between the pre- and post-compression signals can be written as

$$\rho_{1,2} = \frac{\int_{t_1}^{t_2} w^2(t)\langle s_1(t)s_2^*(t)\rangle/\bar{I}dt}{\int_{t_1}^{t_2} w^2(t)dt} = \frac{\int_{-L/2}^{L/2} w^2(\xi)\langle s_1(\xi)s_2^*(\xi)\rangle/\bar{I}d\xi}{\int_{-L/2}^{L/2} w^2(\xi)d\xi} = \frac{\int_{-L/2}^{L/2} w^2(\xi)\rho(\xi)d\xi}{\int_{-L/2}^{L/2} w^2(\xi)d\xi}, \tag{21}$$

where L is the window length for the RF echo signal segment, $\bar{I}=\langle s\cdot s^*\rangle$ is the mean signal intensity. ξ is the distance from scatterers in the volume giving rise to the RF segment to the point of interest. The ρ parameter inside the integral can be either the numerical result of Eq. (14), or its approximated version in Eq. (20).

Figure 3 shows a comparison of the correlation coefficient obtained using the numerical integration of Eq. (14) and the approximated expression using Eq. (20). The following parameter values were used for the calculations: $L=3$ mm, $D=1.8$ cm, center frequency $f_c=5$ MHz with a 50% bandwidth, and $z_{01}=3$ cm. To approximate the lateral PSF, we used $\sigma_x=0.2$ mm in Eq. (15). As shown in Fig. 3, the approximated curves (dashed lines) are very close to the numerical results (solid lines) at small beam angles ($<10°$). However, the deviation of the approximated results becomes larger as the beam angle increases. This result demonstrates that the approximation is accurate only for the small beam angle, and has to be used with caution to model the decorrelation from larger beam angles.

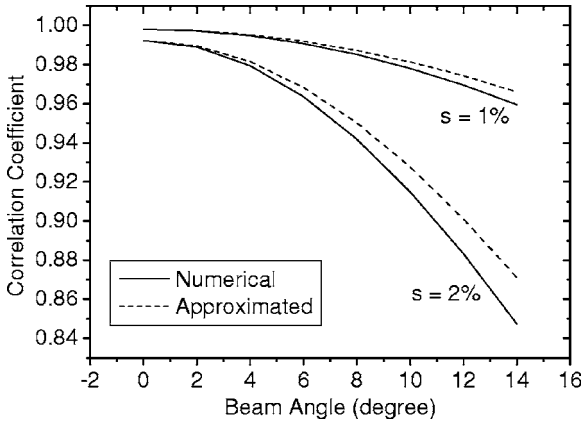


FIG. 3. Comparison between the exact numerical and approximated theoretical expression for the correlation coefficient. The correlation coefficient curves are plotted versus the beam angle.

III. EXPERIMENTAL VALIDATION

A. Method

The theoretical results derived in the previous section are verified and validated using experimental data acquired using a clinical ultrasound scanner. Pre- and post-compression RF data were obtained using a uniformly elastic tissue-mimicking phantom of size $10\times10\times10$ cm^3, manufactured in our laboratory.[37] The phantom was scanned using an Ultrasonix 500RP (Ultrasonix Medical Corporation, Bothell, WA and Vancouver, BC, Canada) real-time scanner equipped with a 5 MHz linear-array transducer with an approximate 60% bandwidth. The Ultrasonix 500RP is equipped with an ultrasound research interface (URI) that readily allows an expert programmer to alter and change the standard operating conditions and introduce new echo signal processing techniques. The Ultrasonix system setup used in our experiment enables beam steering within the angle range from −15° to 15°. In order to acquire RF data at different beam angles, we developed a URI client program, to communicate with the Ultrasonix URI and software server to control the beam steering algorithm. The URI client program enables the operator to input the maximum angle and the angular increment, and the machine will automatically scan the phantom at the specified angles during the angular sweep. In our experiment, the phantoms were scanned from 0° to 15° with a minimum angular increment of 0.75°.

The stepper motor controlled quasistatic compression of the phantom is also controlled by the URI client program on the Ultrasonix 500RP system. The program controls the stepper motor apparatus that enables synchronized acquisition of both the pre- and post-compression RF data sets after a

quasi-static compression. A compression plate with a rectangular slot fitted with the transducer was mounted on a linear stage (Velmex Inc., Bloomfield, NY) driven by a stepper motor. The compression plate is larger than the phantom surface to provide a uniform compression of the phantom. The compression was applied along the axial direction, i.e., angle 0°, without beam steering. Echo signals were acquired, from a 40×40 mm^2 ROI (starting at a depth of 10 mm under the transducer) for applied strains ranging from 0.5 to 3% at steps of 0.5%.

We have implemented an automated beam-steering and data acquisition algorithm on the Ultrasonix system. The algorithm first acquires the precompression data along the specified angular sweep (0° to 15°) at 0.75° increments. The stepper motor is then activated to compress the phantom at a specified compression increment (0.5%), following which the post-compression RF data are acquired following the same angular scan sequence. In order to obtain statistically independent results, pre- and post-compression data were acquired over ten independent realizations along the different scanning planes and at different pre-compression levels.

Each of the angular pre- and post-compression RF frames acquired were analyzed separately to calculate the correlation coefficient at the specified angle. Cross-correlation analysis using a window size of 3 mm and 75% overlap of consecutive windows was used to generate the normalized cross-correlation function. The correlation coefficient is obtained from the peak of the normalized cross-correlation function.

B. Computation of the effective aperture

The actual aperture of the transducer during RF data acquisition was not available during the experiment. Therefore, we calculated the correlation between two parallel RF echo signal segments for the zero-angle frame to estimate the PSF of the system, which in turn would yield an effective aperture estimate.[36] We selected a 4 mm RF segment centered at a depth of 3 cm for each beam line and calculated the correlation coefficient value versus the beam line separation as shown in Fig. 4. The error bars denote the standard deviation of the mean estimates over 20 different lateral positions. Since the transmit focus is set to a depth of 1.0 cm with a dynamic receive focus, the transmit PSF at a 3-cm depth is significantly wider than the receive PSF. The combined PSF, a product of transmit and receive PSF, can be approximated by the receive PSF, which can be written as

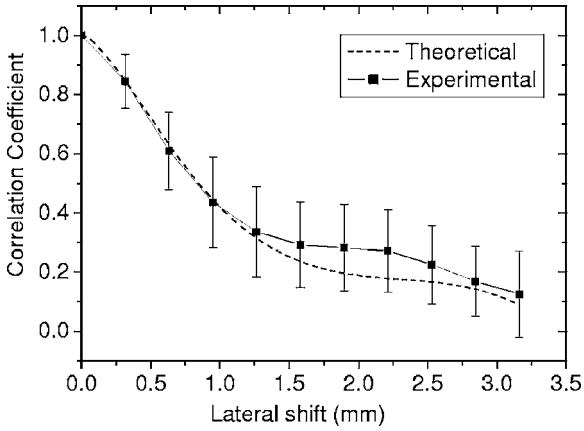


FIG. 4. Correlation coefficient for echo data obtained from parallel RF beam lines. The solid line is the measured correlation coefficient estimate and the dashed curve denotes the approximated theoretical curve generated by assuming a transducer aperture size of 8 mm.

$$p_x(x) = \sin(\pi f_0 x)/(\pi f_0 x). \tag{22}$$

Note the difference between Eqs. (22) and (8), where we assume that the transmit and receive foci are located at the same position as the origin of the echo signal; therefore, the PSF in that case would be the product of two identical sinc functions.

We use the method described by Chen *et al.*[36] to compute the correlation between two parallel beam lines and to determine the effective aperture. We estimated that for an effective receive aperture size of $D=8$ mm, the theoretical prediction fits the experimental data (Fig. 4), where the best fit curve is plotted as a dashed line. With this effective aperture size, we can predict the decorrelation of the pre- and post-compression RF signal segments with an increase of beam angle, and compare the theoretical predictions with the experimental data.

IV. RESULTS

To verify the accuracy of the proposed theory, we use experimental RF data to compute estimates of the correlation coefficient and to compare these results to the theoretical prediction. We selected a location in the phantom at a depth of 3 cm along the central beam line for a 0 deg insonification angle. For each subsequent insonification angle, we then locate the beam line that intersects or passes closest to that point and chose the RF data segment centered at that point. A rectangular window was used to segment or gate the RF data.

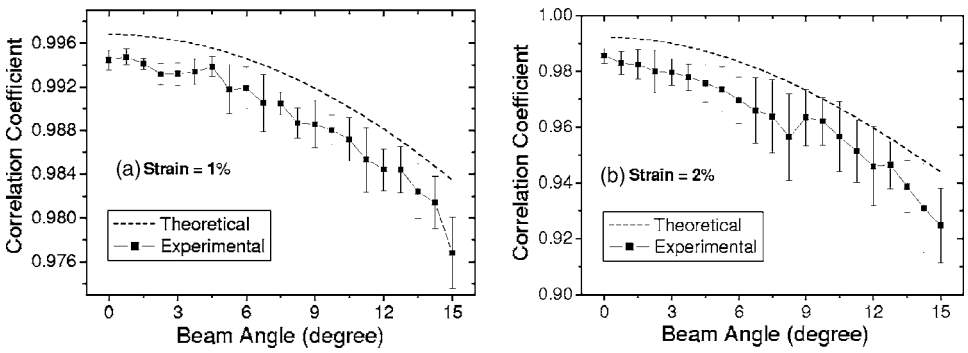


FIG. 5. Comparison between theoretical prediction and experimental results for the correlation coefficient of pre- and post-compression RF segments acquired at different beam angles. Results are shown for applied strains of (a) 1%, and (b) 2%. The correlation coefficient was obtained for a 5 MHz center frequency, using a 3-mm RF data segment centered at a depth of 3 cm.

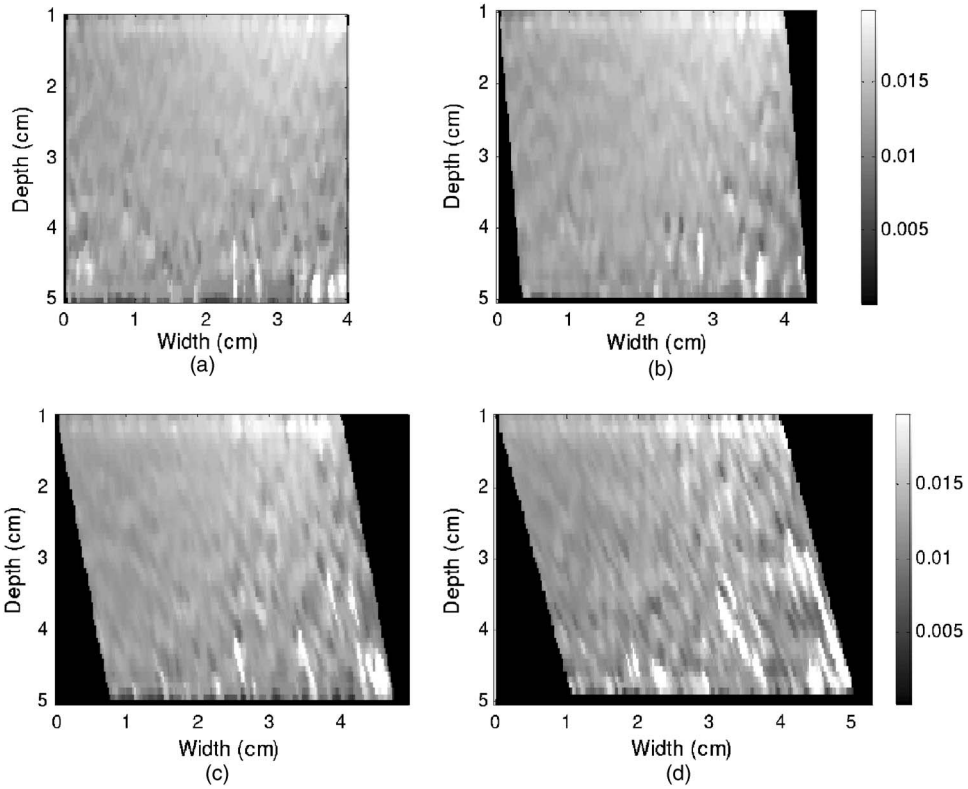


FIG. 6. Angular strain images obtained at (a) 0°, (b) 4.5°, (c) 10.5°, and (d) 15° angles under an applied strain of 1.5%. The colorbar denotes strain, where 1% strain is displayed as 0.01.

Since the window length was selected to be 3 mm and signal intensity variations due to attenuation and focusing effects are ignored, the window function $w(t)$ can be set to 1. The resultant correlation coefficient curves are plotted as a function of the beam angle in Fig. 5 for applied compressions of (a) 1% and (b) 2%. The error bars denote the standard deviation of the mean estimates over ten independent data sets. The theoretical prediction, obtained by numerically calculating Eq. (14), is plotted as a dashed line. Note that the theoretical curve is just above the errorbars of the experimental curve in most beam angle cases. This is because our theoretical model assumes that the displacement along the beam direction has been accurately tracked, in which case the correlation coefficient achieves the maximum value at that beam angle. The experimental estimates of the displacement, however, always contain errors due to the motion tracking algorithm and sampling precision. Therefore, the correlation coefficient obtained from experiment is lower than the theoretical prediction.

The reduction in the correlation coefficient with beam angle manifests as increased decorrelation errors leading to noisy estimation of the displacement and subsequently the local strain in the angular strain images. Examples of angular strain images obtained from the uniformly elastic phantom are shown in Fig. 6, for angles of 0°, 4.5°, 10.5°, and 15°. Note the presence of increased noise artifacts in the strain images with an increase in the insonification or beam angle.

V. DISCUSSION

Spatial-angular compounding can reduce noise artifacts in elastograms, but at the expense of the additional processing time required for displacement and strain estimations along angular directions. The most efficient spatial compounding procedure would be obtained by averaging independent, uncorrelated angular elastograms or strain estimates.[33] The effective number of independent strain estimates obtained depends on two factors: the angular incre-

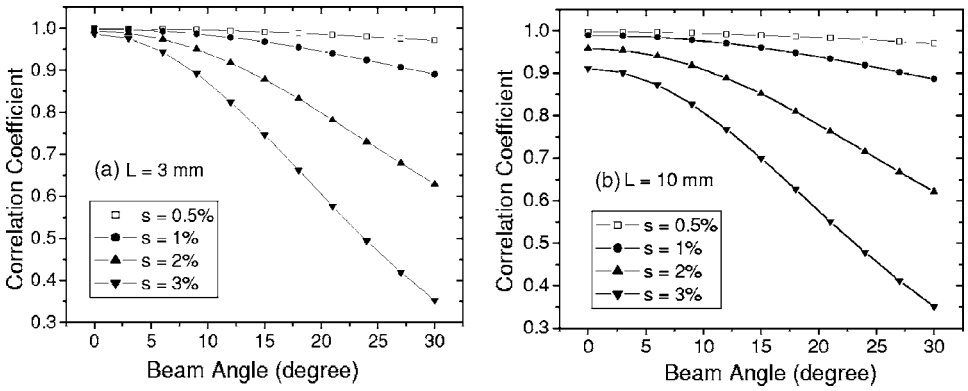


FIG. 7. Plots of the theoretical prediction of the correlation coefficient versus the beam angle for different applied strain and RF data segment lengths of (a) 3 mm and (b) 10 mm. The RF data segment was centered at a 3-cm depth.

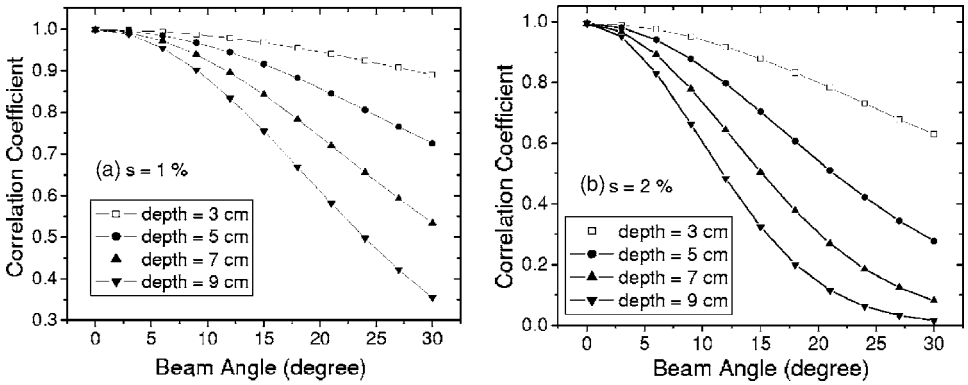


FIG. 8. Plots of the theoretical prediction of the correlation coefficient versus the beam angle for RF data segments obtained at different depths. Results are shown for applied strains of (a) 1%, and (b) 2%. A 3-mm RF segment length is used in the calculation.

ment and the maximum angle used for angular compounding. Generally, increasing the maximum angle for compounding reduces the strain variance and improves the SNRe. However, increased decorrelation noise artifacts are present for the elastograms obtained over larger insonification angles as illustrated in this paper. It is useful to understand how the beam angle and other system factors affect the correlation between pre- and post-compression signals.

Figure 7 shows approximated theoretical correlation coefficient curves as a function of beam angle using RF data segment lengths of (a) 3 mm, and (b) 10 mm, for applied strain values of 0.5, 1, 2, and 3 %, respectively. Here the RF data segment was assumed to be centered at a 3-cm depth in the phantom, and the aperture of the transducer was computed to be 1.5 cm. As illustrated in the figure, the decorrelation rate of the RF signal pairs increase with the value of the applied strain. This increased decorrelation rate is due to the fact that with increased strain more scatterers would leave the beam, leading to increased decorrelation, especially at the larger beam angles. Note that this is also the reason that the applied strain has a smaller effect on the correlation coefficient at smaller beam angles, as shown in Fig. 7. Variations in the correlation coefficient for different gated RF signal segment lengths are also observed in Figs. 7(a) and 7(b). As expected, when echo signals that are further away from the center of the initial data segment are included within the gated segment, increased signal decorrelation occurs between the pre- and post-compression signals. This effect is more pronounced for the smaller beam angles and reduces with an increase in the beam angle. This is due to the fact that at the larger beam angles the major contribution to the signal decorrelation comes from the scatterers moving out of beam and the effect of the data segment length alone appears to be negligible.

Figure 8 shows the theoretical prediction for the correlation coefficient curves for RF signals obtained at different depths. Results are shown for applied strains of (a) 1%, and (b) 2%. A 3-mm RF segment length was used in the calculation, and the aperture of the transducer was dynamically changed to enable the beam to be focused at the different depths. As expected, deeper segments of RF signals decorrelate more rapidly with corresponding increases in the beam angle. This result is explained by the geometric location of the scatterers and beam lines before and after compression, illustrated in Fig. 1. Considering the scatterers that contribute to the θ-angled RF segment centered at point O_1, after compression, these scatterers move to the locations centered at point O_1', which is essentially outside the beamwidth, leading to signal decorrelation. The distance that the scatterers move increases with both the depth and the applied strain. Therefore, the correlation coefficient decreases at a faster rate at deeper locations with larger applied strains. Note that an increased depth does not have an appreciable impact on signal decorrelation at the 0-deg beam angle since the scatterers remain within the beam as they are displaced or deformed along the axial direction.

Figure 9 shows plots of the correlation coefficient versus the beam angle for different insonification frequencies. Results are obtained using 3-mm RF segments centered at a depth of 3 cm for 1% applied strain. As illustrated in the figure, the signal decorrelation rate of RF signal pairs increases with the center frequency. This is primarily due to the fact that the ultrasound beam becomes narrower with the increased center frequency, enabling scatterers that were within the precompression beam to leave the beam and newer scatterers to come within the beam after compression when the beam becomes narrower, especially at larger beam angles. A similar increase in the decorrelation rate is observed when the aperture of the transducer is changed to obtain a narrower beam. Figure 10 plots the correlation coefficient curves for different transducer apertures for a 5 MHz center or insonification frequency. A 3-mm RF data segment centered at a 3-cm depth with 1% applied strain, is

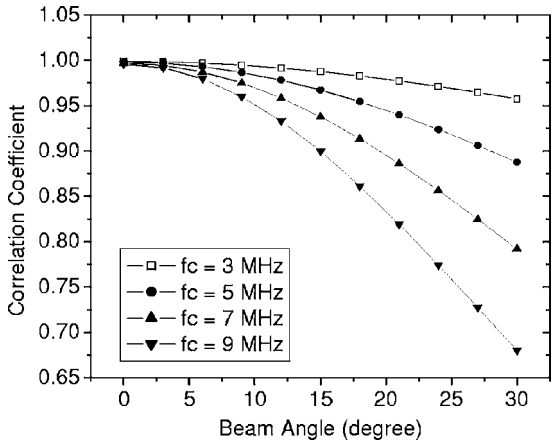


FIG. 9. Plots of the theoretical prediction of the correlation coefficient versus the beam angle for different center frequencies. Results are obtained using 3-mm RF segments centered at a depth of 3 cm for 1% applied strain, and a transducer aperture of 1.5 cm.

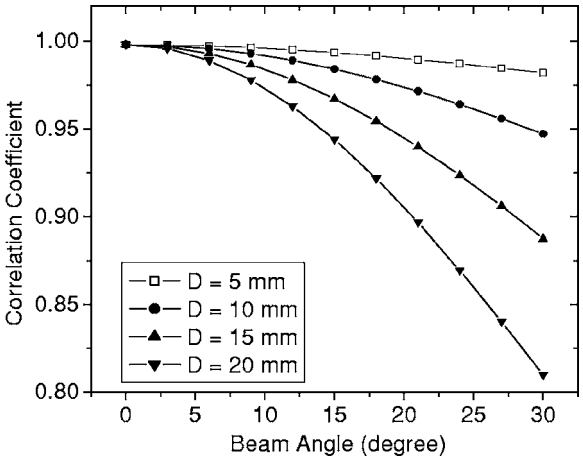


FIG. 10. Plots of the theoretical prediction of the correlation coefficient versus the beam angle for different transducer aperture sizes. Results are obtained using 3-mm RF segments centered at a depth of 3 cm for 1% applied strain, and a center frequency of 5 MHz.

utilized for Fig. 10. For larger apertures, the correlation coefficient curve falls off quickly with beam angle due to the narrower beam.

VI. CONCLUSIONS

A theoretical expression has been derived for the signal decorrelation between pre- and post-compression RF echo signals obtained at different angular insonifications, using beam steering to obtain angular elastograms. This work is based on previous theoretical results presented by Chen *et al.*,[36] where the analysis of the correlation between RF signals acquired from the same location but at different angles were presented. The analysis is extended to study the correlation between pre- and post-compression RF signals at a specified beam angle. The theoretical prediction matches well with the experimental results obtained using beam steering on a linear array transducer. The theoretical results derived in this paper is useful for evaluating the strain estimation performance of angular compounding and shear strain estimation using angular RF data[22,24] in elastography. For angular compounding and shear strain imaging, the theoretical results would enable the estimation of optimal maximum insonification angle to streamline the data processing. To find the optimum angular increment, more work is required to analyze the correlation between the angular elastograms obtained at different angles.

ACKNOWLEDGMENTS

This work is supported in part by NIH Grant No. R21 EB003853. The authors would also like to thank Dr. Laurent Pelissier for the loan of the Ultrasonix 500 RP system used on this research.

[1]J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography—A quantitative method for imaging the elasticity of biological tissues," Ultrason. Imaging **13**, 111–134 (1991).
[2]R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic-resonance elastography by direct visualization of propagating acoustic strain waves," Science **269**, 1854–1857 (1995).
[3]J. Ophir, S. K. Alam, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, and T. Varghese, "Elastography: Ultrasonic estimation and imaging of the elastic properties of tissues," Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **213**, 203–233 (1999).
[4]T. Varghese, J. Ophir, E. Konofagou, F. Kallel, and R. Righetti, "Tradeoffs in elastographic imaging," Ultrason. Imaging **23**, 216–248 (2001).
[5]A. Pesavento, A. Lorenz, S. Siebers, and H. Ermert, "New real-time strain imaging concepts using diagnostic ultrasound," Phys. Med. Biol. **45**, 1423–1435 (2000).
[6]I. Cespedes, J. Ophir, H. Ponnekanti, and N. Maklad, "Elastography: Elasticity imaging using ultrasound with application to muscle and breast in vivo," Ultrason. Imaging **15**, 73–88 (1993).
[7]M. O'Donnell, A. R. Skovoroda, B. M. Shapo, and S. Y. Emelianov, "Internal displacement and strain imaging using ultrasonic speckle tracking," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 314–325 (1994).
[8]K. J. Parker, S. R. Huang, R. A. Musulin, and R. M. Lerner, "Tissue response to mechanical vibrations for sonoelasticity imaging," Ultrasound Med. Biol. **16**, 241–246 (1990).
[9]M. Bertrand, M. Meunier, M. Doucet, and G. Ferland, "Ultrasonic biomechanical strain gauge based on speckle tracking," *IEEE Ultrasonics Symposium* (Montreal, Quebec, Canada, 1989) pp. 859–864.
[10]T. A. Krouskop, D. R. Dougherty, and F. S. Vinson, "A pulsed Doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue," J. Rehabil. Res. Dev. **24**, 1–8 (1987).
[11]M. F. Insana, L. T. Cook, M. Bilgen, P. Chaturvedi, and Y. Zhu, "Maximum-likelihood approach to strain imaging using ultrasound," J. Acoust. Soc. Am. **107**, 1421–1434 (2000).
[12]K. Nightingale, M. Scott Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: in vivo demonstration of clinical feasibility," Ultrasound Med. Biol. **28**, 227–235 (2002).
[13]A. Pesavento, A. Lorenz, S. Siebers, and H. Ermert, "New real-time strain imaging concepts using diagnostic ultrasound," Phys. Med. Biol. **45**, 1423–1435 (2000).
[14]D. B. Plewes, I. Betty, S. N. Urchuk, and I. Soutar, "Visualizing tissue compliance with MR imaging," J. Magn. Reson Imaging **5**, 733–738 (1995).
[15]E. Konofagou and J. Ophir, "A new elastographic method for estimation and imaging of lateral displacements, lateral strains, corrected axial strains and Poisson's ratios in tissues," Ultrasound Med. Biol. **24**, 1183–1199 (1998).
[16]E. E. Konofagou and J. Ophir, "Precision estimation and imaging of normal and shear components of the 3D strain tensor in elastography," Phys. Med. Biol. **45**, 1553–1563 (2000).
[17]U. Techavipoo, Q. Chen, T. Varghese, and J. A. Zagzebski, "Estimation of displacement vectors and strain tensors in elastography using angular insonifications," IEEE Trans. Med. Imaging **23**, 1479–1489 (2004).
[18]S. K. Alam, J. Ophir, and E. E. Konofagou, "An adaptive strain estimator for elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 461–472 (1998).
[19]S. K. Alam and J. Ophir, "Reduction of signal decorrelation from mechanical compression of tissues by temporal stretching: Applications to elastography," Ultrasound Med. Biol. **23**, 95–105 (1997).
[20]I. Cespedes and J. Ophir, "Reduction of image noise in elastography," Ultramicroscopy **15**, 89–102 (1993).
[21]T. Varghese, J. Ophir, and I. Cespedes, "Noise reduction in elastograms using temporal stretching with multicompression averaging," Ultrasound Med. Biol. **22**, 1043–1052 (1996).
[22]U. Techavipoo and T. Varghese, "Wavelet denoising of displacement estimates in elastography," Ultrasound Med. Biol. **30**, 477–491 (2004).
[23]U. Techavipoo and T. Varghese, "Improvements in elastographic contrast-to-noise ratio using spatial-angular compounding," Ultrasound Med. Biol. **31**, 529–536 (2005).
[24]U. Techavipoo, Q. Chen, T. Varghese, J. A. Zagzebski, and E. L. Madsen, "Noise reduction using spatial-angular compounding for elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 510–520 (2004).
[25]M. Rao, Q. Chen, H. Shi, and T. Varghese, "Spatial-angular compounding for elastography using beam steering on linear array transducers," Med. Phys. **33**, 618–626 (2006).
[26]J. Meunier and M. Bertrand, "Ultrasonic texture motion analysis—Theory and simulation," IEEE Trans. Med. Imaging **14**, 293–300 (1995).
[27]M. Bilgen and M. F. Insana, "Deformation models and correlation analysis in elastography," J. Acoust. Soc. Am. **99**, 3212–3224 (1996).
[28]F. Kallel and J. Ophir, "Three-dimensional tissue motion and its effect on

image noise in elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1286–1296 (1997).
[29]T. Varghese, M. Bilgen, and J. Ophir, "Multiresolution imaging in elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 65–75 (1998).
[30]M. Bilgen, "Dynamics of errors in 3D motion estimation and implications for strain-tenser imaging in acoustic elastography," Phys. Med. Biol. **45**, 1565–1578 (2000).
[31]R. F. Wagner, M. F. Insana, and S. W. Smith, "Fundamental correlation lengths of coherent speckle in medical ultrasonic images," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 34–44 (1988).
[32]R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," IEEE Trans. Sonics Ultrason. **30**, 156–163 (1983).
[33]M. Odonnell and S. D. Silverstein, "Optimum displacement for compound image generation in medical ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 470–476 (1988).
[34]S. W. Smith, R. F. Wagner, J. M. Sandrik, and H. Lopez, "Low contrast detectability and contrast detail analysis in medical ultrasound," IEEE Trans. Sonics Ultrason. **30**, 164–173 (1983).
[35]C. B. Burckhardt, "Speckle In Ultrasound B-Mode Scans," IEEE Trans. Sonics Ultrason. **25**, 1–6 (1978).
[36]Q. Chen, A. L. Gerig, U. Techavipoo, J. Zagzebski, and T. Varghese, "Correlation of RF Signals During Angular Compounding," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 961–970 (2005).
[37]E. L. Madsen, G. R. Frank, T. A. Krouskop, T. Varghese, F. Kallel, and J. Ophir, "Tissue-mimicking oil-in-gelatin dispersions for use in heterogeneous elastography phantoms," Ultrason. Imaging **25**, 17–38 (2003).