CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 121, No. 4 April 2007

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

SPEECH PERCEPTION [71]

(Continued)

CONTENTS—*Continued from preceding page*

Applying nonlinear resonant ultrasound spectroscopy to improving thermal damage assessment in concrete

C. Payan, V. Garnier, and J. Moysan
Laboratoire de Caractérisation Non Destructive, Université de la Méditerranée, IUT Aix-en-Provence, Avenue Gaston Berger, 13625 Aix-en-Provence Cedex 1, France
cedric.payan@univmed.fr, garnier@iut.univ-aix.fr, moysan@iut.univ-aix.fr

P. A. Johnson
Geophysics Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 875454, USA
paj@lanl.gov

Abstract: Nonlinear resonant ultrasound spectroscopy (NRUS) consists of evaluating one or more resonant frequency peak shifts while increasing excitation amplitude. NRUS exhibits high sensitivity to global damage in a large group of materials. Most studies conducted to date are aimed at interrogating the mechanical damage influence on the nonlinear response, applying bending, or longitudinal modes. The sensitivity of NRUS using longitudinal modes and the comparison of the results with a classical linear method to monitor progressive thermal damage (isotropic) of concrete are studied in this paper. In addition, feasibility and sensitivity of applying shear modes for the NRUS method are explored.

PACS numbers: 43.25.Ba, 43.25.Gf, 43.35.Zc [MH]
Date Received: September 1, 2006 **Date Accepted:** January 17, 2007

1. Introduction

Nonlinear acoustics based methods offer promising means for nondestructive evaluation because of their sensitivity in comparison with linear methods (velocity, attenuation). Methods have been, and are currently, in development to apply nonlinear means to detect and image localized damage with, for example, time reversal nonlinear elastic wave spectroscopy (TR NEWS[1]), and distributed damage with NRUS[2] as well as other nonlinear methods. Concrete is a structural heterogeneous and microcracked material exhibiting strong elastic nonlinearity similar to rock[3] and geomaterials[4] in general, including granular media.[5] In addition to classical Landau and Lifschitz[6] theory, their nonlinear response may be physically explained at different scales by dislocations, rupture, and recovery of intergrain cohesive bonds, porosity, opening/closing of micro-cracks, etc. As a result, these materials exhibit hysteresis in their pressure-strain response, the phenomenon of slow dynamics, and are thought to also exhibit end point memory.[7,8] A phenomenological description based on the Preisach-Mayergoyz space representation describing both second- and higher-order nonlinearity and hysteretic behavior has been proposed.[7,8] Note that this model does not contain the slow dynamics (a time dependant recovery process of elastic properties occurring after a disturbance) present in these materials. A nonlinear and hysteretic modulus[9] in the stress strain relationship in one dimension can be written

$$M(\varepsilon,\dot{\varepsilon}) = M_0(1 - \beta\varepsilon - \delta\varepsilon^2 - \cdots - \alpha(\Delta\varepsilon + \mathrm{sign}(\dot{\varepsilon})\varepsilon)), \quad (1)$$

where M_0 is the linear modulus, ε is strain, $\dot{\varepsilon}$ the strain rate, β and δ the second and third order nonlinearity, α being the nonlinear hysteretic parameter, $\Delta\varepsilon$ the average strain amplitude, and the sign function equals +1 if the strain rate is positive and −1 if negative.

This model predicts a softening or hardening of the material with increasing driving amplitude depending on the signs of β, δ, and α. If the net effect is negative (as it is in geomaterials, for instance), the resonant frequency decreases as a function of wave amplitude. At large strain amplitude levels in these materials, much empirical evidence suggests that the nonlinear hysteretic behavior proportional to α dominates,[2] and a first order approximation gives

$$\frac{f_0 - f}{f_0} \approx \alpha \Delta\varepsilon, \tag{2}$$

where f_0 is the linear resonant frequency and f the resonant frequency for an increasing driving amplitude. The evaluation of this linear (slope α) relative frequency shift dependence with strain amplitude is the basis of the NRUS method.

Some studies have already explored the potential of nonlinear methods on evaluating the physical/mechanical properties of concrete. For instance, curing of concrete has been monitored by harmonic generation[10] and damage evaluation has been studied by the nonlinear wave modulation[11] method. NRUS has already been employed in mechanically damaged concrete,[12,13] providing promising results which indicate that the method has potential to monitor thermal damage.

NRUS on damaged concrete exploits longitudinal[13] (P) or flexural[12] mode to estimate the nonlinear α parameter.

To our knowledge, the nonlinear hysteretic behavior of concrete has not been studied applying shear (S) waves. Potentially, S waves propagating in nonlinear hysteretic material should be efficient for nondestructive evaluation.[14] We can reasonably expect that sliding of rough contacts at grain boundaries and microcracks lips may be hysteretic. Note that excitation of these phenomena take place in P modes by coupling between P and S waves due to Poisson effect, nonlinear processes,[15] and scattering[16] from inhomogeneities.

The aim of this paper is to study the evolution of concrete thermal damage applying NRUS and comparing the results to ultrasonic velocities. We then examine S wave sensitivity to thermal damage by applying the NRUS method for shear.

2. Thermal damage process of concrete

Concrete is a complex multiphasic solid material composed, before curing, of anhydrous cement, aggregates, sand, and water. Anhydrous cement is principally composed of silica (SiO_2), alumina (Al_2O_3), lime (CaO), and calcium sulphate ($CaSO_4$). Most of the contained aggregates are limestone and silica. The aggregate size is generally between 3 and 16 mm. Cohesion of concrete is guaranteed by a water cement ratio (w/c) of typically $0.3 < w/c < 0.6$. Chemical processes occur with heat generated during curing, producing an increase of porosity and mi-

Table 1. Chemical process occurring in concrete while increasing temperature. The top three lines are the temperature range studied here.

Temperature	Chemical process
$\rightarrow$105 °C	$\langle\langle$Free$\rangle\rangle$ water evaporation
$\rightarrow$300 °C	First step of dehydration. Breaking of cement gel and uprooting of water molecules into hydrated silicates
400$\rightarrow$500 °C	Portlandite decomposition: $Ca(OH)_2 \rightarrow CaO + H_2O$
600 °C	Structural transformation of quartz α into β—swelling of quartziferous aggregates
$\rightarrow$700 °C	Second dehydration step: dehydration of hydrated calcium silicates
$\rightarrow$900 °C	Limestone decomposition: $CaCO_3 \rightarrow CaO + CO_2$
1300 °C	Aggregates and cement paste fusion

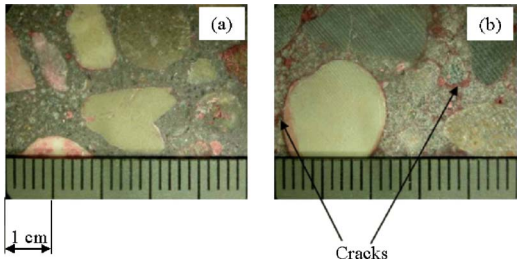


Fig. 1. (Color online) Macrography of intact sample (a) and thermally damaged sample (b) (Ref. 18).

crocracks. Thermal damage process of concrete is well known[17] and synthesized in Table 1. Evidence of cracking is obtained applying macrography[18] which provides the means for estimating the crack density (Fig. 1). For intact concrete we observe 10^3 cracks/m^2. For 200 °C thermally damaged concrete (held at temperature for 3 hours) we observe 33×10^3 cracks/m^2. These measures reveal two essential observations: (i) there is no preferential cracking direction validating our hypothesis of isotropic damage; (ii) most of cracks appear at the cement-aggregate interface and in the cement matrix but never inside the aggregates, following the chemical process described in Table 1 (the first aggregate transformation appears at 600 °C).

3. Experiments

Four samples were studied. The first is a reference (20 °C), while three others have been (1) heated for 3 hours, to 120 °C; (2) to 250 °C, and (3) to 400 °C, respectively. These samples are parallelepipeds of dimension $10 \times 10 \times 5$ cm. P wave transducers (Panametrics V1012, central frequency: 100 kHz) are glued (Salol) on both polished sides of the sample (Fig. 2) and driven by a function generator with high voltage output. In order to find the first compressional resonance mode, a P wave time of flight t measurement is performed. Due to the free surface boundary conditions, the resonant frequency is given by

$$f_0 = 1/2t. \tag{3}$$

For each amplitude (at least 7), a monochromatic tone burst is transmitted. The duration of the burst is selected so as to perform an RMS measurement at steady-state conditions (order Q-cycles, or about 100 cycles). The frequency of the tone burst is fixed around f_0 to obtain a resonance curve. The same scheme is repeated at each amplitude level. Figure 3 presents typical NRUS curves. The system linearity was checked with a reference steel sample using the identical system. We exploit measured RMS amplitude V_{RMS}, which is proportional to the strain amplitude

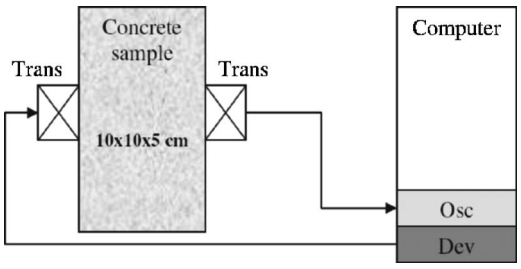


Fig. 2. Scheme of the NRUS experiment. Osc: A/D converter; Dev: high voltage ultrasonic device; Trans: Panametrics transducers (V1012 for P modes and V1548 for S modes).

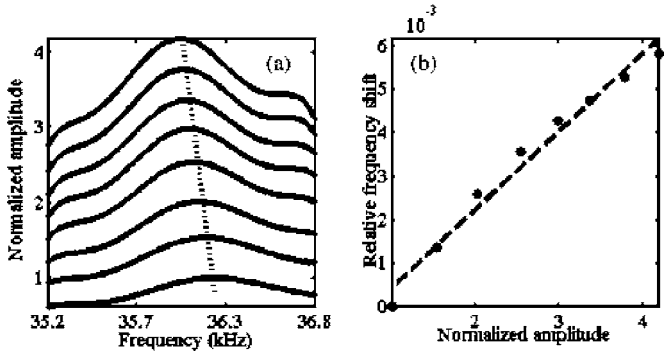


Fig. 3. 120 °C damaged sample frequency shift (a) and extraction of α from the slope of the frequency change with amplitude (b).

$$V_{\mathrm{RMS}} = K\Delta\varepsilon, \tag{4}$$

with K the transducer constant. The value of nonlinear parameter αK is obtained in Fig. 3 by Eq. (2).

In order to compare the sensibility of the NRUS with a linear parameter, velocity is obtained via the linear resonant frequency

$$\nu = 2Lf_0, \tag{5}$$

with L the length of the sample.

As expected, results show the high sensitivity of NRUS to thermal damage applying compression (Fig. 3). Its dynamic evolution is far greater than the classical linear method (Fig. 4). The relative variation of α is 230% while relative velocity variation is only 35%.

The implementation of S modes for NRUS follows the same scheme. The only difference is that the mode is selected so that the half wavelength corresponds to a third of the sample length (third bulk S-resonance mode). This mode is used in order to employ the S-wave transducers (Panametrics V1548, central frequency 100 kHz) near their central frequency, and to be

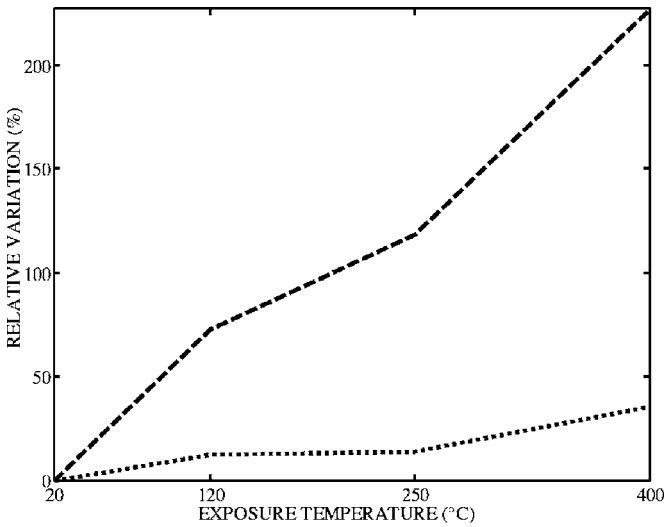


Fig. 4. Relative variation of nonlinear α parameter for first Young mode (dashed line) compared to relative variation of velocity (dotted line) in function of exposure temperature.

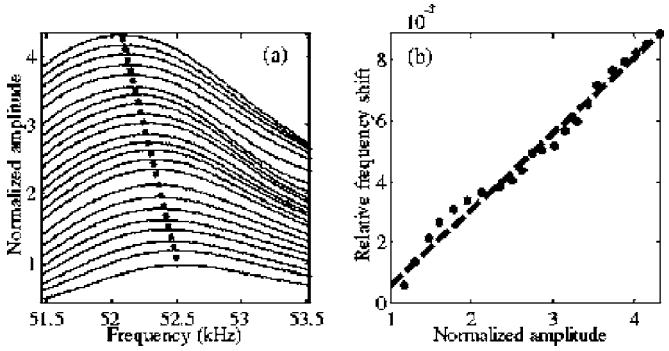


Fig. 5. 250 °C damaged sample frequency shift (a) and extraction of α parameter (b) in shear mode.

sure that the mode explored is the mode expected. Higher modes, for this particular geometry, are not exploitable because of increasing mode density with frequency. Time of flight measurement of S waves is more difficult because S transducers generate a small P wave as well (less than 30 dB/S wave) and concrete causes mode conversion by multiple scattering. Thus the arrival is masked by P-wave coda. For our frequency range (~50 kHz) and length of sample (~5 cm), it is nearly impossible to separate S and P waves. Therefore, the time-of-flight is measured at higher frequency (500 kHz) with another transducer (Panametrics V151).

The feasibility of applying S modes for NRUS method is achieved (Fig. 5). Moreover, sensitivity to thermal damage of the nonlinear α parameter extracted from the S mode (Fig. 5), is very close, less than 8% to that of the P one (Fig. 6).

Note that the fits of the change in frequency vs amplitude for extraction of α in both the compressional [Fig. 3(b)] and shear experiments [Fig. 5(b)] are not perfect, and could be fit with other functions. The shear result is particularly complex. In future experiments we will explore in more detail these behaviors and whether they may change with increasing damage. It may be that the simple model presented here based on hysteresis is only partially correct.

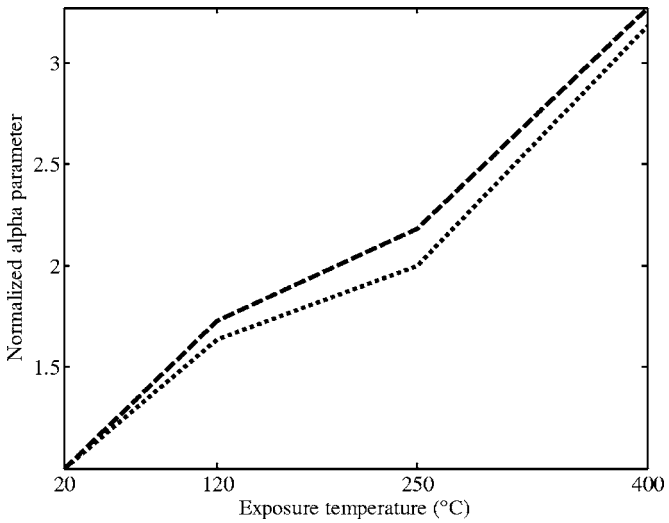


Fig. 6. Comparison of nonlinear α parameter for the P mode (dashed line) with the S mode (dotted line) as a function of exposure temperature.

4. Conclusions and prospects

The significant sensitivity of the nonlinear response to thermal damage in concrete is demonstrated. The method, when compared with linear velocity measurement, exhibits greater sensitivity which should be useful for nondestructive evaluation.

Shear modes have also been tested. Their feasibility for NRUS method and their sensitivity to thermal damage have been illustrated. Qualitative values of the nonlinear α parameter have been obtained for both P and S modes and their dynamic evolutions are very similar for this isotropic damage. Therefore, the same study should be performed to monitor the evolution of the P and S modes responses for different anisotropic mechanical states.

References and links

[1] T. J. Ulrich, P. A. Johnson, and A. Sutin, "Imaging nonlinear scatterers applying the time reversal mirror," J. Acoust. Soc. Am. **119**, 1514–1518 (2006).

[2] K. Van Den Abeele, J. Carmeliet, J. TenCate, and P. A. Johnson, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. Part II: Single-mode nonlinear resonance acoustic spectroscopy," Res. Nondestruct. Eval. **12**, 31–43 (2000).

[3] R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**, 30–35 (1999).

[4] L. Ostrovsky and P. A. Johnson, "Dynamic nonlinear elasticity in geomaterials," Riv. Nuovo Cimento **24**, 1–46 (2001).

[5] P. A. Johnson and X. Jia, "Nonlinear dynamics, granular media and dynamic earthquake triggering," Nature (London) **437**, 871–874 (2005).

[6] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon Press, New York, 1959).

[7] K. R. McCall and R. A. Guyer, "Equation of state and wave propagation in hysteretic nonlinear elastic material," J. Geophys. Res. **99**, 23887–23897 (1994).

[8] R. A. Guyer, K. R. McCall, and G. N. Boitnott, "Hysteresis, discrete memory, and nonlinear wave propagation in rock: A new paradigm," Phys. Rev. Lett. **74**, 3491–3494 (1995).

[9] K. Van Den Abeele, P. A. Johnson, and A. Sutin, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. Part I: Nonlinear wave modulation spectroscopy (NWMS)," Res. Nondestruct. Eval. **12**, 17–30 (2000).

[10] J. C. Lacouture, P. A. Johnson, and F. Cohen-Tenoudji, "Study of critical behavior in concrete during curing by application of dynamic linear and nonlinear means," J. Acoust. Soc. Am. **113**, 1325–1332 (2003).

[11] K. Warnemuende and H. C. Wu, "Actively modulated acoustic nondestructive evaluation of concrete," Cem. Concr. Res. **34**, 563–570 (2004).

[12] K. Van Den Abeele and J. De Visscher, "Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration techniques," Cem. Concr. Res. **30**, 1453–1464 (2000).

[13] M. Bentahar, H. El Aqra, R. El Guerjouma, M. Griffa, and M. Scalerandi, "Hysteretic elasticity in damaged concrete: Quantitative analysis of slow and fast dynamics," Phys. Rev. B **73**, 014116 (2006).

[14] V. Gusev, C. Glorieux, W. Lauriks, and J. Thoen, "Nonlinear bulk and surface shear acoustic waves in materials with hysteresis and end-point memory," Phys. Lett. A **232**, 77–86 (1997).

[15] A. Goldberg, "Interaction of plane longitudinal and transverse elastic waves," Sov. Phys. Acoust. **6**, 306–310 (1960).

[16] V. V. Varadan and V. K. Varadan, "Scattering matrix for elastic waves. III. Application to spheroids," J. Acoust. Soc. Am. **75**, 896–905 (1979).

[17] N. A. Noumowé, "Effet de hautes températures (20 °C–600 °C) sur le béton. Cas particulier du BHP ("Effect of high temperatures (20 °C–600 °C) on high performance concrete")," Ph.D. thesis, INSA de Lyon, 1995.

[18] J. F. Chaix, "Caractérisation non destructive de l'endommagement de bétons: apport de la multidiffusion ultrasonore ("Nondestructive evaluation of concrete damage: Contribution of the ultrasonic multiple scattering")," Ph.D. thesis, Université de la Méditerranée, 2003.

The perception of phonemic contrasts in a non-native dialect

Sophie Dufour
Laboratoire de Psycholinguistique Expérimentale, University of Geneva, Switzerland
Sophie.Dufour@pse.unige.ch

Noël Nguyen
Laboratoire Parole et Langage, CNRS & Aix-Marseille University, Aix-en-Provence, France
noel.nguyen@lpl.univ-aix.fr

Ulrich Hans Frauenfelder
Laboratoire de Psycholinguistique Expérimentale, University of Geneva, Switzerland
ulfrich.frauenfelder@pse.unige.ch

Abstract: This study examined the impact on speech processing of regional phonetic/phonological variation in the listener's native language. The perception of the /e/-/ɛ/ and /o/-/ɔ/ contrasts, produced by standard but not southern French native speakers, was investigated in these two populations. A repetition priming experiment showed that the latter but not the former perceived words such as /epe/ and /epɛ/ as homophones. In contrast, both groups perceived the two words of /o/-/ɔ/ minimal pairs (/pom/-/pɔm/) as being distinct. Thus, standard-French words can be perceived differently depending on the listener's regional accent.

PACS numbers: 43.71.Hw [DOS]
Date Received: September 19, 2006 **Date Accepted:** January 15, 2007

1. Introduction

Adult listeners find it difficult to discriminate speech sounds not present in their native language. Models such as Best's "Perceptual Assimilation Model" Best *et al.* (1988) and Flege's "Speech Learning Model" (1995) propose that non-native speech perception abilities can best be explained by making reference to the native phonetic space. For example, adult Japanese listeners have difficulties discriminating between American English [l] and [ɹ] (Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura, 1975), presumably because Japanese has only one liquid phoneme (/ɾ/), to which both [l] and [ɹ] are assimilated by Japanese listeners. Difficulties in perceiving non-native speech sounds are not only observed for a foreign language, but also for the second language of highly fluent bilinguals. In a study on the perception of the /e/-/ɛ/ contrast found in Catalan but not in Spanish, Pallier *et al.* (2001) showed that bilinguals who acquired Spanish first but learned Catalan before 6 years of age, merged this contrast and processed Catalan words like /perə/ and /pɛrə/ as homophones. Hence, even for fluent bilinguals, speech perception depends upon the phonemic categories of the first language.

While considerable work has accumulated on the perception of non-native phonemic contrasts, the impact on speech perception of *within*-language phonetic/phonological variation has only recently been investigated. Moreover, conflicting results have emerged about how phonemic contrasts specific to one language's regional variety are perceived by listeners of another regional variety. In a recent study of the perception of vowels from two different British English accents, Evans and Iverson (2004) showed that listeners can adjust their vowel categorization decisions according to the accent of the carrier sentence. Also, Cutler *et al.* (2005) had speakers of Australian English identify vowels in CV and VC syllables produced by an American English talker and found that the Australian speakers' performance was similar to that of speakers of

American English. In contrast, Conrey *et al.* (2005) found that speakers of an American-English dialect with a /i/-/ɛ/ vowel merger were less accurate than speakers of an unmerged dialect in discriminating between minimal pairs of words (e.g., *pin/pen*) contrasting in these vowels (see also Labov *et al.* 1991, and Janson and Schulman, 1983 for Swedish dialects). Further evidence that regional phonetic/phonological variation affects speech perception has been obtained by Floccia *et al.* (2006), who showed that French words were recognized more slowly when spoken in an unfamiliar accent.

To gain a better understanding of the effect of regional accent on speech perception, we have examined the word identification performance by listeners of two varieties of French, standard French and southern French which have different phonemic inventories. Standard French has three mid vowel pairs, /e/-/ɛ/, /ø/-/œ/ and /o/-/ɔ/ which are all contrastive, as is the case of the minimal pairs *épée* /epe/ "sword" vs *épais* /epɛ/ "thick" or *côte* /kot/ "hill" vs *cote* /kɔt/ "rating". Conversely, there is no contrastive distinction between /e/-/ɛ/, /ø/-/œ/, and /o/-/ɔ/ in southern French, which is viewed as having three mid vowel phonemes only, namely, /e/, /ø/, and /o/ (Durand, 1990). [ɛ], [œ], and [ɔ] do appear at the phonetic level, but they are in complementary distribution with respect to the corresponding mid-high variants: mid-high vowels occur in open syllables and mid-low vowels in closed syllables and whenever the next syllable contains schwa /ə/ (Durand, 1990). Thus, *épée* and *épais* will both be pronounced [epe], while *côte* and *cote* will both be pronounced [kɔtə].

How do listeners of southern French perceive standard French forms such as [epɛ], compared with [epe], or [kot], compared with [kɔt]? To address this question, a long lag repetition priming paradigm was used as in Pallier *et al.* (2001). The repetition priming effect refers to the fact that participants respond more rapidly when a word is encountered a second time. Of particular interest is the amount of priming obtained between members of minimal pairs such as /epe/-/epɛ/ or /kot/-/kɔt/. Standard-French listeners should clearly perceive these words as being different and show no facilitation on the processing of the second word of the pair. In contrast, since [e]-[ɛ] as well as [o]-[ɔ] are conditioned variants associated with a single phonemic category in southern French, these listeners should map the two words of a minimal pair onto the same lexical representation, regardless of the height (mid-high vs mid-low) of the critical vowel. Hence, unlike standard-French listeners, southern-French listeners should take the second word of the minimal pair to be a repetition of the first and show facilitation on the second word. We used a lexical decision task in which participants had to discriminate between words and nonwords.

2. Method

2.1 Participants

Sixty-nine native speakers of French participated in the experiment for course credit. Thirty-two were standard-French listeners from the University of Geneva. The other 37 were southern-French listeners from the Aix-en-Provence area. All participants reported no hearing disorders.

2.2 Materials

Sixty-four (C)V.CV bisyllabic words making up 32 minimal pairs based on the /e/-/ɛ/ phonemic contrast were selected. This vowel contrast occurred in word-final position. These 32 pairs were split into two sets according to their morphological/semantic relation. For one set, the words forming a pair were morphologically/semantically related (e.g., /pike/ *to prick*–/pikɛ/ *stake*). For the other, no morphological or semantic relation existed between the words (e.g., /epe/ *sword*–/epɛ/ *thick*). The results of related and unrelated Pairs will be analyzed separately, since it has been shown that a morphological/semantic relation between prime and target can facilitate target word processing (Meunier and Segui, 2002). Sixteen CVC monosyllabic words forming eight minimal pairs containing the /o/-/ɔ/ phonemic contrast were also selected (e.g., /pom/ *palm*–/pɔm/ *apple*). Finally, we included 16 CV monosyllabic words forming eight minimal pairs based on the /ø/-/y/ phonemic contrast (e.g., /fø/ *fire*–/fy/ *barrel*) which exists in both

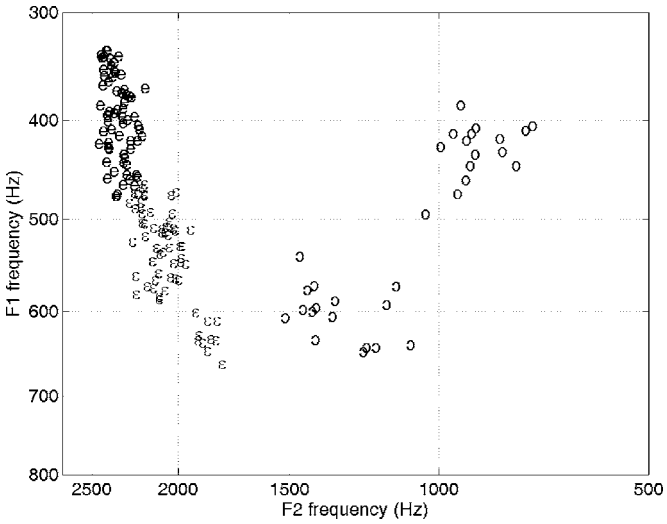


Fig. 1. F1 and F2 frequencies for the critical vowel in each of the /e/-/ɛ/ and /o/-/ɔ/ words.

regional varieties. Only eight minimal pairs were found for the /o/-/ɔ/ and /ø/-/y/ phonemic contrasts, because pairs based on these contrasts are rare in French.

For the purpose of the lexical decision task, 96 nonwords were created. Sixty-four were bisyllabic and formed 32 minimal pairs based on the /e/-/ɛ/ contrast. The nonwords were created by replacing the first syllable of test words by the first syllable of another word present in the list to avoid a strategy for word responses based on the repetition of the first syllable. The remaining 32 nonwords were monosyllabic, forming eight minimal pairs based on the /o/-/ɔ/ contrast and eight minimal pairs based on the /ø/-/y/ contrast. Monosyllabic nonwords were created by replacing the first phoneme of the test words. Finally, an additional 108 words and 108 nonwords were included as filler items in the experimental lists.

Four counterbalanced lists of 408 items were created such that each member of a minimal pair was repeated or followed by the other member of the minimal pair. Primes and targets were separated by 8 to 17 items. In each list, the words forming a pair appeared in the same positions.

2.3 Acoustic stimuli

The stimuli were recorded by a female native speaker of standard French with a Parisian accent. Each minimal pair was produced twice, and the pair for which the acoustic differences in the critical vowel were the largest between the two words was presented to participants. Figure 1 shows the frequency of the second formant (F2) as a function of the first formant (F1) for the critical vowel in each of the /e/-/ɛ/ and /o/-/ɔ/ words. As can be seen, both /e/-/ɛ/ and /o/-/ɔ/ contrasts were produced distinctly: /ɔ/ was systematically associated with higher frequencies of both the first and second formants than /o/, and /ɛ/ was characterized by both a higher first formant frequency and a lower second formant frequency than /e/. Although the regions associated with /e/ and /ɛ/ in the F2 vs F1 acoustic plane were close to each other, there was little overlap between the two.

3. Procedure

Participants were asked to make a lexical decision as quickly and accurately as possible and to give "word" responses using their dominant hand. Response times (RTs) were measured from the onset of the test item. An interval of 2500 ms elapsed between the participant's response and the presentation of the next stimulus. If participants failed to respond within 1800 ms of a

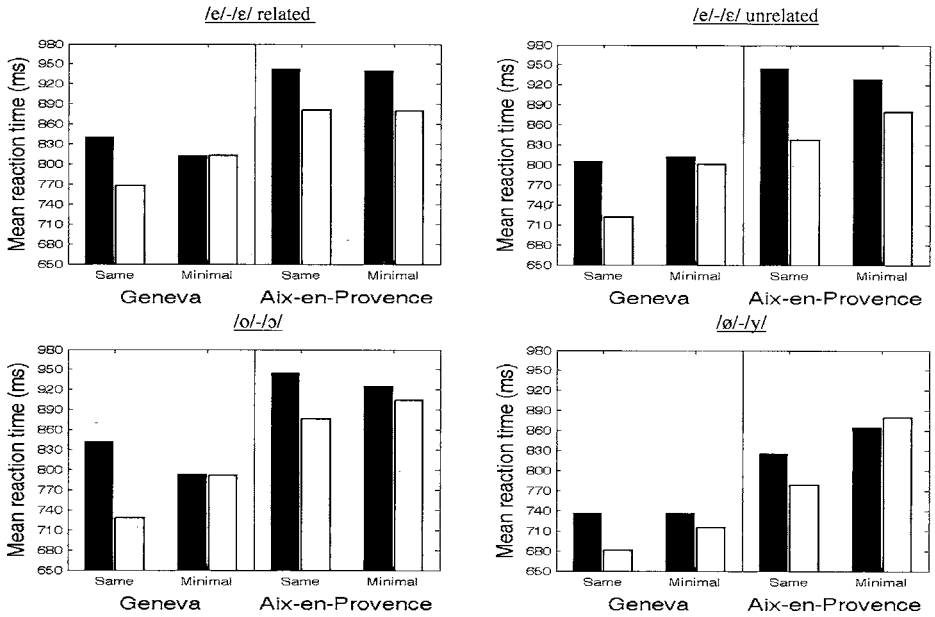


Fig. 2. Mean reaction times (in ms) for the first (black bars) and the second (white bars) occurrence as a function of pair type and population for each contrast.

stimulus, no response was recorded and a next item was presented. The participants were tested on only one experimental list and began the experiment with 12 practice trials.

4. Results

Mean correct response times (RTs) in each condition were computed and are shown in Fig. 2. The data from one participant from Aix-en-Provence were not included in the statistical analyses given the high percentage of errors on the task (more than 25%). Four word pairs (two, one, and one, for the /e/-/ɛ/ related, /e/-/ɛ/ unrelated, and /ø/-/y/ contrast, respectively) giving rise to more than 30% of errors were also excluded from the analyses. For each contrast, planned comparisons on the RTs assessed the priming effect (first vs second occurrence) within each pair (same, minimal) and for each group (Geneva, Aix-en-Provence). F values are reported by participants ($F1$) and items ($F2$).

/e/-/ɛ/ *related contrast*: A priming effect was observed for the same pairs in both Geneva [$F1(1,66)=17.81$, $p<.001$; $F2(1,40)=16.34$, $p<.001$] and Aix-en-Provence [$F1(1,66)=14.14$, $p<.001$; $F2(1,40)=10.46$, $p<.01$]. RTs were shorter for target words encountered for the second time compared with the first time. A priming effect was observed for the minimal pairs, with RTs being faster for the second than for the first occurrence in Aix-en-Provence [$F1(1,66)=14.47$, $p<.001$; $F2(1,40)=8.58$, $p<.01$] but not in Geneva (Fs<1). Note that the lack of a priming effect for the minimal pairs in Geneva suggests that the morphological/semantic relation between words had no impact.

/e/-/ɛ/ *unrelated contrast*: A priming effect was observed for the same pairs in both Geneva [$F1(1,66)=21.53$, $p<.001$; $F2(1,40)=26.18$, $p<.001$] and Aix-en-Provence [$F1(1,66)=40.16$, $p<.001$; $F2(1,40)=29.61$, $p<.001$]. RTs were shorter for target words encountered for the second time compared with the first time. A priming effect was observed for the minimal pairs, with RTs being shorter for the second than for the first occurrence in Aix-en-Provence [$F1(1,66)=10.31$, $p<.01$; $F2(1,40)=5.57$, $p<.05$] but not in Geneva [$F1<1$; $F2(1,40)=1.96$, $p=.17$].

/o/-/ɔ/ *contrast*: A priming effect was observed for the same pairs in both Geneva [$F1(1,66)=28.59$, $p<.001$; $F2(1,40)=27.34$, $p<.001$] and Aix-en-Provence [$F1(1,66)=11.67$, $p<.01$; $F2(1,40)=9.24$; $p<.01$]. RTs were shorter when the target words were encountered for the second time relative to the first time. No priming effect was observed for the minimal pairs in either Geneva (Fs <1) or Aix-en-Provence (Fs <1).

Finally, neither group showed a priming effect for the minimal pairs based on the /ø/-/y/ contrast which was common to both regional French varieties [in Geneva ($F1<1$; $F2(1,40)=1.32$; $p>.20$); in Aix-en-Provence (Fs <1)].

5. Discussion

This study has shown differences in how French listeners perceive words contrasting in only one phonetic feature depending on their regional accent. When presented with French minimal pairs containing either a /e/-/ɛ/ or a /o/-/ɔ/ standard phonemic contrast, standard French listeners showed no priming between members of a minimal pair. Hence, they perceived the two members of a minimal pair as being different. In contrast, southern-French listeners showed a priming effect on members of the minimal pair with the /e/-/ɛ/ contrast. This suggests that they treated the second member as a repetition of the first. These findings are consistent with the Best *et al.* (1988) model and suggest that listeners are insensitive to phonemic contrasts when the speech sounds can be assimilated perceptually to a single category in their phonemic inventory.

Another significant outcome of this study was that southern-French listeners did not respond in the same way to the /e/-/ɛ/ and /o/-/ɔ/ contrasts, since a priming effect was found for the former but not for the latter. The absence of a priming effect for the /o/-/ɔ/ minimal pairs suggests that the two words were perceived as being distinct by southern-French listeners. One potential explanation for this discrepancy lies in how both contrasts were produced. It may be the case, as Fig. 1 suggests, that /e/ differed from /ɛ/ to a lesser extent than /o/ from /ɔ/ at the phonetic level, hence allowing the first word to facilitate the recognition of the second word in the /e/-/ɛ/ but not in the /o/-/ɔ/ pairs. However, the lack of a priming effect for the /e/-/ɛ/ minimal pairs for the standard French speakers confirms that differences between /e/ and /ɛ/ were clearly perceivable.

The differential response pattern for the two contrasts in southern French listeners could also be related to the phonological status of these contrasts in today's standard French. The /o/-/ɔ/ contrast is a particularly well-established phonological feature of standard French which is, as such, known to the listeners of southern French, even if this contrast is neutralized in their own productions. Indeed it probably stands as one of the shibboleths allowing southern listeners to recognize standard-French speakers. The explicit knowledge that southern French listeners have of the /o/-/ɔ/ contrast could explain why they perceive the two members of /o/-/ɔ/ minimal pairs differently.

By comparison, the /e/-/ɛ/ contrast, as it occurs in word-final position in standard French, is characterized by greater complexity both across and within speakers (e.g., Coveney, 2001; Fagyal *et al.* 2006; Tranel, 1987). For example, Tranel (1987) points out that the distribution of /e/ and /ɛ/ is partly speaker-dependent (e.g., *quai* "platform" may be pronounced [ke] by some speakers and [kɛ] by others), and that the pronunciation of the vowel in grammatical words such as *les* "the," plural form, or *est* "is" commonly varies between [e] and [ɛ] even for a given speaker. In addition, there is experimental evidence showing that /e/ and /ɛ/ in word-final position are in the process of merging in Parisian French (e.g., Fagyal *et al.* 2002). For example, young Parisian speakers tend to no longer maintain a distinction in the pronunciation of *épée* vs *épais*. Both the complexity of the pronunciation of the /e/ and /ɛ/ vowels and the merging process of the two vowels in standard French make it less likely that this opposition is represented in the southern listeners' receptive phonological knowledge of standard French. As a consequence, southern French listeners tend to assimilate the [e] and [ɛ] vowels to their [e] native category and perceived the two members of /e/-/ɛ/ minimal pairs as homophones. It thus appears that the listener's linguistic experience influences speech processing, even within dialects of the same language.

Acknowledgments

This research was partly supported by the TCAN interdisciplinary CNRS program. Thanks to Cécile Fougeron for recording the speech material. We are also grateful to Robert Espesser for sharing his statistical expertise and to Jacques Durand, Zsuzsanna Fagyal, and Bernard Laks for helpful comments. We also thank the reviewers for their helpful comments.

References and links

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. **14**, 345–360.

Conrey, B., Potts, G. F., and Niedzielski, N. A. (**2005**). "Effects of dialect on merger perception: ERP and behavioral correlates," Brain Lang **95**, 435–449.

Coveney, A. (**2001**). *The Sounds of Contemporary French: Articulation and Diversity* (Elm Bank Publications, Exeter, UK).

Cutler, A., Smits, R., and Cooper, N. (**2005**). "Vowel perception: Effects of non-native language vs non-native dialect," Speech Commun. **47**, 32–42.

Durand, J. (**1990**). *Generaive and Non-Linear Phonology* (Longman, London, UK).

Evans, B. G., and Iverson, P. (**2004**). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," J. Acoust. Soc. Am. **115**, 352–361.

Fagyal, Z., Hassa, S., and Ngom, F. (**2002**). "L'opposition [e]-[ɛ] en syllabes ouvertes en fin de mot en français parisien: étude acoustique préliminaire ("The [e]-[ɛ] opposition in open syllables at word offsets in Parisian French: A preliminary acoustic study")," *Proceedings of the XXIVémes Journèes d'Études sur la Parole* (Nancy, France, 24–27 June), pp. 165–168.

Fagyal, Z., Jenkins, F., and Kibbee, D. (**2006**). "French: A Linguistic Introduction," in *Phonetics and phonology* (Cambridge University Press, Cambridge, UK).

Flege, J. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-language research*, edited by W. Strange and J. Jenkins (York Press, Timonium, MD), pp. 233–277.

Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (**2006**). "Does a regional accent perturb speech processing?," J. Exp. Psychol. Hum. Percept. Perform. **32**, 1276–1293.

Janson, T., and Schulman, R. (**1983**). "Non-distinctive features and their use," Linguistics **19**, 321–336.

Labov, W., Karan, M., and Miller, C. (**1991**). "Near-mergers and the suspension of phonemic contrast," Lang. Var. Change **3**, 33–74.

Meunier, F., and Segui, J. (**2002**). "Cross-modal morphological priming in French," Brain Lang **81**, 89–102.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of /r/ and /l/ by native listeners of Japanese and English," Percept. Psychophys. **18**, 331–340.

Pallier, C, Colomé, A., and Sebastián-Gallés, N. (**2001**). "The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries," Psychol. Sci. **12**, 445–449.

Tranel, B. (**1987**). *The Sounds of French* (Cambridge University Press, Cambridge, UK).

Experimental study on subharmonic and ultraharmonic acoustic waves in water-saturated sandy sediment

Byoung-Nam Kim, Kang Il Lee, and Suk Wang Yoon
Department of Physics and Institute of Basic Science, Sungkyunkwan University, Suwon 440-746, Republic of Korea
bnkim@skku.edu, akustica@skku.edu, swyoon@skku.ac.kr

Abstract: Experimental observations of the subharmonic and ultraharmonic acoustic waves in water-saturated sandy sediment are reported in this paper. Acoustic pressures of both nonlinear acoustic waves strongly depend on the driving acoustic pressure at a transducer. The first ultraharmonic wave reaches a saturation value as the driving acoustic pressure increases. The acoustic pressure levels of both nonlinear acoustic waves exhibit some fluctuations in comparison with that of the primary acoustic wave as the receiving distance of hydrophone increases in sediment. The subharmonic and the ultraharmonic phenomena in this study show close resemblance to those produced in bubbly water.

PACS numbers: 43.30.Ma [JFL]
Date Received: October 31, 2006 **Date Accepted:** January 26, 2007

1. Introduction

When a primary acoustic wave with a finite amplitude propagates in a medium, the nonlinear acoustic waves such as subharmonic, ultraharmonic, second harmonic, and higher harmonic acoustic waves can be generated not only due to nonlinearity of a medium but also due to finite amplitude of a primary acoustic field (Ekimov *et al.*, 1996; Solodov *et al.*, 2002; Korshak *et al.*, 2002; Bazhenova *et al.*, 2005). If two primary acoustic waves of different frequencies propagate in a medium, nonlinear acoustic waves at the sum and difference frequencies can be also generated in the medium (Sutin *et al.*, 1998; Ostrovsky *et al.*, 2003). The generation of the subharmonic ($f_0/2$) and the ultraharmonic ($3f_0/2, 5f_0/2, 7f_0/2, \ldots$) acoustic waves is very interesting because they can be generated in the medium only when the acoustic pressure of the primary acoustic wave exceeds a certain threshold value.

The subharmonic and the ultraharmonic acoustic waves can provide important information with other nonlinear acoustic waves in order to detect cracks and defects in solid media (Ekimov *et al.*, 1996; Solodov *et al.*, 2002; Korshak *et al.*, 2002). Ekimov *et al.* (1996) showed that these nonlinear acoustic waves could be applied to the diagnosis of ice cover in a natural fresh water lake. Recently, Solodov *et al.* (2002) and Korshak *et al.* (2002) showed the generation of both nonlinear acoustic waves due to cracks in laminated solid samples. The subharmonic and the ultraharmonic acoustic waves can also provide important information related to resonance which is used to detect bubbles in bubbly granular media. However, the investigation of both nonlinear acoustic waves in this granular media has not been performed because the primary acoustic wave is heavily attenuated. Furthermore, even in water-saturated granular media, the behavior of both nonlinear acoustic waves has not been investigated; this may provide background information for the study on the subharmonic and the ultraharmonic phenomena in bubbly granular media.

The objectives of this paper are, first, to experimentally investigate the generation of the subharmonic and the first ultraharmonic acoustic waves in water-saturated sandy sediment (a water-saturated granular medium) and, second, to experimentally investigate the acoustic

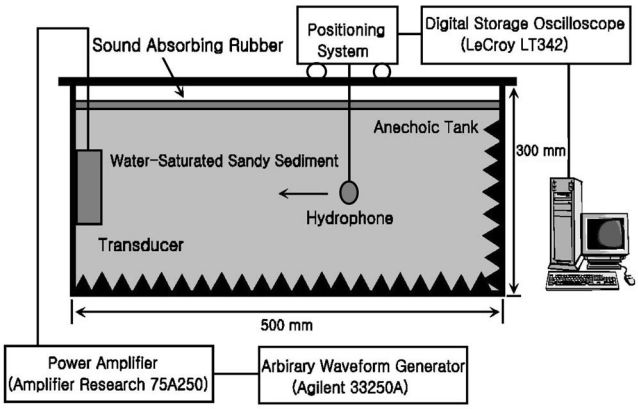


Fig. 1. Experimental setup to measure the subharmonic and the first ultraharmonic acoustic waves.

characteristics such as the dependence on the driving acoustic pressure and on the receiving distance. Most likely mechanisms of the subharmonic and the ultraharmonic generation in water-saturated sandy sediment are also discussed here.

2. Material and method

2.1 Water-saturated sandy sediment

An anechoic water tank of volume $500 \times 280 \times 300$ mm^3 was prepared to pack water-saturated sandy sediment, which was filled with fresh water. The water temperature in the tank was maintained in the range 11 °C and 15 °C for all measurements. To avoid any inclusion of small bubbles within water-saturated sandy sediment, the sediment was slowly packed through a large sieve installed in the anechoic water tank. The porosity of water-saturated sandy sediment and the density of sand grains were $40.8 \pm 1.3\%$ and 2559 ± 52 kg/m^3, respectively. The diameters of sand grains were between 250 and 500 μm.

2.2 Experimental measurements

Figure 1 shows a schematic diagram of the experimental setup to measure the subharmonic and the first ultraharmonic acoustic waves in water-saturated sandy sediment. A transducer with a diameter of 80 mm was used to transmit the signals and was buried in water-saturated sandy sediment. The driving frequency was 76 kHz, which was the first major resonance frequency of the transducer. The signals transmitted through water-saturated sandy sediment were sinusoidal tone burst signals with a pulse duration of 1 ms and repetition time of 100 ms. The tone burst driving method was selected to get a continuous wave condition with a very high driving acoustic pressure. The driving acoustic pressure was defined as the acoustic pressure measured at a distance of 1 mm from the transducer in water-saturated sandy sediment. It increased from 346 to 507 kPa for the subharmonic and the first ultraharmonic acoustic waves in sediment, respectively.

An arbitrary waveform generator (Agilent 33250A) and a power amplifier (Amplifier Research AR 75A 250) were used to drive the transducer. The transmitted signals in water-saturated sandy sediment were received by a hydrophone (B&K 8103). The hydrophone had an omni-directional receiving beam pattern of receiving sensitivity -211.3 dB *re* 1 V/μPa within ± 2 dB between 1 Hz and 150 kHz. Movement of the hydrophone in sediment was controlled by a positioning system. To minimize any variation in the structural composition of the sediment for all measurements, the hydrophone was always moved towards, rather than away from, the transducer as shown in Fig. 1. The received signals were acquired using a 500 MHz digital storage oscilloscope (LeCroy LT342) and stored on a computer for off-line analysis.

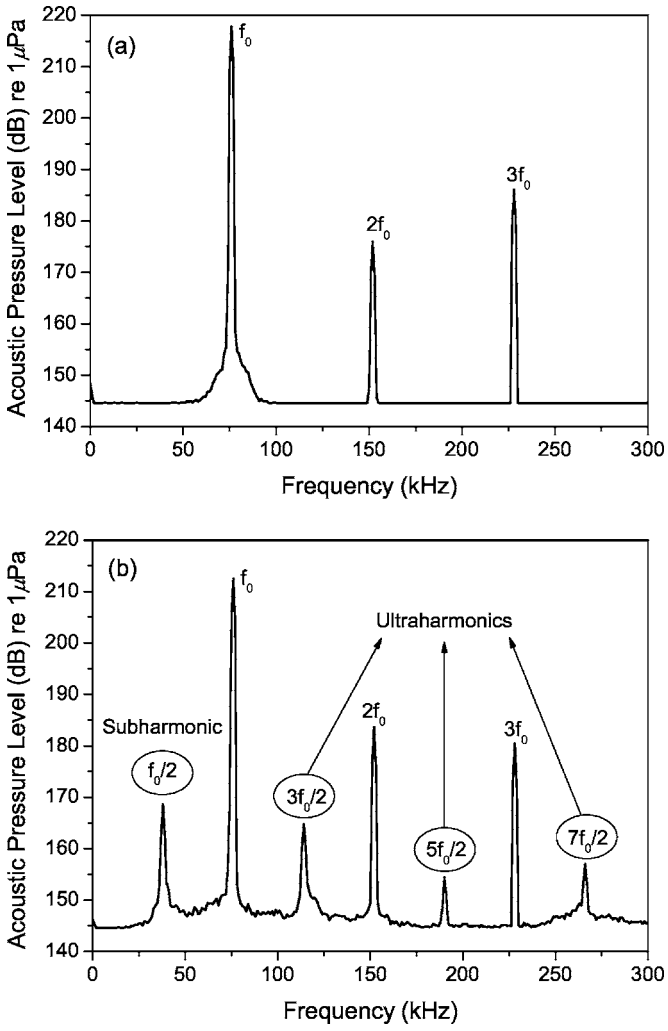


Fig. 2. Frequency spectra of the signals transmitted through (a) pure water and (b) water-saturated sandy sediment at maximum driving acoustic pressure and a distance of 100 mm from the transducer with a resonance frequency of 76 kHz.

3. Results

Figure 2 shows the frequency spectra of signals transmitted through pure water and water-saturated sandy sediment at maximum driving acoustic pressure and a distance of 100 mm from the transducer in the anechoic tank. As shown in Fig. 2(b), the subharmonic (38 kHz) and the ultraharmonic (114, 190, and 266 kHz) acoustic waves were generated due to the nonlinearity of the water-saturated sandy sediment at the primary frequency of 76 kHz. The acoustic pressure level of the subharmonic acoustic wave in sediment was about 24 dB higher over the background noise level, while the levels of the ultraharmonic acoustic waves were between 9 and 20 dB higher.

Figure 3 shows the acoustic pressure variations of the primary, the subharmonic, and the first ultraharmonic acoustic waves as a function of the driving acoustic pressure at the primary frequency of 76 kHz, at a distance of 100 mm from the transducer in water-saturated sandy sediment. As shown in Fig. 3, the acoustic pressure of the primary acoustic wave was

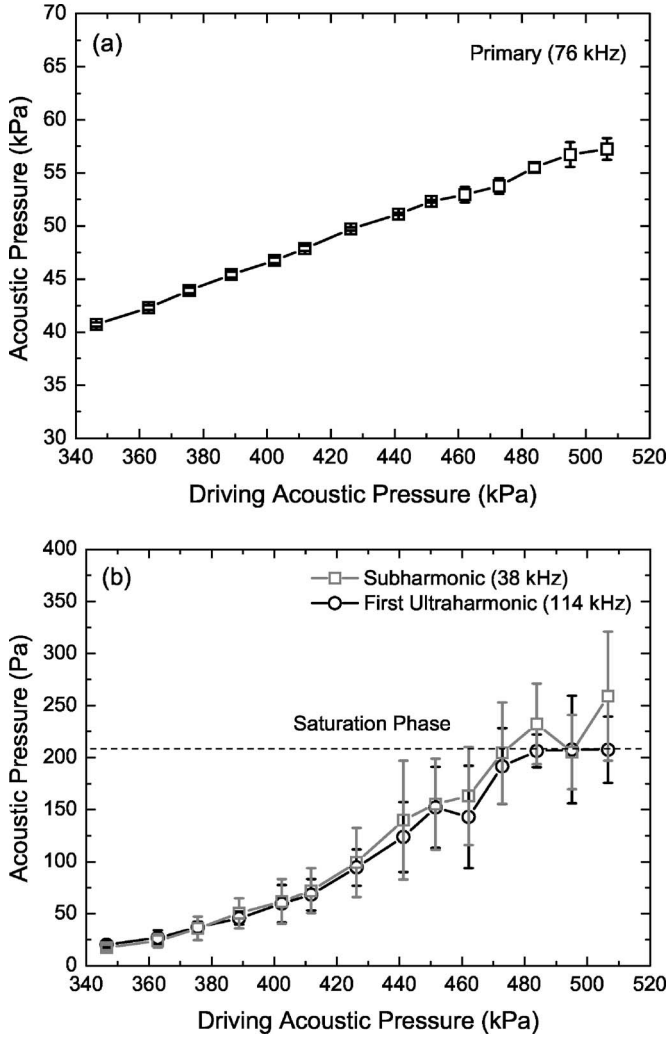


Fig. 3. The acoustic pressure variations of (a) the primary (76 kHz), (b) the subharmonic (38 kHz) and the first ultraharmonic (114 kHz) acoustic waves as a function of the driving acoustic pressure at distance of 100 mm from the transducer in water-saturated sandy sediment.

linearly increased as the driving acoustic pressure increased, while the acoustic pressure of the subharmonic acoustic wave was gradually and exponentially increased. In Fig. 3(b), the acoustic pressure of the first ultraharmonic acoustic wave was also gradually and exponentially increased as the driving acoustic pressure increased up to 484 kPa. However, it approached a saturation value as the driving acoustic pressure became greater than 484 kPa.

Figure 4 shows the acoustic pressure level variations of the primary, the subharmonic, and the first ultraharmonic acoustic waves as a function of receiving distance. The solid, the dashed, and the dot-dashed lines indicate exponential fitting lines for each data, respectively. The acoustic pressure levels for the subharmonic and the first ultraharmonic acoustic waves more rapidly decreased than that for the primary acoustic wave as the receiving distance increased.

To confirm the subharmonic and the ultraharmonic phenomena at another primary frequency in water-saturated sandy sediment, another transducer (RESON TC2122, 180 mm in diameter) with a resonance frequency of 33 kHz was used as the driving transducer and was

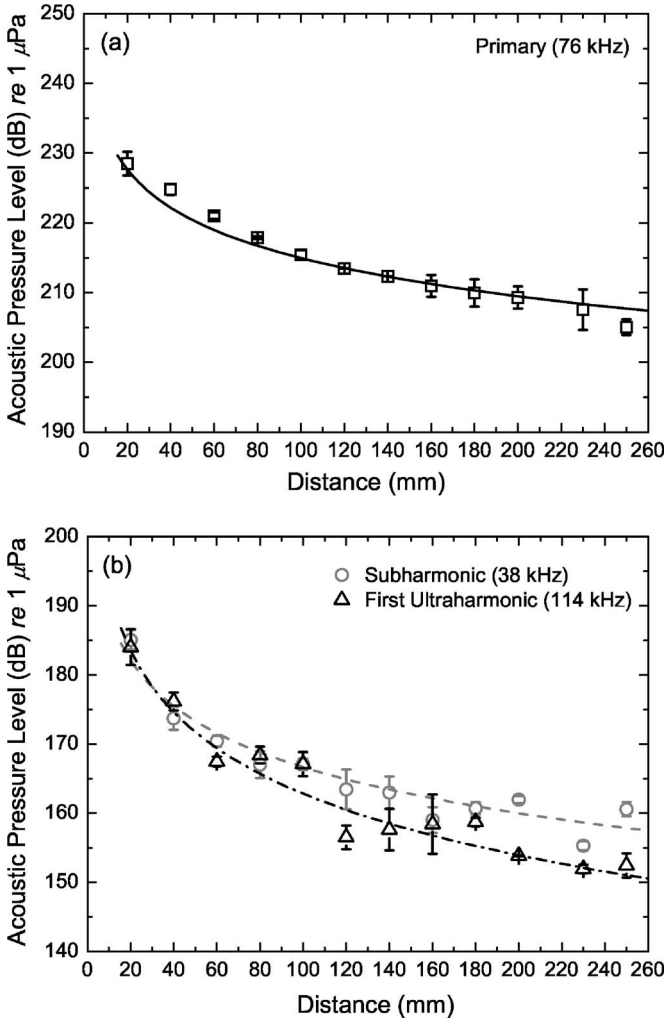


Fig. 4. The acoustic pressure level variations of (a) the primary (76 kHz), (b) the subharmonic (38 kHz) and the first ultraharmonic (114 kHz) acoustic waves in water-saturated sandy sediment as a function of receiving distance.

driven at a maximum acoustic pressure of 246 kPa. Figure 5 shows the frequency spectra of the signals transmitted through pure water and water-saturated sandy sediment at a distance of 100 mm. As shown in Fig. 5(b), the subharmonic (16.5 kHz) and the ultraharmonic (49.5, 82.5, and 115.5 kHz) acoustic waves were generated due to the nonlinearity of the water-saturated sandy sediment. The acoustic pressure levels of the subharmonic and the ultraharmonic acoustic waves in sediment were about 17 dB higher than the background noise level.

4. Discussion

The driving acoustic pressure of 346 kPa in Fig. 3(b) is the minimum acoustic pressure to generate the subharmonic and the first ultraharmonic acoustic waves in water-saturated sandy sediment. Therefore, the driving acoustic pressure of 346 kPa can be considered as the threshold acoustic pressure to generate the nonlinear acoustic waves when the primary acoustic wave propagates in water-saturated sandy sediment at a frequency of 76 kHz. Generally, it is well known that the acoustic pressures of the subharmonic and the first ultraharmonic acoustic waves saturate as the driving acoustic pressure increases (Lostberg *et al.*, 1996; Shankar *et al.*,

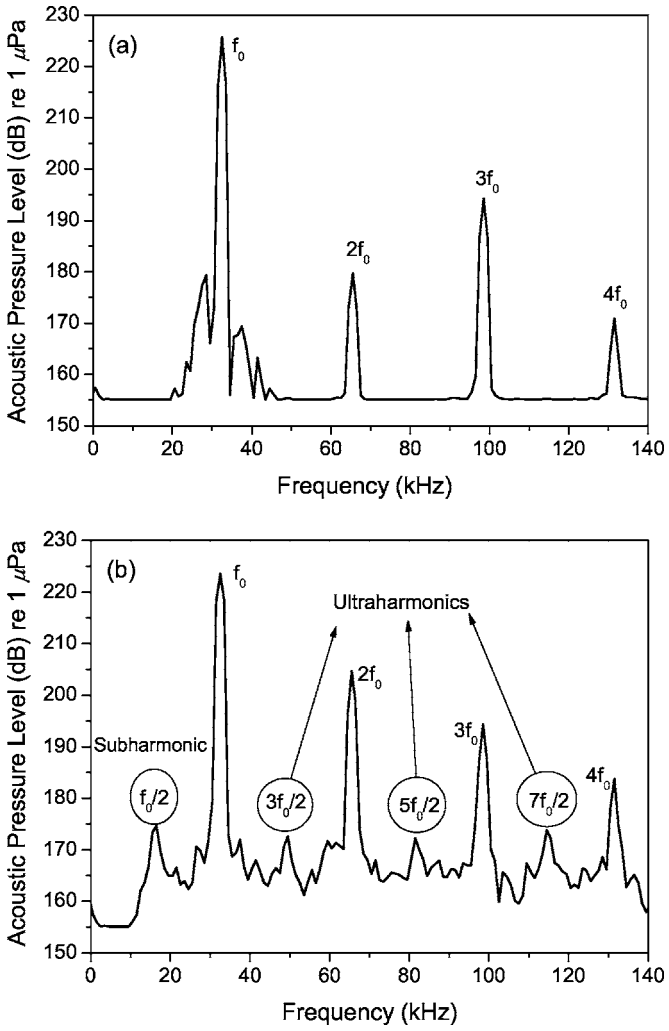


Fig. 5. Frequency spectra of signals transmitted through (a) pure water and (b) water-saturated sandy sediment at distance of 100 mm from the transducer with a resonance frequency of 33 kHz.

1999). In this study, the saturation phase of the acoustic pressure was only observed for the first ultraharmonic acoustic wave as shown in Fig. 3(b). For the subharmonic acoustic wave in Fig. 3(b), the saturation phase of the acoustic pressure might be observed at a driving acoustic pressure greater than 507 kPa. However, this could not be confirmed because of the limitation of the transmitting response of the transducer.

In Fig. 4, the subharmonic and the first ultraharmonic acoustic waves exhibited more fluctuations in pressure level than the primary acoustic wave as a function of receiving distance. Since the source of the primary acoustic wave in water-saturated sandy sediment was obviously the transducer, the primary acoustic wave coherently radiates from the vibrating plane of the transducer to the water-saturated sandy sediment. Therefore, the acoustic pressure level fluctuation of primary acoustic wave was hardly observed in sediment as shown in Fig. 4(a). However, since the nonlinear sources of the subharmonic and the first ultraharmonic acoustic waves could be randomly distributed in the nonlinear acoustic interaction zone of the water-saturated sandy sediment, the nonlinear acoustic waves might exhibit both coherent and incoherent properties. In this case, the acoustic pressure levels of both nonlinear acoustic waves might fluctuate

as shown in Fig. 4(b). Since high driving acoustic pressure amplitude at the transducer was required to generate the subharmonic and the first ultraharmonic acoustic waves in water-saturated sandy sediment, the rapid decreases of the acoustic pressure levels of both nonlinear acoustic waves in Fig. 4(b) might indicate that the nonlinear sources for the generation of the subharmonic and the ultraharmonic acoustic waves were mainly distributed in the vicinity of the transducer.

In water-saturated granular media, structural inhomogeneities, such as defects and discontinuities, are considered to be parametric resonance sources which generate the nonlinear acoustic waves. Specific investigation for structural inhomogeneities in granular media has been performed by Ostrovsky *et al.* (2000). They considered the granular media as media with soft and hard phases. The soft phases occupy small volumes in the media and they produce strong deformations for acoustic waves with finite amplitude, whereas the hard phases produce significantly less deformations. Since these strong deformations in granular media give rise to strong nonlinear acoustic responses, the deformations may be considered to be structural inhomogeneities. The contacts among individual grains in granular media are much softer than the grains. Therefore, the concentration of stress due to external pressure at the contact boundaries of individual grains may cause strong nonlinearity as a result of deformations in the media. Deformations at the contact boundaries of individual grains can ultimately be considered as deformations of pores in granular media, because the contact boundary surfaces between individual grains are parts of the pore surface in the media. Since the transducer in Fig. 1 was placed in water-saturated sandy sediment, the contacts on the boundary between the transducer and the sediment might be also considered as the nonlinear sources for the generation of the subharmonic and the ultraharmonic acoustic waves. However, it could not explain the subharmonic and the ultraharmonic phenomena observed through water-saturated sandy sediment slab separated from the transducer in water. Therefore, the subharmonic and the ultraharmonic phenomena in this study might be caused by various mechanisms based on the contacts between individual sediment grains and the contacts on the boundary between the transducer and the sediment.

The subharmonic and the ultraharmonic phenomena were first observed in water by Korpel and Adler (1965). Generally, the observations of these nonlinear phenomena are difficult in water because the nonlinear parameter of water is very small, around 3.5 (Beyer, 1998). However, they showed that the nonlinear phenomena could be observed in water if a standing wave pattern was made in the primary acoustic wave field. The theoretical approaches for this system have been performed by Adler and Breazeale (1970) and Hughes (1977). The subharmonic and the ultraharmonic phenomena can also be observed in parametric resonance systems, such as bubbly water and solid media with cracks, defects, and discontinuities (Ekimov *et al.*, 1996; Lostberg *et al.*, 1996; Shankar *et al.*, 1999; Solodov *et al.*, 2002; Korshak *et al.*, 2002). The Rayleigh-Plesset type equation (Lostberg *et al.*, 1996; Shankar *et al.*, 1999) for the motion of the bubbles is widely known as the nonlinear oscillation equation which predicts the subharmonic and the ultraharmonic phenomena in bubbly water. In solid media, such theoretical model equations are not yet well known, except nonlinear oscillation equations, such as Duffing (Fyrillas and Szeri, 1998) and nonlinear Mathieu (Boston, 1971) equations. They can be used for a phenomenological understanding of the subharmonic and the ultraharmonic phenomena in solid media. They have been practically used for understanding nonlinear phenomena in physical oscillation systems (Hayashi *et al.*, 1960; Fyrillas and Szeri, 1998).

5. Conclusions

Nonlinear acoustic waves were observed in water-saturated sandy sediment at the subharmonic and the ultraharmonic frequencies; they resulted from nonlinearity of the sediment. Such nonlinearity might be generated from the contacts between individual sediment grains and the contacts on the boundary between the transducer and the sediment. These nonlinear acoustic waves strongly depended on driving acoustic pressure. The experimental results in this study show that

the subharmonic and the ultraharmonic phenomena in water-saturated sandy sediment can be significantly observed when the driving acoustic pressure is greater than a threshold value. These phenomena show close resemblance to those produced in bubbly water.

Acknowledgments

The authors would like to thank Dr. Alexander M. Sutin at ARTANN Laboratories, Dr. Igor N. Didenkulov at Institute of Applied Physics in Russia, and anonymous reviewers for their valuable comments. This work was supported by the Agency for Defense Development, Republic of Korea.

References and links

Adler, L., and Breazeale, M. A. (**1970**). "Generation of fractional harmonics in a resonant ultrasonic wave system," J. Acoust. Soc. Am. **48**, 1077–1083.

Bazhenova, E. D., Vil'man, A. N., and Esipov, I. B. (**2005**). "Fluctuations of acoustic field in a granular medium," Acoust. Phys. **51**, S37–S42.

Beyer, R. T. (**1998**). "The parameter B/A," in *Nonlinear Acoustics*, edited by M. F. Hamilton, and D. T. Blackstock (Academic Press, San Diego), pp. 25–40.

Boston, J. R. (**1971**). "Response of a nonlinear form of the Mathieu equation," J. Acoust. Soc. Am. **49**, 299–305.

Ekimov, A. E., Lebedev, A. V., Ostrovsky, L. A., and Sutin, A. M. (**1996**). "Nonlinear acoustic effects due to cracks in ice cover," Acoust. Phys. **42**, 51–54.

Fyrillas, M. M., and Szeri, A. J. (**1998**). "Control of ultra- and subharmonic resonances," J. Nonlinear Sci. **8**, 131–159.

Hayashi, C., Nishikawa, Y., and Abe, M. (**1960**). "Subharmonic oscillations of order one half," IRE Trans. Circuit Theory **7**, 102–111.

Hughes, B. (**1977**). "Some theoretical aspects of the generation of surface ripples by parametric subharmonic resonance with sound waves," J. Acoust. Soc. Am. **61**, 407–412.

Korpel, A., and Adler, L. (**1965**). "Parametric phenomena observed on ultrasonic waves in water," Appl. Phys. Lett. **7**, 106–108.

Korshak, B. A., Solodov, I. Yu., and Ballad, E. M. (**2002**). "DC effects, sub-harmonics, stochasticity and 'memory' for contact acoustic nonlinearity," Ultrasonics **40**, 707–713.

Lostberg, O., Hovem, J. M., and Aksum, B. (**1996**). "Experimental observation of subharmonic oscillations in Infoson bubbles," J. Acoust. Soc. Am. **99**, 1366–1369.

Ostrovsky, L. A., Johnson, P. A., and Shankland, T. J. (**2000**). "The mechanism of strong nonlinear elasticity in earth solids," in *Nonlinear Acoustics at the Turn of the Millennium*, edited by W. Lauterborn and T. Kurz (AIP Press, New York), pp. 75–84.

Ostrovsky, L. A., Sutin, A. M., Soustova, I. A., Matveyev, A. L., Potapov, A. I., and Kluzek, Z. (**2003**). "Nonlinear scattering of acoustic waves by natural and artificially generated subsurface bubble layers in sea," J. Acoust. Soc. Am. **113**, 741–749.

Shankar, P. M., Krishna, P. D., and Newhouse, V. L. (**1999**). "Subharmonic backscattering from ultrasound contrast agents," J. Acoust. Soc. Am. **106**, 2104–2110.

Solodov, I. Yu., Krohn, N., and Busse, G. (**2002**). "CAN: an example of nonclassical acoustic nonlinearity in solids," Ultrasonics **40**, 621–625.

Sutin, A. M., Yoon, S. W., Kim, E. J., and Didenkulov, I. N. (**1998**). "Nonlinear acoustic method for bubble density measurements in water," J. Acoust. Soc. Am. **103**, 2377–2384.

Silent research vessels are not quiet

Egil Ona, Olav Rune Godø, Nils Olav Handegard, Vidar Hjellvik, Ruben Patel, and Geir Pedersen
Institute of Marine Research, P.O. Box 1870 Nordnes, 5817 Bergen, Norway
egil.ona@imr.no, olavrune@imr.no, nilsolav@imr.no, vidar.hjellvik@imr.no, ruben.patel@imr.no, geir.pedersen@imr.no

Abstract: Behavior of herring (*Clupea harengus*) is stimulated by two ocean-going research vessels; respectively designed with and without regard to radiated-noise-standards. Both vessels generate a reaction pattern, but, contrary to expectations, the reaction initiated by the silent vessel is stronger and more prolonged than the one initiated by the conventional vessel. The recommendations from the scientific community on noise-reduced designs were motivated by the expectation of minimizing bias on survey results caused by vessel-induced fish behavior. In conclusion, the candidate stimuli for vessel avoidance remain obscure. Noise reduction might be necessary but is insufficient to obtain stealth vessel assets during surveys.

PACS numbers: 43.80.Nd, 43.80.Ev, 43.30.Sf, 43.40.Rj [CM]
Date Received: December 12, 2006 **Date Accepted:** January 11, 2007

1. Introduction

Vessel-induced fish behavior during acoustic density estimation[1,2] and trawl sampling[3,4] may bias survey estimates of stock abundance. As noise has been considered a major stimulus, fisheries research institutions worldwide are investing in new silent research vessels[5] in accordance with recommendations from the International Council for the Exploration of the Sea (ICES).[6] The lack of fish avoidance observed from a stealth vessel[7] has been considered a result of reducing vessel noise,[7,8] but no direct comparison with a traditional research vessel has demonstrated this. Nevertheless, the scientific community has tacitly accepted that the major avoidance stimulus originates from the sound characteristics of the vessels, which is also the basis for the ICES recommendations.[1,6,9]

In 2003 the new Norwegian diesel-electric propulsioned research vessel "G. O. Sars" (GS) [Gross Registered Tonnage (GRT) 4067 tons, Length Overall (LOA) 77,5 m], fulfilling the ICES demands for a silent vessel, was put into operation. Several earlier reports have documented vessel avoidance of Norwegian spring spawning herring (*Clupea harengus*), showing that this stock is underestimated acoustically when distributed in the upper 100 m of the water column.[1,2] Currently the acoustic observations are used to establish an abundance index. Assuming similar or randomly varying conditions for observation among years, the time series of indices gives a relative change in abundance from one year to the next that is utilized in the stock evaluation. A vessel comparison was therefore an absolute necessity before the new GS could be used in data collection for the official annual stock assessments. Vessel comparison experiments were carried out in December 2004 in the Ofotfjord in northern Norway between GS and the previous standard vessel "Johan Hjort" (JH) (GRT 1828 tons, LOA 64,4 m). JH is a traditional research vessel with sound emission above the ICES standard[6] and thus far noisier than GS.

2. Methods

The two vessels followed each other at standard cruising speed and maximum distance along the exact same triangular cruise track [Fig. 1(a)]. Both vessels collected acoustic data according to the standard protocol for acoustic surveys, and, along the pursue track, an upward-looking echosounder and an acoustic Doppler current profiler (ADCP) were placed. This allowed us to record herring density by depth as well as the mean swimming velocity of the fish layer during

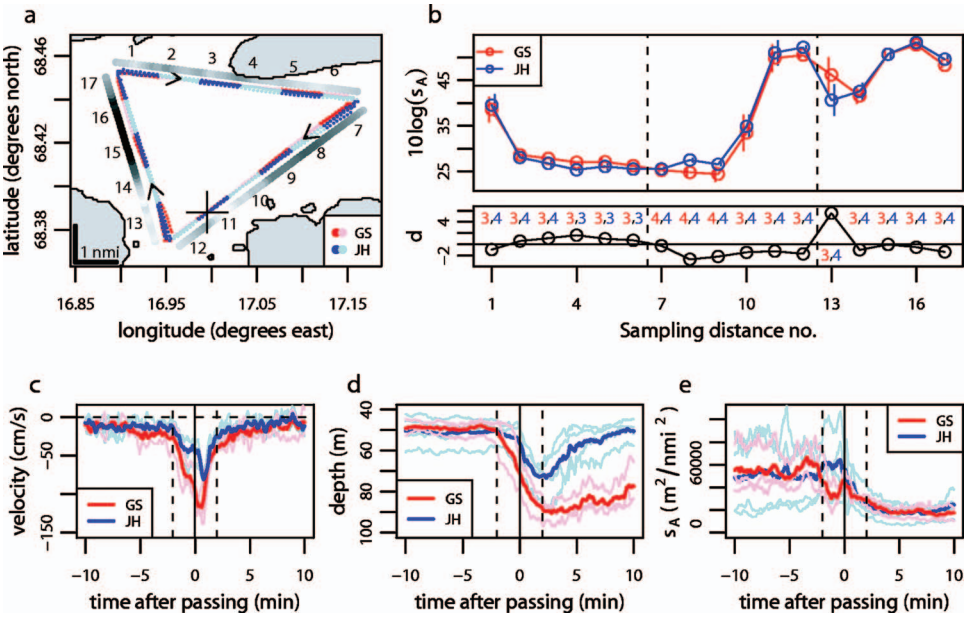


Fig. 1. Results from the pursuit experiment and mooring passages. (a) The numerical-density vessel-transects integrated over intervals of 1 nmi, shown as numbered light and dark sequences of points. Each point indicates the start of a 0.1 nmi sampling distance. Gray points indicate bottom depth (lightest: about 60 m, darkest: about 520 m). The cross at sampling distance no. 11 indicates the position of the moored platforms. Arrows indicate cruising direction. (b), upper panel. Numerical density along the transects. Each point represents the average of the log-transformed numerical densities for one vessel at one interval. The error bars show ±2 standard errors of the averages. (b) lower panel: The corresponding vessel differences in numerical density, and the number of passages for each vessel. Dashed vertical lines indicate the corners of the triangle in (a). (c) Vertical swimming velocity component when GS (red curves) or JH (blue curves) passed the moored ADCP. Thin lines denote single passages, and thick lines denote the averages over all passages. Vertical dotted lines are drawn 2 min before and after the point of passage. (d) Vertical fish distribution (median depth) when GS (red curves) or JH (blue curves) passed the moored echosounder. Thin lines denote single passages and thick lines denote the averages over all passages. Vertical dashed lines are drawn 2 min before and after the point of passage. (e) Average s_A when GS (red curves) or JH (blue curves) passed the moored echosounder. Thin lines denote single passages and thick lines denote the averages over all passages. Vertical dotted lines are drawn 2 min before and after the point of passage.

and between passages of the two vessels. The mooring was passed four times by JH and three times by GS at moderately dense recordings of herring at 40–80-m depth during the night. The two vessels were completely darkened during the experiments. First, we present the method and protocol for collecting and analyzing the data from the vessel-mounted echosounders. Then, the method to collect and analyze the data from the bottom-mounted platforms is presented, including noise level confirmation for the vessels.

Raw echosounder data were recorded using the Simrad EK500 (Kongsberg Gruppen, Kongsberg, Norway), 18 kHz on JH, and the Simrad EK60, 18 kHz on GS. The calibrated raw data were directly transferred to the format of the postprocessing system BEI, and scrutinized by standard procedures for herring surveys by the same two operators. In fjord surveys with high densities, this basically involves the removal of bottom detection errors and isolating the herring layers by integration boundaries. The processed herring density data were stored in a database with 0.1 nmi × 10-m depth bins in absolute, linear units for the "nautical area scattering coefficient," s_A (m^2/nmi^2), a standard unit in fisheries acoustics.[10]

The s_A values from the two vessels were compared using a standard method,[11] slightly modified since the method is designed for two vessels following parallel transects, whereas in our experiment each part of the survey transect was traversed 3–4 times by each vessel. Also, we have chosen to replace $\ln(x)$ in the published method with $10\log(x)$. The modified method is as

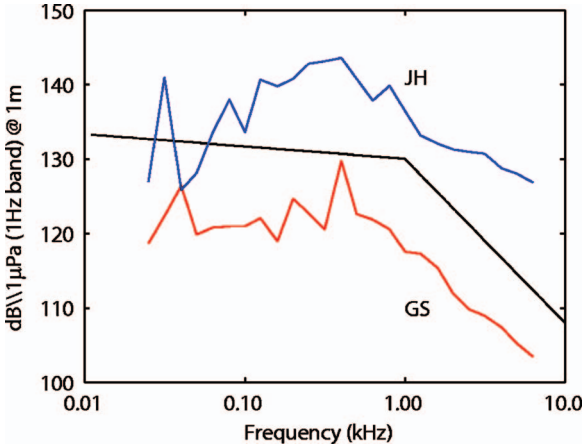


Fig. 2. Noise signatures of JH (blue) and GS (red), respectively, measured in the keel aspect of both ships using the Naxys hydrophone. The maximum recommended levels of noise from a free-running survey vessel at any speed up to and including 11 knots according to ICES CRR 209[6] is also included in the figure (black curve). Due to a possible inaccuracy in the hydrophone positioning, the absolute levels may be 3 dB off. This does not affect the relative difference between the vessels.

follows: Let ρ_i be the true fish density at elementary sampling distance unit i, and $x_{ij} = \alpha_j \rho_i 10^{\sigma\varepsilon_{ij}}$ the density measured at sampling distance i by vessel j, $j=\{1,2\}$, averaged over all passages (there is no visible trend in the measured densities over the time period used in the analysis). Here, α_j is a vessel-dependent bias, σ^2 is the variance, and ε_{ij} is random noise. Defining $d_i = 10\log(x_{i1}) - 10\log(x_{i2})$, we have $d_i = \delta + \sigma\varepsilon_i'$, where $\delta = 10\log(\alpha_1/\alpha_2)$ and $\varepsilon_i' = 10(\varepsilon_{i1} - \varepsilon_{i2})$. Testing the null hypothesis that $H_0: \alpha_1 = \alpha_2$ is equivalent with testing $H_0: \delta = 0$, and if the d_i are independent, this can be done using a two-sided t-test. The ratio α_1/α_2 is estimated by $10^{\bar{d}/10}$, where $\bar{d} = \Sigma_i d_i$.

The horizontal resolution of the data is 0.1 nmi, but using this resolution in the analysis would yield autocorrelated d_i, and the t-test could not be used. We have therefore aggregated the data so that each sampling distance is 1.0 nmi [Fig. 1(a)]. This has another advantage as well: the sampling distances for the two vessels do not match exactly, but the mismatch is smaller relative to the sampling length when this is increased (or resolution is decreased).

At a selected position in both surveys [Fig. 1(a)], the vessels passed directly above a bottom-moored platform, carrying a calibrated, upward-looking EK60, 38 kHz echosounder, and a calibrated underwater hydrophone (Naxys A/S, Bergen, Norway) with a computer for digital sound recordings. The hydrophone was used to verify the noise levels from the two vessels [Fig. 2]. An upward-looking bottom-moored 75-kHz acoustic Doppler current profiler (Teledyne RD Instruments, San Diego) 50 m perpendicular to the track line was used to measure the swimming speed of the herring layer. The raw backscattering data from all of the four ADCP beams were used to create a mask to isolate the herring layers from the surrounding water, and to calculate the resulting vertical and horizontal velocity components.

Avoidance reactions were analyzed using data from seven vessel passages over the moored platforms, four by JH and three by GS, including data from 10 min before each passage to 10 min after. The acoustic backscatter from the moored echosounder was integrated over the fish layer and is presented in s_A units. Herring depth distribution and s_A were available with a resolution of about 3 pings per second, and ADCP data with one recording every fifth second. To remove some of the random noise, the depth and s_A data were smoothed using a moving average with a window of 10 s. The ADCP data were smoothed using a window of 15 s for the vertical component and 25 s for the horizontal component. Vessel differences and effects of vessel passages were tested for using two-sided t-tests with $n=3$ (JH), 4 (GS), or 3+4 (both

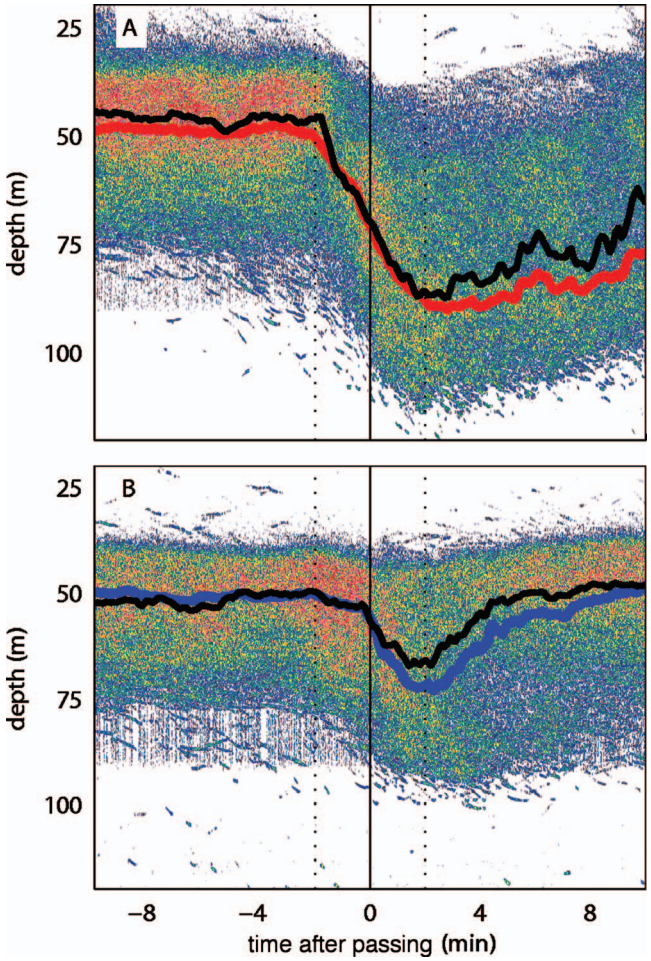


Fig. 3. The echogram for a single passage over the moored echosounder when passed by GS (a) and JH (b), respectively. The black line is the median depth distribution for this passage, and the blue and red lines are the means of the median depth distributions for all passages for JH and GS, respectively.

vessels). For example, when testing for a vessel difference in vertical displacement, the difference $D_{\mathrm{JH},i}=\bar{q}_{50,i,\mathrm{before}}-\bar{q}_{50,i,\mathrm{after}}$ between the median depths averaged over the time periods 5−2 min before and 2−5 min after passage i, $i=\{1,\ldots,4\}$, was calculated for JH, and similarly, $D_{\mathrm{GS},j}$, $j=\{1,\ldots,3\}$ was calculated for GS. A difference between D_{JH} and D_{GS} was then tested for using a standard t-test with $n=3+4$, not assuming equal variance.

3. Results and discussion

First, we show that the herring numerical densities measured from the two vessels were similar. The average numerical density recorded by GS during the pursuit experiment was 97% of the average recorded by JH, which is a nonsignificant difference ($p=0.75$ under the null hypothesis of no vessel differences).

Next, we show that GS initiated a more intense and prolonged avoidance reaction than JH by analyzing the results from the moored instrumentation. The vertical mean swimming velocity as estimated by the ADCP during the period of passage (from 2 min before to 2 min after) was significantly higher for GS (−70.5 cm/s) than for JH (−42.7 cm/s) [t-test, $p=0.009$, $n=3+4$, Fig. 1(c)]. For GS the mean vertical velocity of the fish layer corresponds to

$2-3$ bl s^{-1} (body lengths per second), while the maximum recorded corresponds to about 4 bl s^{-1}. Since the horizontal velocity component is not taken into account, and as variation is expected for individuals over the depth range of the layer, this is a strong reaction compared to the maximum individual swimming speed for herring of $6-7$ bl s^{-1}.[12] The horizontal swimming speed in the same period was similar for the two vessels.

From the moored echosounder (Fig. 3), the diving is observed as a vertical displacement of the herring layer [Fig. 1(d)]. From about 2 min before passage and to 2 min after passage a median displacement of about 20 and 40 m is seen for JH and GS, respectively [Fig. 1(d) and Fig. 3]. The difference between the vessels in terms of change in median depth distribution from $5-2$ min before passage to $2-5$ min after passage is significant on a 1% level (t-test, $n=3+4$, $p=0.003$). The fish also needed more time to return to its original distribution after being disturbed by GS [Fig. 1(d)].

The numerical density when passing the mooring was on average similar to that before vessel passage [Fig. 1(e)]. The variability in numerical density was large during the experiment, and it is difficult to separate vessel-induced effects from natural variations on a detailed level based on only seven passages. Nevertheless, there was a clear decrease in numerical density after the vessels had passed. Comparing the intervals $2-5$ min before passage and $2-5$ min after passage, the decrease is significant on a 5% level both for GS (t-test, $p=0.022$, $n=3$) and JH (t-test, $p=0.045$, $n=4$).

4. Conclusion

Although the moored echosounder did not record any significant difference in numerical density before and during passage for any vessel, the fish reaction pattern is strong and clearly vessel dependent. The differences in behavioral response and the magnitude of the response demonstrate the potential to cause severe bias, particularly seen in the perspective of earlier experience.[1,2] This illustrates the complexity of the vessel avoidance behavior, but more important, the results show that a stimulus other than noise, as defined by ICES, must be responsible for the reaction. Silent vessels have many advantages for reliable acoustic surveys, e.g., improving signal-to-noise ratio. Reducing the vessel noise may be necessary but is not a sufficient measure, and as long as influential candidate stimuli for fish avoidance remain obscure, the ICES goal of a establishing a stealth vessel design appears unrealistic.

Acknowledgments

This work was partially supported by grants from the Norwegian Research Council. We are grateful to the engineers Ronald Pedersen, Ingvald Svellingen, Atle Totland, and Terje Torkelsen who established and operated the instrumentation. Captains and crews on G. O. Sars and Johan Hjort are thanked for their patience and cooperation during the experiments.

References and links

[1]K. Olsen, J. Angell, F. Pettersen, and A. Løvik, "Observed fish reactions to a surveying vessel with special reference to herring, cod, capelin and polar cod," FAO Fish. Rep. **300**, 131–138 (1983).
[2]R. Vabø, K. Olsen, and I. Huse, "The effect of vessel avoidance of wintering Norwegian spring spawning herring," Fish. Res. **58**, 59–77 (2002).
[3]E. Ona and O. R. Godø, "Fish reaction to trawling noise: The significance for trawl sampling," Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **189**, 159–166 (1990).
[4]N. O. Handegard and D. Tjøstheim, "When fish meets a trawling vessel: Examining the behaviour of gadoids using a free floating buoy and acoustic split-beam tracking," Can. J. Fish. Aquat. Sci. **62**, 2409–2422 (2005).
[5]J. Moore, "Stealth ship sets sail for a quiet life fishing for data," Nature (London) **423**, 7 (2003).
[6]R. B. Mitson, "Research vessel standards: Underwater noise of research vessels, Review and Recommendations," ICES Co-op. Res. Rep. **209**, 1–61 (1995).
[7]P. G. Fernandes, A. S. Brierley, E. J. Simmonds, N. W. Millard, S. D. McPhail, F. Armstrong, P. Stevenson, and M. Squires, "Fish do not avoid survey vessels," Nature (London) **404**, 35–36 (2000).
[8]P. G. Fernandes, A. S. Brierley, E. J. Simmonds, N. W. Millard, S. D. McPhail, F. Armstrong, P. Stevenson, and M. Squires, "Fish do not avoid survey vessels (addendum)," Nature (London) **407**, 152 (2000).
[9]R. B. Mitson and H. P. Knudsen, "Causes and effects of underwater noise on fish abundance estimation," Aquat. Liv. Res. **16**, 255–263 (2003).
[10]D. N. Mac Lennan, P. G. Fernandes, and J. Dalen, "A consistent approach to definitions and symbols in

fisheries acoustics," ICES J. Mar. Sci. **59**, 365–369 (2002).
[11]R. Kieser, T. J. Mulligan, N. J. Williamson, and M. O. Nelson, "Intercalibration of two echo integration systems based on acoustic backscattering measurements," Can. J. Fish. Aquat. Sci. **44**, 562–572 (1987).
[12]J. H. S. Blaxter and W. Dickson, "Observations of the swimming speeds of fish," J. Cons., Cons. Int. Explor. Mer **24**, 472–479 (1959).

Sex differences in the length of the organ of Corti in humans

James D. Miller
Communication Disorders Technology, Inc., Indiana University Research Park, 501 N. Morton Street, Suite 215, Bloomington, Indiana 47404
jamdmill@indiana.edu

Abstract: Sato *et al.* [Acta. Otolaryngol. **111** (6), 1037–1040 (1991)] reported that the human cochlea is, on average, 15% longer for males than females. This corresponds to 4.7 mm in length and to 2.78 standard deviations (SD). Anatomical measurements of the lengths of cochleas from 148 heads (194 cochleas) from eleven sources are reviewed and summarized. A sex difference of 3.36% is observed. This corresponds to 1.11 mm in length and to 0.49 SD. The mean lengths of the male and female cochleas are approximately 34 and 33 mm, respectively, and the population SD is 2.28 mm. The statistical significance of the observed difference is questionable.

PACS numbers: 43.64.Dw, 43.66.Ba [BLM]
Date Received: September 29, 2006 **Date Accepted:** January 23, 2007

1. Introduction

The length of the organ of Corti (OC) in relation to range of hearing has been strongly implicated as a predictor of the frequency resolving power of the ear (Békésy,[1] Békésy and Rosenblith,[2] and see Fay[3] for a review). This position is strengthened by the fact that in human beings the density of outer hair cells does not change with cochlear length, and the density of inner cells also does not change with cochlear length, except for the most apical few mm (Wright *et al.*).[4] Bohne and Carr[5] found a similar result for the chinchilla. This means that longer cochleas have more inner and outer hair cells than do shorter cochleas. Bohne *et al.*[6] conclude their study of myelinated nerve fibers in the chinchilla cochlea with the statement, "In view of the present results, it is reasonable to expect longer cochleas (which contain more sensory cells) to have more spiral ganglion cells." Nadol,[7] a leading student of the human eighth nerve, concluded that for humans, longer organs of Corti with more inner and outer hair cells may have more eighth-nerve afferent fibers. For these reasons, a possible sex difference in the length of the cochlea is important as it might imply important sex differences in the numbers of sensory cells, numbers of primary afferent fibers, and in auditory function. One study of the human cochlea (Sato *et al.*[8]) reported a large sex difference in length of the organ of Corti. Electrophysiological measurement of response delay times combined with assumptions about cochlear mechanics led Don *et al.*[9] to a similar conclusion. Casual examination of other sets of anatomical measurements did not indicate to the present author that such a large sex difference existed, which led to the following review and summary of studies that measured the anatomical lengths of the cochlea for both women and men.

2. Methods used to measure the anatomical length of the cochlea

Four anatomical methods have been used to measure the length of the OC. (1) **The surface preparation method**. In the middle 1800's, Retzius[10] found that the organ of Corti and the basilar membrane were of the same length and that the cochlear duct (scala media) was another 1.5–1.8 mm longer. Retzius used a dissection method similar to the surface preparation method that was later used by Bredberg,[11] Ulehlova *et al.*,[12] Wright *et al.*,[4] Leake,[13] and Leake *et al.*[14] By this method, one looks down on the organ of Corti and measures its length along the clearly defined junction of the outer pillar cell with the first outer hair cell. Individual pieces of the OC are cut and laid flat, each piece measured, and the total length taken as the sum of the lengths of

the pieces. Takagi and Sando[15] state that this method is accurate, the only problem being the possible loss of tissue during the process of cutting the pieces to be measured. (2) **The serial section method**. This method is based on serial sections and a projection of the loci of the junction between the heads of the pillar cells onto a plane (a 2D representation). The earliest version of this method was described by Guild[16] and is called the Guild method. According to Bredberg,[11] the Guild method ignored about 1 mm of the basal hook of the cochlear duct. The Guild method was modified by Schuknecht[17] to include complete measurement of the basal hook and is known as the Guild/Schuknecht method. Both of these two-dimensional methods result in shorter lengths than the surface preparation method because: (1) the shorter radius of the cochlear spiral (the junction of the heads of the pillars is more medial than the junction of the outer pillar with first outer hair cell), (2) the rise in the elevation of the cochlear spiral is not included in the 2D projection, and (3) the possibility that, if the sections are not exactly parallel to the mid-modiolar axis, the projections may be foreshortened (Takagi and Sando).[15] It is found here that measures made by the Guild method can be brought into agreement with those made by the surface preparation by adding 1.0 mm for the hook, as suggested by Bredberg,[11] and then multiplying by 1.039 to take into account the radius, elevation, and possible foreshortening. Measures made by the Guild/Schuknecht method only need to be multiplied by 1.039 to be brought into agreement with the measures made from surface preparations. These corrections were found by trial and error to bring the means of the Guild and Guild/Schuknecht methods close to those found by the surface preparation method. (3) **The 3D reconstruction method**. This method was introduced by Takagi and Sando.[15] Their method uses serial sections of the cochlea, but the 3D coordinates of the junctions of the pillar heads in each serial section are entered into a computer in relation to reference points that do not require that the sections be exactly parallel to the axis of the modiolus. A three-dimensional representation of the line formed by the path of the junction of the pillar heads is created in the computer and its length calculated. In a study of a single cochlea, they found good agreement with the Guild/Schuknecht method when the serial sections were corrected by computer to be parallel to the mid-modiolar axis. This 3D method was applied by Sato *et al.*[8] to study sex differences. However, unlike previous investigators, they measured along the inner and outer borders of the basilar membrane (BM) and took the average of the two lengths to represent cochlear length. Their 3D reconstruction of these measurements includes the elevation of the cochlear spiral and has a larger radius than the surface method, as halfway between the inner and outer edges of the BM may fall nearer to the second or third row of outer hair cells than to the junction of the outer pillar with the first outer hair cell. It will be shown that Sato *et al.*[8] found the male average to be about 3 mm longer and the female average to be about 1 mm shorter than the same averages found by other methods. The reasons for these differences are unknown. (4) **The CT method**. This method is a 3D reconstruction of the cochlea based on *in vivo* CT scans of the temporal bone and was introduced by Ketten *et al.*[18] and Skinner *et al.*[19] With this method, none of the soft tissue of the cochlea is visible. The centroid of the bony cochlear canal is located in each section of the cochlea and a mathematical spiral is "fitted" the 3D array of centroids so generated.

3. Handling of the gleaned measurements

The literature was searched for measured sex-identified cochleas. If both right and left cochleas were measured for the same individual, the average of two lengths was used. This was deemed appropriate as Bohne *et al.*[20] found for 151 chinchilla cochleas that the correlation between right and left lengths was 0.96, and it was found here for 46 human cochleas that the correlation between right and left lengths was 0.74. In cases where only one cochlea was measured for an individual, possible laterality effects were ignored as the proportion of right and left cochleas did not differ substantially between the sexes (60% right for males and 55% right for females). In addition, when all 46 cases for which both right and left cochleas were measured are considered, the average difference was 0.5 mm in favor of the right ear. This difference was not statistically significant and only amounted to 0.20 standard deviation units.

Table 1. Lengths of male and female cochleas in mm.

Citation/ Method	Male	N	Female	N	σ[a]	M-F	((M-F)/σ)[a]	M/F
21/Guild	33.69	37	33.18	9	2.46	0.51	0.21	1.02
22/Guild/ Schuknecht	34.90	5	32.86	5	1.98	2.04	1.03	1.06
23/Guild/ Schuknecht	34.25	12	33.93	4	2.33	0.32	0.14	1.01
24/Guild/ Schuknecht	30.01	4	29.09	1	3.357	0.92	0.27	1.03
10/Surface	33.65	2	32.00	1	0.50	1.65	3.30	1.05
11/Surface	34.43	21	33.34	5	1.24	1.09	0.88	1.03
13,14/Surface	33.62	6	32.15	3	2.11	1.47	0.70	1.05
18/CT	33.44	7	32.75	13	2.37	0.69	0.29	1.02
19/CT	34.66	6	34.58	7	1.287	0.08	0.06	1.00
8/3D	37.09	9	32.37	9	1.70	4.72	2.78	1.15
Sums		109		57				
Weighted Averages	34.13		33.02		2.28[b]	1.11	0.49	1.03

[a]Pooled estimate.
[b]Calculated by combining all M-data and all F-data and finding the pooled estimate of σ.

4. Results

Measurements from eleven sources are summarized in Table 1. Note that the lengths from Hardy[21] were converted by adding 1 mm and then multiplying by 1.039. Lengths from Walby,[22] Hinojosa *et al.*,[23] and Pollak *et al.*[24] were multiplied by 1.039. As described in the discussion of the serial section methods above, these conversions make the measurements using the Guild and Guild/Schuknecht methods comparable to those made by the surface methods. Also, note that no data from Ulehlova *et al.*[12] and from Wright *et al.*[4] were included, as the former only reported on male cochleas and the latter did not identify the sex of the cochleas.

As can be seen in Table 1, the mean length of the cochlea for each of the eleven sets of measurements is longer for males than for females. However, by *t* test none of these differences are statistically significant ($p=0.05$) except for the results of Sato *et al.*[8] They find a substantial difference of 15% and 2.78 standard deviation units. Sato *et al.*[8] find the average length of the male cochlea to be 37.1 mm, whereas the average length for the ten other studies is 33.9 mm. Sato *et al.*[8] find the average length of the female cochlea to be 32.4 mm, whereas the average length for the other ten studies is 33.1 mm. There are three possible explanations of this discrepancy. (1) The Sato *et al.*[8] results simply represent a Type I sampling error, in which case they should be averaged with the others. (2) The Sato *et al.*[8] data may contain some unknown factor or error that leads to an overestimation the of lengths of the male cochleas. (3) The Sato *et al.*[8] data are, in fact, the most accurate data and represent the true state of sexual dimorphism in the lengths of the human cochlea. It seems prudent to assume that explanation 1 is correct, as it represents all of the available data. Explanation 3 can only be verified by careful, comparative studies of two or more of the methods as applied to many cochleas. Until such studies are conducted and prove otherwise, it appears that there may be a sex difference in cochlear length of 1.11 mm, which amounts to about 3.36% and represents 0.49 standard deviation units. However, whether the observed difference is statistically significant is questionable. One method of analysis was to pool the measurements, as if they had been collected in one study of 109 male and 57 female cochleas, and conduct a *t* test with 164 degrees of freedom. This method resulted in a statistically significant difference with $p=0.003$. If the data of Sato *et al.*[8] are dropped from the pool, the *t* value failed to reach significance with $p=0.064$. Also, a statistical meta-analysis

was conducted using a method described by Hedges and Olkin.[25] An unbiased estimate of the effect (d defined on p. 81 of Ref. 25) was calculated for each of the eleven sets of data, and the weighted average (d_+ defined on p. 111 of Ref. 25) was found. The effect size, so defined, was found to be 0.37 units with a 95% confidence interval of ± 1.50 units (defined on p. 112–113 of Ref. 25). By this method, the null hypothesis could not be rejected.

In summary, the average observed length of the male cochlea is about 34 mm, and the average observed length of the female cochlea is about 33 mm. The population standard deviation is 2.28 mm. The range of the 166 cochlear lengths that comprise the data base studied here is 13.78 mm, which is consistent with calculated 6 SD range of 13.68 mm. Thus, the observed sex difference in cochlear length is small in comparison to the observed variability of the lengths of normal cochleas. It is tentatively concluded that there may be a small difference in the lengths of male and female human cochleas even though statistical analyses of the data are not decisive.

Acknowledgments

The author wishes to thank P. A. Leake of the UCSF School of Medicine for providing cochlear measurements. Also, the following people provided thoughtful reviews of the manuscript: Barbara A. Bohne and Gary W. Harding of Washington University School of Medicine, Dennis McFadden of the University of Texas, and J. C. Saunders of the University of Pennsylvania Medical School.

References and links

[1]G. von Békésy, "Über die mechanische Frequenz-analyse in der Schnecke verscheider ("On the mechanisms of frequency analysis in various cochleas")," Tiere Akust. Z. **9**, 3–11 (1944).
[2]G. von Békésy and W. A. Rosenblith, "The mechanical properties of the ear," in *The Handbook of Experimental Psychology*, edited by S. S. Stevens (Wiley, New York, 1951), pp. 1075–1115.
[3]R. R. Fay, "Structure and function in sound discrimination among vertebrates," in *The Evolutionary Biology of Hearing*, edited by A. N. Popper, R. R. Fay, and D. B. Webster (Springer-Verlag, New York, 1992), pp. 229–263.
[4]A. Wright, A. Davis, G. Bredberg, L. Ulehlova, and H. Spencer, "Hair cell distributions in the normal human cochlea,"Acta Oto-Laryngol., Suppl. **444**, 1–48 (1987).
[5]B. A. Bohne and C. D. Carr, "Location of structurally similar areas in chinchilla cochleas of different lengths,"J. Acoust. Soc. Am. **66**, 411–414 (1979).
[6]B. A. Bohne, A. Kenworthy, and C. D. Carr, "Density of myelinated nerve fibers in the chinchilla cochlea,"J. Acoust. Soc. Am. **72**, 102–107 (1982).
[7]J. B. Nadol, "Quantification of human spiral ganglion cells by serial section reconstruction and segmental density estimates,"Am. J. Otolaryngol. **9**, 47–51 (1988).
[8]H. Sato, I. Sando, and H. Takahashi, "Sexual dimorphism and the development of the human cochlea: Computer 3-D measurement,"Acta Oto-Laryngol. **111**, 1037–1040 (1991).
[9]M. Don, C. W. Ponton, J. J. Eggermont, and A. Masuda, "Gender differences in cochlear response time: An explanation amplitude differences in the unmasked brain-stem response,"J. Acoust. Soc. Am. **94**, 2135–2148 (1993).
[10]G. Retzius, *Das Gehororgan der Wirbeliere, Vol. II: Das Gehororgan der Reptilian, der Vogel, und der Saugethiere (The hearing organs of vertebrates, Vol. II: The hearing organs of reptiles, birds, and mammals)* (Samson & Wallin, Stockholm, 1884), p. 368.
[11]G. Bredberg, "Cellular pattern and nerve supply of the human organ of Corti," Acta Oto-Laryngol., Suppl. **236**, 1–135 (1968).
[12]L. Ulehlova, L. Voldrich, and R. Janisch, "Correlative study of sensory cell density and cochlear length in humans,"Hear. Res. **28**, 149–151 (1987).
[13]P. A. Leake, Personal communication (2006).
[14]P. A. Leake, O. A. Stakhovskaya, and S. Sridhar, "Protective and Plastic Effects of Patterned Electrical Stimulation on the Deafened Auditory System" (Progress Report: Dept. Otolaryngol-HNS, Univ. of California, San Francisco) (2005), pp. 1–16.
[15]A. Takagi and I. Sando, "Computer-aided three-dimensional reconstruction: A method of measuring temporal bone structures including the length of the cochlea,"Ann. Otol. Rhinol. Laryngol. **98**, 515–522 (1989).
[16]S. R. Guild, "A graphic reconstruction method for the study of the organ of Corti," Anat. Rec. **22** 141–157 (1921).
[17]H. F. Schuknecht, "Techniques for the study of cochlear function and pathology in experimental animals," Arch. Otolaryngol. **58**, 377–397 (1953).
[18]D. R. Ketten, M. W. Skinner, G. Wang, M. W. Vannier, G. A. Gates, and J. G. Neely, "In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays," Ann. Otol. Rhinol.

Laryngol. **107**, 1–16 (1998).
[19]M. W. Skinner, D. R. Ketten, L. K. Holden, G. W. Harding, P. G. Smith, G. A. Gates, J. G. Neely, G. R. Kletzker, B. Brunsden, and B. Blocker, "CT-derived estimation of cochlear morphology and electrode array position in relation to word recognition in Nucleus-22 recipients," JARO-J. Assoc. Res. Otolaryngol. **03**, 332–350 (2002).
[20]B. A. Bohne, D. G. Bozzay, and G. W. Harding, "Interaural correlations in nomal and traumatized cochleas: Length and sensory cell loss," J. Acoust. Soc. Am. **80**, 1729–1736 (1986).
[21]M. Hardy, "The length of the organ of Corti in man," Am. J. Anat. **62**, 291–311 (1938).
[22]A. P. Walby, "Scala tympani measurement," Ann. Otol. Rhinol. Laryngol. **94**, 393–397 (1985).
[23]R. Hinojosa, R. Seligsohn, and S. A. Lerner, "Ganglion cell counts in the cochleae of patients with normal audiograms.," Acta Oto-Laryngol. **99**, 8–13 (1985).
[24]A. Pollak, H. Felix, and A. Schrott, "Methodological aspects of quantitative study of spiral ganglion cells," Acta Oto-Laryngol., Suppl. **436**, 37–42 (1987).
[25]L. V. Hedges and I. Olkin, *Statistical Methods for Meta-Analysis* (Academic Press, San Diego, 1985), pp. 75–113.

Breathing noise elimination in through-water speech communication between divers

B. Woodward and H. Sari
Department of Electronic and Electrical Engineering, Loughborough University, LE11 3TU, United Kingdom
b.woodward1@lboro.ac.uk, h.sari@lboro.ac.uk

Abstract: Breathing noise and bubble noise are the main factors affecting the subjective quality of through-water speech signals in communications between divers wearing full-face masks or aural-nasal masks. Only breathing noise is considered here, which can be gated out by applying a combination of zero-crossing detection and energy measurements to noisy speech signals above predetermined threshold values. The signals are picked up by a microphone placed close to the diver's mouth in the air cavity of the mask. Results were obtained during diving trials with four different types of masks.

PACS numbers: 43.50.Ed, 43.60.Cg, 43.60.Dh, 43.72.Dv [MRS]
Date Received: November 23, 2006 **Date Accepted:** January 29, 2007

1. Introduction

In diver communications, clear speech transmission is essential for safe diving practice and for operational efficiency.[1–3] Among the factors affecting speech quality are noise and distortion, which are introduced at the source, in the water channel and at the receiver. Distortion at the source depends on the type of microphone used, the shape of the diver's mask, the breathing gas constituents, the ambient pressure and the difficulty of producing speech underwater. The effect of the mask on speech quality has been the subject of a separate study, aimed at designing a digital filter to compensate for the mask response.[3] In this paper, no attempt is made to compensate for channel noise, which can be from extraneous sounds produced by boats and animals, bubbles in the water column, wave action, rain on the water surface, and many other factors. Only source noise produced by a diver is considered here.

The two types of source noise affecting speech quality for a diver wearing conventional Self-Contained Underwater Breathing Apparatus (SCUBA) with a full-face mask or an auxiliary aural mask are:

(i) breathing noise, which is produced by air flow through the regulator (demand valve) during inhalation and which, due to its high amplitude, has a significant effect on the quality of communications;
(ii) bubble noise, which is generated during both exhalation and speech by air released from the regulator, and although emanating from outside the mask, it can be detected by a microphone inside the mask cavity.

For a diver wearing a helmet or band-mask, there is also noise from free-flowing air.

When speaking in air normally, unlike in a diving mask, there is no noise generated that is comparable to "breathing noise," hence there has been no need for the kind of processing adopted here. It may be possible to use active noise control to cancel the breathing noise produced underwater, but it would be more complex to implement because it would require the generation of a wave form that matched the range of frequencies of the noise and in precise anti-phase with it.

Breathing noise suppression has been attempted by using analogue methods applied in *hardwired* communication systems.[1,4] One method was to feed the input signal to a bandpass filter tuned to the center frequency of the breathing noise, then apply rectification and integration. When the signal exceeded a preset amplitude threshold, indicating the presence of breathing noise, an attenuator was activated to eliminate it. This method depended on the breathing

Fig. 1. (Color online) Divers' masks used during underwater speech processing tests.

noise amplitude being higher than any part of the speech signal. Another method employed two microphones, one for speech placed inside the mask, the other for noise placed inside the regulator. The noise microphone, which was encapsulated so as not to pick up speech and bubble noise, detected when the diver inhaled then activated an attenuator. The main problem with these methods was that the diver's mask or regulator needed to be modified with extra hardware, which could have safety implications.

Unlike the earlier techniques, the aim here is to apply a digital signal processing method to achieve improved underwater acoustic voice communications by reducing or eliminating the breathing noise whilst leaving the diver's speech undistorted.[2,3,5]

2. Speech recording and processing

For speech recording trials, four commercially available diving masks (*Wet Mask*, *Aga*, *Exo-26*, and *Aqua Lung*, shown in Fig. 1) and their associated regulators were worn in turn by a diver submerged in a laboratory tank (9 m long, 5.5 m wide, 2 m deep). The diver's speech was detected by an electret microphone (Type CF 2949, Knowles Ltd.), which can compensate for ambient pressure changes, and recorded on an instrumentation-quality tape recorder (Nagra III), with a frequency response of ±1 dB from 30 Hz to 18 kHz and a signal-to-noise ratio of 75 dB at a tape speed of 15 inches per second. The speech signals were fed to the audio input port of a computer, which had a "flat response" low-pass filter with a cut-off frequency of 3.4 kHz and an analog-to-digital converter operating at a sampling rate of 8 kHz with 8-bit resolution.

3. Breathing noise elimination

It is clear that the speech signal and the breathing noise are mutually exclusive in the time domain, since a diver cannot speak and breathe in simultaneously. Generally, the breathing noise amplitude is higher than the speech signal amplitude, a factor that can have a significant effect on the clarity of communications. In the frequency domain, breathing noise is characterized as broadband throughout the 1–4 kHz band, whereas for speech signals, voiced sounds

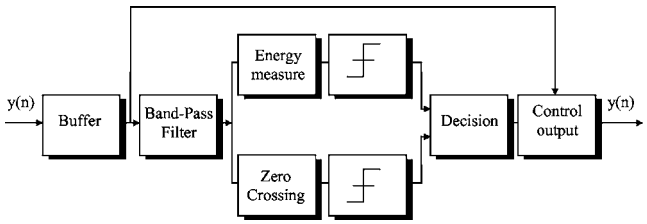


Fig. 2. Breathing noise elimination system.

occur below 1 kHz and unvoiced sounds occur in the 1–3 kHz band. This implies that a method used to detect unvoiced sound intervals may be implemented to recognize breathing noise. Measurement of zero crossings in a frame of speech signal is a common method of unvoiced sound recognition,[6] and is, therefore, a potential candidate for detecting breathing noise. Another factor is that the air velocity in the regulator generates high pressure sound levels inside the mask cavity, hence the energy magnitude of the breathing noise is relatively high compared to that of the speech signal. This suggests that energy measurements may also be applied to detect breathing noise.

To discriminate breathing noise intervals from speech signals, energy or zero-crossing measurements alone do not provide an adequate solution, as shown in the results below. If only energy measurements are made, voiced sounds may be identified as breathing noise. If only zero-crossing measurements are made, unvoiced sounds may also be identified as breathing noise. Figure 2 shows a combination of both methods to form a noise elimination system. The input comprises recorded speech signals, including breathing noise and bubble noise. Since the breathing noise occupies a broadband, typically 1.5–4 kHz, frames of the input signal are first fed to a 20-tap finite impulse response (FIR) 1–4 kHz band-pass filter. Voiced sounds are therefore attenuated and detection errors introduced by them are minimized.

The short-time energy, E, and zero-crossing rate, ZCR, are computed by splitting the speech signal into 22.5 ms frames and applying the following equations:

$$E(k) = \sum_{n=kN}^{(k+1)N-1} y(n)^2, \tag{1}$$

$$ZCR(k) = \sum_{n=kN+1}^{(k+1)N-1} |\text{sign}[y(n)] - \text{sign}[y(n-1)]|, \tag{2}$$

where $y(n)$ is the input signal, $\text{sign}[y(n)]=1$ for $y(n)>0$, $\text{sign}[y(n)]=-1$ for $y(n)<0$, N is the number of samples (180) in a frame, and k is the number of frames (maximum of 512). The decision block in Fig. 2 combines these results.

4. Results

Figure 3(a) shows the energy in about 510 frames of data comprising speech with breathing noise and bubble noise present, while Fig. 3(b) shows the corresponding zero-crossing rate. The data corresponds to the time axis shown in Fig. 4, e.g., the first breathing noise interval is centered on about 50 frames or 1.125 s. It is clear that while the energy of the breathing noise is high compared to the speech between the noise intervals, the zero-crossing rates can be at least as high for both speech and breathing noise. Thus, when both the energy and zero-crossing values for each speech frame are simultaneously above or equivalent to predetermined threshold levels, the frame is assumed to be breathing noise and its value is set to zero (along with several following input signal frames).

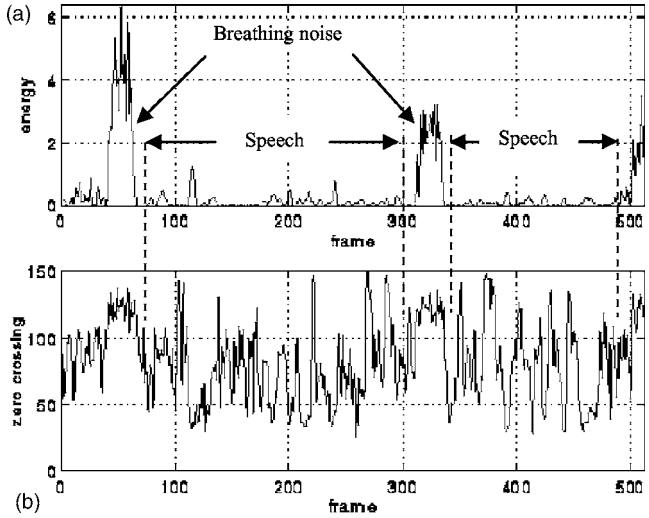


Fig. 3. Normalized speech signal, showing (a) energy, (b) zero-crossing rates, for a series of speech frames.

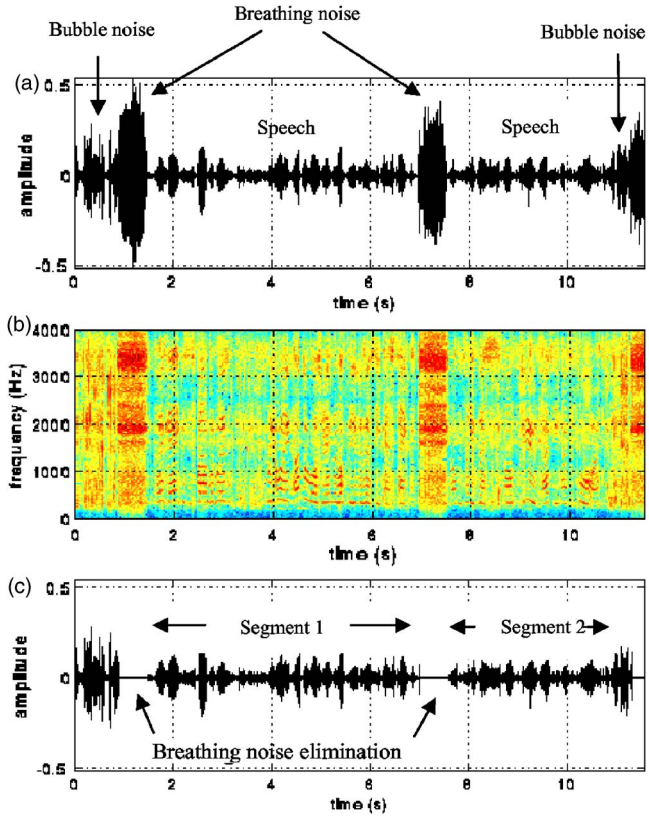


Fig. 4. (Color online) Speech signal (a) in time domain, (b) in frequency domain, (c) in time domain after breathing noise elimination. (b) is intended for color viewing. The information may not be properly conveyed in a black and white printout.

Table 1. Estimation of thresholds for energy and zero-crossing measurements (Ref. 3).

	Measurements				Threshold	
Mask type	mean *E*	std *E*	mean *ZC*	std *ZC*	*E*	*ZC*
Wet Mask	0.32	0.13	103.92	8.57	0.16	93
Aga	0.15	0.09	90.82	6.64	0.03	82
Exo-26	1.04	0.69	97.83	5.40	0.16	91
Aqua Lung	2.26	0.85	128.73	10.61	1.17	115

To illustrate the breathing noise elimination method, the speech signal in Fig. 4(a) was processed. A spectrogram of this signal is shown in Fig. 4(b). Choosing appropriate threshold levels is an important factor for successful noise cancellation. If they are incorrectly defined, unvoiced or voiced sounds are identified as breathing noise.

To estimate threshold levels for energy and zero-crossing measurements, a list of words was read out underwater to evaluate the response of each mask and statistical analysis was then applied.[3] The threshold values for all the masks were defined for 90% acceptance, leading to approximate threshold values of 1.2 and 90, respectively, derived from the mean and standard deviation (std) values of these intervals, as shown in Table 1.

Decisions on these thresholds for the energy and zero-crossing data led to successful identification of breathing noise intervals. When this method was implemented, breathing noise was eliminated, as shown in Fig. 4(c), leaving only the speech and bubble noise. The algorithm's effectiveness was determined subjectively by listening to the processed speech. The bubble noise was completely eliminated so there were gaps in the speech previously occupied by the noise. This affected the speech "quality" by making it sound less natural but the speech "intelligibility" was good in that the words could be identified clearly.

5. Conclusions

A breathing noise elimination method has been presented as a way of improving the subjective speech quality of a diver using a digital voice communication system. Tests with three types of mask, the *wet mask*, *EXO-26* mask, and *Aqua Lung* mask, demonstrated that breathing noise can be eliminated by measuring the energy and zero-crossing rate of the noisy speech signal using carefully chosen amplitude threshold values. Tests with an *Aga* mask were less successful and revealed that the breathing noise magnitude was too low for reliable detection. Further quality enhancement may be achieved by minimizing or eliminating bubble noise, which is the subject of a further study.

Acknowledgment

The authors thank Graseby Dynamics Ltd. for the loan of masks used in diving trials.

References and links

[1]D. J. Meares, "Broadcast-quality speech from diving helmets," J. Audio Eng. Soc. **37**, 927–933 (1989).
[2]B. Woodward and H. Sari, "Digital underwater acoustic voice communications," IEEE J. Ocean. Eng. **21**, 181–192 (1996).
[3]H. Sari, "Underwater acoustic voice communications using digital techniques," Ph.D. thesis, Loughborough University (1997).
[4]C. D. Mathers, and M. D. M. Baird, "A simple means of improving the quality of speech from a diving helmet," Report BBC RD 1989/16, Engineering Division (1989).
[5]B. Woodward, S. Datta, and A. Welsford, "Quality enhancement of a diver's speech signal," *Proceedings of 7th European Conference on Underwater Acoustics*, Delft, The Netherlands, 2004, pp. 979–984.
[6]J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals* (Macmillan, New York, 1993).

Perception of roughness by listeners with sensorineural hearing loss

Jennifer B. Tufts
Department of Communication Sciences, University of Connecticut, 850 Bolton Ave., Unit 1085, Storrs, Connecticut 06269
jennifer.tufts@uconn.edu

Michelle R. Molis
National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans Hospital Road, Portland, Oregon 97207
michelle.molis@va.gov

Abstract: The perception of auditory roughness presumably results from imperfect spectral or temporal resolution. Sensorineural hearing loss, by affecting spectral resolution, may therefore alter roughness perception. In this study, normal-hearing and hearing-impaired listeners estimated the roughness of amplitude-modulated tones varying in carrier frequency, modulation rate, and modulation depth. Their judgments were expected to reflect effects of impaired spectral resolution. Instead, their judgments were similar, in most respects, to those of normally-hearing listeners, except at very slow modulation rates. Results suggest that mild-to-moderate sensorineural hearing loss increases the roughness of slowly fluctuating signals.

PACS numbers: 43.66.Sr, 43.66.Jh, 43.66.Lj [QJF]
Date Received: November 22, 2006 **Date Accepted:** January 15, 2007

1. Introduction

Auditory roughness is an aspect of timbre associated with repeated and rapid fluctuations in amplitude or frequency, and is often described with reference to such qualities as harshness, raspiness, and hoarseness. It is used as a metric for evaluating sound quality in industrial applications, such as automobile manufacturing (e.g., Gonzalez *et al.*, 2003), and in the clinical evaluation of voice quality (e.g., Webb *et al.*, 2004).

Roughness is important in musical contexts as well. Judgments of the perceived tension of chords show a significant positive correlation with the chords' estimated roughness values (Bigand *et al.*, 1996). This result is important because the interplay of musical tension and relaxation is a critical element of the structure of tonal music (Lerdahl and Jackendoff, 1983). Even in nontonal music, in which tension and relaxation are not conveyed by traditional relationships among chords, Pressnitzer *et al.* (2000) found that judgments of the roughness of chords were strongly correlated with judgments of tension. Thus, in both tonal and nontonal music, roughness plays a role in creating the tension so important for musical expression.

The roughness of amplitude-modulated tones is assumed to result from the inability of the auditory system to resolve the components of the stimulus (i.e., the carrier and two sidebands) either spectrally or temporally (Zwicker and Fastl, 1990). Based on earlier work by Terhardt and others, Zwicker and Fastl (1990) proposed that roughness is limited by frequency resolution at low carrier frequencies and by temporal resolution at higher carrier frequencies. For carriers below 2000 Hz, the modulation rate at which peak roughness occurs, the height of the peak, and the modulation rate at which roughness vanishes, all grow with increasing carrier frequency. This is presumably due to the widening of auditory channels as the carrier frequency increases, allowing for interaction among more widely-spaced stimulus components. At carrier frequencies at and above 2000 Hz, Zwicker and Fastl (1990) reported that roughness reaches a peak at modulation rates of approximately 70–80 Hz and vanishes at approximately 250 Hz, independent of carrier frequency, suggesting that temporal resolution is the limiting factor.

Sensorineural hearing loss (SNHL) is often accompanied by impaired frequency and/or temporal resolution. Impaired frequency resolution may mean that the perception of roughness at lower carrier frequencies will be altered for listeners with SNHL, such that stimulus components that would normally be resolved for a normal-hearing (NH) listener would continue to interact and create sensations of roughness for the hearing-impaired (HI) listener. For higher carrier frequencies, where frequency resolution is less important, the perception of roughness by HI listeners may be relatively unaffected.

A recent study addressed this question indirectly. Tufts *et al.* (2005) asked NH and HI listeners to judge the dissonance of musical intervals composed of two harmonic complexes geometrically centered at 500 Hz. Although dissonance and roughness are not synonymous, they are closely related (Terhardt, 1974), with roughness being applicable to both musical and nonmusical sounds. The pattern of the dissonance judgments made by the HI subjects suggested that they did not distinguish differences in dissonance among chords as clearly as the NH listeners did. This result may be partially explained by the wider auditory filter bandwidths measured for the HI subjects at 500 Hz (though not at 2000 Hz), allowing greater interaction among components that are resolved in NH listeners. Alternatively, reduced pitch strength, which often accompanies SNHL (Leek and Summers, 2001), may have affected the judgments of HI listeners by lessening the degree of fusion of tone combinations considered highly consonant by the NH listeners, thereby leading to reduced contrast between consonant and dissonant intervals.

If the perception of roughness is altered by SNHL, such a finding may have implications for the perception of tension in music, as well as timbre and sound quality in musical and nonmusical sounds. The present study examined perceived roughness for SAM tones in listeners with SNHL. Four carrier frequencies were chosen, spanning the range from 250 to 3000 Hz. SAM tones were presented at various modulation rates at two modulation depths. NH subjects were tested at two levels (an equal SPL condition and an equal SL condition) for comparison with HI subjects. It was expected that any differences in roughness perception between NH and HI listeners would be observed primarily for carrier frequencies below 2000 Hz.

2. Method

Twelve subjects participated. Six subjects (1 M, 5 F; mean age=31 years; SD=11.8) had normal hearing in the test ear (i.e., air-conduction thresholds $\leqslant$20 dB HL from 0.25 to 4 kHz; re: ANSI, 1996). The other six subjects (4 M, 2 F; mean age=73.7 years, SD=4.8) had bilateral hearing losses, with a mild to moderate sensorineural hearing loss in the test ear (i.e., air-conduction thresholds between 30 and 60 dB HL from 0.25 to 3 kHz, air-bone gaps of $\leqslant$10 dB from 0.5 to 4 kHz, and a normal tympanogram). None of the subjects had training in music theory or ear training, and none had perfect pitch. All testing took place in a double-walled sound-treated booth. Participation time was approximately two to three hours per subject, spread over one or two sessions with rest breaks allowed. All participants provided written informed consent prior to beginning the study. HI participants were paid for their participation.

Stimuli consisted of SAM tones with carrier frequencies of 250, 500, 1000, and 3000 Hz, and modulation depths of 50% and 100%. For each carrier, eleven modulation frequencies were chosen, spanning the range from sensations of fluctuation through maximum roughness to smoothness, as ascertained through informal listening by NH individuals. The lowest modulation rate was 3 Hz for all carriers. Table 1 lists the modulation rates for each carrier.

The SAM tones were 1000 ms in total duration, including 50 ms raised-cosine onset and offset ramps. All stimuli were generated digitally and played through a 24-bit D/A converter (TDT RP2) at a rate of 40 000 samples per second. They were then passed through an attenuator (TDT PA4) and a headphone buffer (TDT HB6) to one channel of a set of calibrated circumaural earphones (Sennheiser, HD540). Presentation level was 95 dB SPL for the HI subjects, to ensure audibility without causing discomfort. For the NH subjects, the SAM tones were presented at 45 and 95 dB SPL in separate blocks. The lower level was chosen to provide an approximately equal sensation level (SL) to that experienced by the HI subjects, while the upper level was chosen to provide an equal SPL for all subjects.

Table 1. Modulation rates for each carrier frequency.

Carrier frequency (Hz)	Modulation rates (Hz)										
250	3	15	30	45	60	75	90	105	120	135	150
500	3	25	50	75	100	125	150	195	200	225	250
1000	3	25	50	75	100	125	150	195	200	225	250
3000	3	30	60	90	120	150	180	210	240	270	300

A magnitude estimation (ME) technique with standard stimulus was employed (Stevens, 1975). The standard stimulus, a 500 Hz sinusoid 100% amplitude-modulated at 25 Hz, was arbitrarily assigned a value of 100. On each trial, the subject heard the standard, followed by a 500-msec silence, and then the stimulus to be judged. The subject assigned a number to each stimulus that best matched its perceived roughness, using the standard as a reference.

All combinations of carrier frequency, modulation rate, and modulation depth were presented in quasirandomized order, with four replications of each stimulus (i.e., 4 carrier frequencies $\times$ 11 modulation rates $\times$ 2 modulation depths $\times$ 4 replications=352 SAM tones). Stimuli were blocked by presentation level for the NH listeners. The total number of stimuli was 352 for the HI subjects and 704 for the NH subjects (352 stimuli $\times$ 2 presentation levels). Prior to data collection, all subjects successfully completed an ME task in which they judged the relative length of lines (Stevens, 1975), to ensure that they could give appropriate estimates in an ME task. Before testing began, subjects completed 24 practice trials using randomly chosen SAM tones.

3. Results

Geometric means of the roughness judgments were calculated for each stimulus within each group (NH and HI) and condition (modulation depth $\times$ presentation level). Figures 1 and 2 show mean roughness judgments as a function of modulation rate for the NH and HI groups, respectively. Data were fit by separate lognormal functions for each carrier within each condition for the NH listeners and for the 50%-AM stimuli for the HI listeners. R^2 values for these fits fell between 0.85 and 0.99. In Fig. 2(B), the curve-fit for the 500 Hz carrier misses the data point at the modulation rate of 3 Hz. However, the overall fit was quite good ($R^2=0.91$). This R^2 was comparable to the R^2 values for the other fits in this figure. Lognormal functions did not provide acceptable fits to the HI data for the 100%-AM stimuli, primarily due to the relatively high roughness estimates of the 3-Hz-modulated stimuli for all four carriers. These data were fit by logistic functions, with R^2 values between 0.85 and 0.98. Roughness judgments in both listener groups showed the usual bandpass characteristic, with judgments for each carrier rising to a maximum value and then declining as the modulation rate increased. The single exception occurred for 50%-AM 3000 Hz tones at 45 dB SPL, for which the NH subjects' estimates changed very little as modulation rate increased from 60 Hz.

In each condition for both subject groups, greater roughness was perceived over a wider range of modulation rates for higher carriers. This trend was examined via separate Friedman tests for each condition within each group, with overall roughness operationally defined as the sum of the mean roughness judgments for a single carrier in a particular condition. All six tests were significant ($p<0.05$), indicating that overall roughness differed among at least two of the carriers in each condition. Five of the six analyses rank-ordered roughness according to increasing carrier. The only exception occurred for the HI group listening to 100%-AM tones. In that condition, the 500 Hz carrier was ranked as having least overall roughness, followed by the 250 Hz carrier, and then the 1000 and 3000 Hz carriers, which were ranked equally rough.

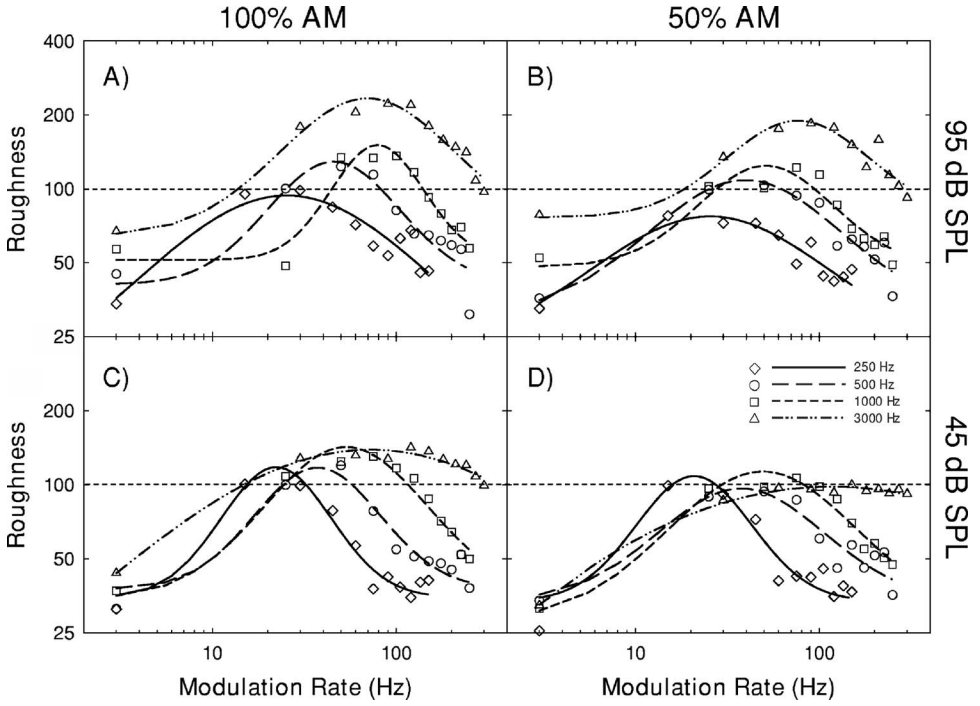


Fig. 1. Geometric means of the roughness judgments of the normal-hearing listeners (N=6) as a function of modulation rate for 100%-modulated and 50%-modulated pure tones presented at 95 dB SPL (panels A and B) or 45 dB SPL (panels C and D). The parameter is carrier frequency.

The modulation rate at which each fitted curve reached its maximum value (i.e., peak roughness) is plotted in Fig. 3. For comparison, data points adapted from Zwicker and Fastl (1990), showing the modulation rates at which peak roughness occurred for SAM tones, and from Miskiewicz *et al.* (2006) showing the beat rates at which peak roughness occurred for two simultaneous tones, are shown. The peak roughness estimates of the NH group were generally consistent with these earlier data. For most conditions, the modulation rate at peak roughness

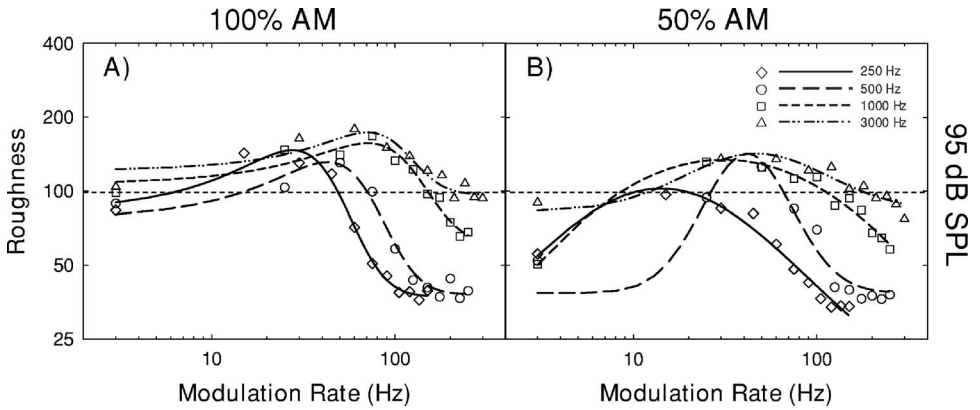


Fig. 2. Geometric means of the roughness judgments of the hearing-impaired listeners (N=6) as a function of modulation rate for 100%-modulated (panel A) and 50%-modulated (panel B) pure tones presented at 95 dB SPL. The parameter is carrier frequency.

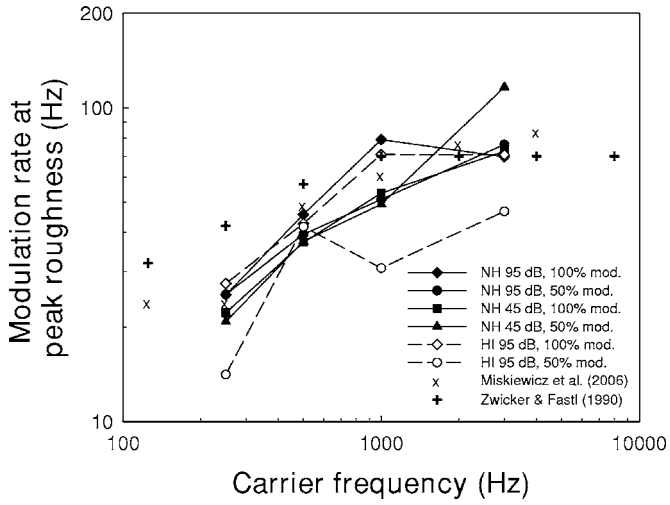


Fig. 3. Modulation rates at which lognormal functions fitted to subject data reached their maximum values (i.e., peak roughness) as a function of carrier frequency. Each line in the figure represents a single stimulus condition for either the normal-hearing or hearing-impaired listeners. Data points adapted from Zwicker and Fastl (1990), showing the modulation rates at which peak roughness occurred for SAM tones, and from Miskiewicz *et al.* (2006), showing the beat rates at which peak roughness occurred for two simultaneous tones, are also plotted.

increased with carrier, though at a decelerating rate. For the carrier frequency of 3000 Hz, however, peak roughness occurred at a much higher modulation rate for the 50%-AM tones presented at 45 dB SPL compared with the other three stimulus conditions. Although the value of 116 Hz represents the peak in the fitted function, it may not truly reflect peak roughness in the subjects' perception, since judgments in this condition changed very little for modulation rates greater than 60 Hz. Relative to the NH listeners, the HI subjects perceived peak roughness at similar modulation rates for the 100%-AM stimuli, but at somewhat lower modulation rates for the 50%-AM stimuli (except at 500 Hz).

The effect of modulation depth within each listener group was tested via Wilcoxon signed-rank tests on overall roughness collapsed across carrier frequency. Tests were significant for the NH group at both the high ($z=-2.2$, $p<0.05$) and low ($z=-2.2$, $p<0.05$) presentation levels, and for the HI group at the single (high) presentation level ($z=-2.2$, $p<0.05$), indicating that greater roughness was perceived for 100%-AM tones than for 50%-AM tones. The effect of presentation level (NH subjects only) was assessed with a Wilcoxon signed-rank test on overall roughness collapsed across carrier and modulation depth. Roughness estimates were significantly higher for the high presentation level ($z=-1.99$, $p<0.05$) compared with the low presentation level. Differences in overall roughness estimates between the NH and HI groups at equal SPL and approximately equal SL were assessed via Mann–Whitney U tests, with correction for multiple tests. None of the comparisons was significant, although the comparison across groups listening at equal SPL approached significance for the 100%-AM stimuli ($z=-2.08$, $p=0.07$).

4. Discussion and conclusions

NH listeners' roughness judgments were consistent with previous research on the roughness of SAM and beating tones. Briefly, roughness nearly always showed a bandpass characteristic as a function of modulation rate; modulation rate at peak roughness tended to increase with carrier frequency; overall roughness was greater for higher carrier frequencies; 100%-AM stimuli were judged rougher than 50%-AM stimuli; and signals presented at 95 dB SPL were perceived as rougher than those at 45 dB SPL. The latter result was driven mainly by responses to the 3000 Hz tones. At the high presentation level, several NH subjects commented on the aversive,

shrill quality of these stimuli. This may have led to the disproportionately high roughness estimates for these tones, despite instructions to the subjects to ignore all other stimulus variables.

The roughness judgments of the HI group were generally similar to those of the NH group with respect to the relative relationships among the stimulus parameters, regardless of whether the stimuli were presented at equal SPL or at approximately equal SL. Two main differences between the groups were noted. First, 100%-AM stimuli modulated at 3 Hz received higher roughness estimates from the HI subjects, necessitating the fitting of these data with logistic instead of lognormal functions. Second, the HI group showed a trend toward perceiving peak roughness at somewhat lower modulation rates compared with the NH group for the 50%-AM stimuli. The wider auditory channels that usually accompany SNHL presumably allow interactions to occur among stimulus components over a wider frequency range. However, because these differences occurred for stimulus bandwidths well within the auditory filter bandwidth for both subject groups, and for carriers above and below 2000 Hz, they appear not to be directly related to auditory filter bandwidth per se. Furthermore, disregarding the previously discussed 3000 Hz stimuli, NH subjects' judgments at the high and low presentation levels were generally similar, even though their auditory filter bandwidths would presumably be broader—and thus more similar to the auditory filters of the HI subjects—at the high presentation level. No direct relationship exists between degree of hearing loss and frequency selectivity. Since auditory filter bandwidths were not measured, it is, therefore, not possible to determine the extent to which the frequency selectivity of the HI subjects was impaired. However, based on their absolute thresholds, their filter bandwidths in the frequency regions of interest are estimated to have been two to three times broader than normal (Glasberg and Moore, 1986). Recently, the relationship of auditory filter bandwidth to roughness perception has been questioned. Miskiewicz *et al.* (2006), in a study of the perceived roughness of beating tones, concluded that neither the peak in roughness nor the point at which the sensation of roughness disappeared were related to auditory filter bandwidth. Their results, together with those of the current study, indicate there likely is not a strong relationship between auditory filter bandwidth and roughness perception.

Loss of cochlear compression due to outer hair cell damage may have affected the roughness judgments of the HI listeners for the 100%-AM stimuli modulated at 3 Hz. Greater portions of the valleys in the waveforms may have been inaudible to the HI subjects compared with the NH subjects. The longer "silent" gaps, combined with the presumably normal loudness of the waveform peaks due to recruitment, may have distorted the signal, producing a rougher, "staccato" sensation.

The present findings suggest one explanation for the results of Tufts *et al.* (2005), in which HI listeners did not distinguish consonant and dissonant tone combinations as clearly as NH subjects did, and even rated a 2000-Hz pure tone as much more dissonant than NH subjects did. In that study, some of the more consonant ("smooth"-sounding) tone combinations still produced slow beats. The HI subjects in that study, like those in the present study, may have perceived the slow amplitude fluctuations as sounding rougher than did the NH subjects. It is possible that a combination of factors resulting from cochlear damage, including reduced pitch strength, loss of cochlear compression and loudness recruitment, may have caused increased sensations of roughness for these signals. With regard to music listening then, signals that NH listeners perceive as stable or slowly fluctuating may in fact sound rougher to listeners with mild-to-moderate SNHL.

Acknowledgments

This research was supported in part by NIH Grant DC 00626, "Hearing Loss and the Perception of Complex Sounds." The authors wish to thank Marjorie Leek and two anonymous reviewers for their comments on an earlier version of this manuscript.

References and links

American National Standards Institute (**1996**). "American National Standard: Specifications for audiometers," ANSI S3.6-1996 (New York).

Bigand, E., Parncutt, R., and Lerdahl, F. (**1996**). "Perception of musical tension in short chord sequences: The

influence of harmonic function, sensory dissonance, horizontal motion, and musical training," Percept. Psychophys. **58**, 125–141.
Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.
Gonzalez, A., Ferrer, M., de Diego, M., Pinero, G., and Garcia-Bonito, J. (**2003**). "Sound quality of low-frequency and car engine noises after active noise control," J. Sound Vib. **265**, 663–679.
Lerdahl, F., and Jackendoff, R. (**1983**). *A generative theory of tonal music* (MIT Press, Cambridge, MA).
Leek, M. R., and Summers, V. (**2001**). "Pitch strength and pitch dominance of iterated rippled noise in hearing-impaired listeners," J. Acoust. Soc. Am. **109**, 2944–2954.
Miskiewicz, A., Rakowski, A., and Rosciszewska, T. (**2006**). "Perceived roughness of two simultaneous pure tones," Acta. Acust. Acust. **92**, 331–336.
Pressnitzer, D., McAdams, S., Winsberg, S., and Fineberg, J. (**2000**). "Perception of musical tension for nontonal orchestral timbres and its relation to psychoacoustic roughness," Percept. Psychophys. **62**, 66–80.
Stevens, S. S. (**1975**). *Psychophysics: Introduction to its perceptual, neural, and social prospects*, edited by G. Stevens (Wiley, New York), pp. 26–31.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Tufts, J., Molis, M., and Leek, M. (**2005**). "Dissonance perception in people with normal hearing and sensorineural hearing loss," J. Acoust. Soc. Am. **118**, 955–967.
Webb, A., Carding, P., Deary, I., MacKenzie, K., Steen, N., and Wilson, J. (**2004**). "The reliability of three perceptual evaluation scales for dysphonia," Eur. Arch. Otorhinolaryngol. **261**, 429–434.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer-Verlag, New York), pp. 231–236.

Poisson point process modeling for polyphonic music transcription

Paul Peeling, Chung-fai Li, and Simon Godsill
Signal Processing Group, Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom
php23@eng.cam.ac.uk, cfli@cuhk.edu.hk, sjg@eng.cam.ac.uk

Abstract: Peaks detected in the frequency domain spectrum of a musical chord are modeled as realizations of a nonhomogeneous Poisson point process. When several notes are superimposed to make a chord, the processes for individual notes combine to give another Poisson process, whose likelihood is easily computable. This avoids a data association step linking individual harmonics explicitly with detected peaks in the spectrum. The likelihood function is ideal for Bayesian inference about the unknown note frequencies in a chord. Here, maximum likelihood estimation of fundamental frequencies shows very promising performance on real polyphonic piano music recordings.

PACS numbers: 43.60.Uv, 43.75.Xz, 43.60.Pt [JC]
Date Received: January 10, 2007 **Date Accepted:** February 4, 2007

1. Introduction

Music transcription refers to the generation of a musical score from audio data. The transcription of polyphonic music is of particular interest because of the underlying quasiperiodic structure of the individual note components. To exploit this, models describing the fundamental frequency of a note and its harmonics have been proposed.[1–3] Only a relatively small number of parameters is needed for a plausible description of the frequency content of the music. Often the performance of these schemes has been limited because the harmonics of different notes coincide, thus obscuring some of the components present in the data. Methods which account for this tend to either increase the model complexity[2] or use a heuristic approach to recover the missing components.[4]

Here we focus on determining multiple pitches within short frames of polyphonic music. As in many such systems,[4–7] a preprocessing stage is assumed which extracts, or "detects," individual peaks from a short-time frequency representation of the music. These peaks in the time-frequency domain are then modeled in a novel way as a nonhomogeneous Poisson point process.[8] In such a model, the number of peaks detected in each frame is a Poisson random variable. In this formulation a likelihood function may be directly formulated for the observed data, without resorting to any data association task which assigns individual detected peaks to particular note fundamentals or harmonics.[5,6] Thus we avoid the computational complexity of a full probabilistic data association, and also the heuristic approximations of suboptimal data association schemes.

The paper is organized as follows. In Sec. II we describe the basic Poisson process model for musical note clusters. Section III describes the estimation of rate functions for the model. Section IV describes a basic algorithm for implementation of the approach and Sec. V gives some results of application to real musical extracts. Finally, Sec. VI gives concluding remarks and points toward future developments.

2. Poisson point processes

Suppose that M simultaneous pitches are present in a frame of audio, with fundamental frequencies $F=\{f_1,f_2,\ldots,f_M\}$, and this frame exhibits a number of peaks $P=\{p_1,p_2,\ldots\}$ in the frequency domain. The kth element p_k in P is the frequency in Hz of a single peak in the spec-

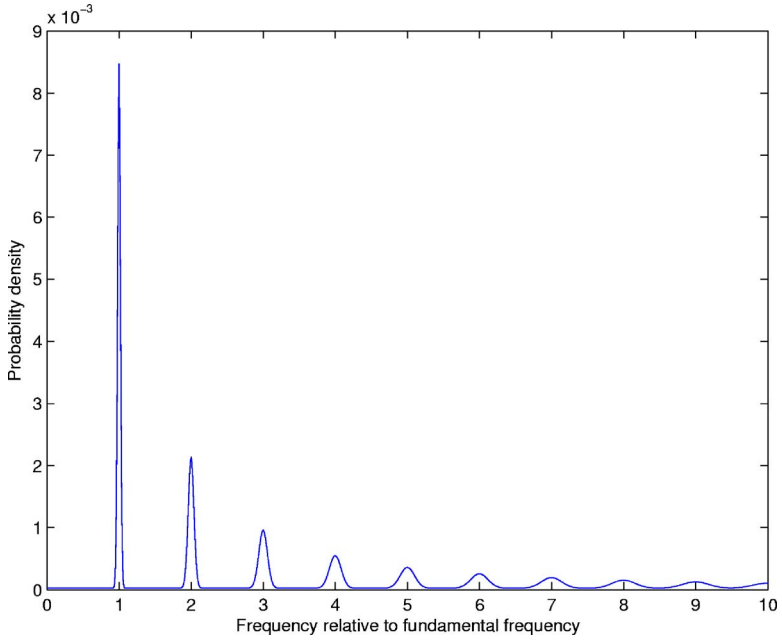


Fig. 1. (Color online) Detection of peaks from raw spectral amplitude data and representation as a point process on the frequency axis.

trum. It is expected that each note f_m will contribute peaks corresponding to its own harmonic structure. Any noise in the signal or inadequacies in the peak detection method may generate additional noise or "clutter" peaks (see Fig. 1).

Our fundamental assumption is that the peaks generated in this fashion are realizations from a *nonhomogeneous Poisson point process*[8] with intensity function $\rho(p|F)$. The number of detected peaks $N_{\mathcal{A}}$ in an interval $\mathcal{A}$ of the frequency domain is a Poisson random variable

$$p(N_{\mathcal{A}} = n) = \frac{e^{-\mu_{\mathcal{A}}}\mu_{\mathcal{A}}^{n}}{n!},$$

where the expected number of peaks in interval $\mathcal{A}$ is

$$\mu_{\mathcal{A}} = \int_{\mathcal{A}} \rho(p|F)dp.$$

All peaks generated by the pitches have been combined into one set—we have technically taken a union of the individual point processes for each pitch. The union process of a number of independent Poisson processes is also a Poisson process, with intensity function given by the sum of the individual intensity functions. Therefore, we can decompose the intensity function into individual note and clutter components, i.e.

$$\rho(p|F) = \rho(p|f_1, f_2, \ldots, f_M) = \rho_C(p) + \sum_{m=1}^{M} \rho(p|f_m), \quad (1)$$

where $\rho_C(p)$ is the predefined intensity function of the clutter process, and $\rho(p|f_m)$ is the intensity function of an individual note f_m.

It should be noted at this stage that this type of model is subtly different from those usually employed in peak modeling, which assume that each harmonic of each note in the spectrum has to be uniquely associated with at most one detected peak,[6,9] which can lead to a com-

binatorial explosion of data association terms when many notes are superimposed. Here, however, each harmonic may generate any number of detected peaks, in accordance with the intensity function $\rho(p|f_m)$. This will lead to substantial simplifications in computation. It also models the fact that individual harmonics may often lead to "split" peaks where several peaks are detected rather than just one. In a tracking setting a similar principle has recently been applied for simplification of the classical data association problem.[10–12]

Peak detections will usually be made over a discrete set of frequency bins. Let $N(k)$ be the number of peaks occurring in the frequency interval $(k\Delta,(k+1)\Delta]$, where Δ is the frequency analysis bin size (easily made variable with frequency in multiscale approaches if required). Then, under the nonhomogeneous Poisson point process assumption, the probability of $N(k)$ peaks occurring is given by

$$P(N(k)=n|f_1,f_2,\ldots,f_M)=\frac{e^{-\mu(k)}\mu(k)^n}{n!},$$

where $\mu(k)$ is defined as the expected number of peaks occurring within the kth bin. Using Eq. (1) we have

$$\mu(k)=\mu_C(k)+\sum_{m=1}^{M}\mu_{f_m}(k), \tag{2}$$

where $\mu_C(k)$ or $\mu_{f_m}(k)$ are defined as the integrals of the intensity functions $\rho_C(p)$ and $\rho(p|f_m)$ within bin k, respectively. We term these components the *rate functions* within bin k since they specify the expected number of peaks contributed by each note in bin k.

We assume that a single detection is made in a particular bin if one or more peaks are present on the continuous frequency scale. We then have the probability of a peak being detected in bin k as

$$P(N(k)\geqslant 1)=1-e^{-\mu(k)}$$

and the probability of no peak being detected is

$$P(N(k)=0)e^{-\mu(k)}.$$

Now, suppose that if a peak is detected at bin k we set observation $y_k=1$, and $y_k=0$ otherwise, $k=0,\ldots,K-1$. Hence for detected peak data $Y=[y_0,y_1,\ldots,y_{K-1}]^T$ we have

$$P(Y|F)=\prod_{k=0}^{K-1}y_k(1-e^{-\mu(k)})+(1-y_k)e^{-\mu(k)}. \tag{3}$$

Having obtained a likelihood function it is now in principle possible to perform inference on the number of notes and their pitches.

3. Rate function estimation

In order to compute the likelihood for a given note combination in Eq. (3) it is necessary to specify a rate function $\mu_f(k)$ for each possible note frequency f and for each frequency bin k. We note from Eq. (2) that the rate functions are expressed in terms of the underlying Poisson intensity functions. These intensity functions can in principle be learned from annotated training data. As an alternative, we here construct the rate functions $\mu_f(k)$ directly, either by learning their form from training data, or by construction from generic modeling principles:

Nonparametric estimation from training data. In this approach, rate functions are estimated from a large database of annotated training data. Peaks in the discrete Fourier transform (DFT) are extracted from each frame of data using a thresholded first-difference operation, with a frequency-dependent threshold determined using a running median filter. Their positions in terms of frequency bins are then histogrammed to determine the rate functions for each note separately.

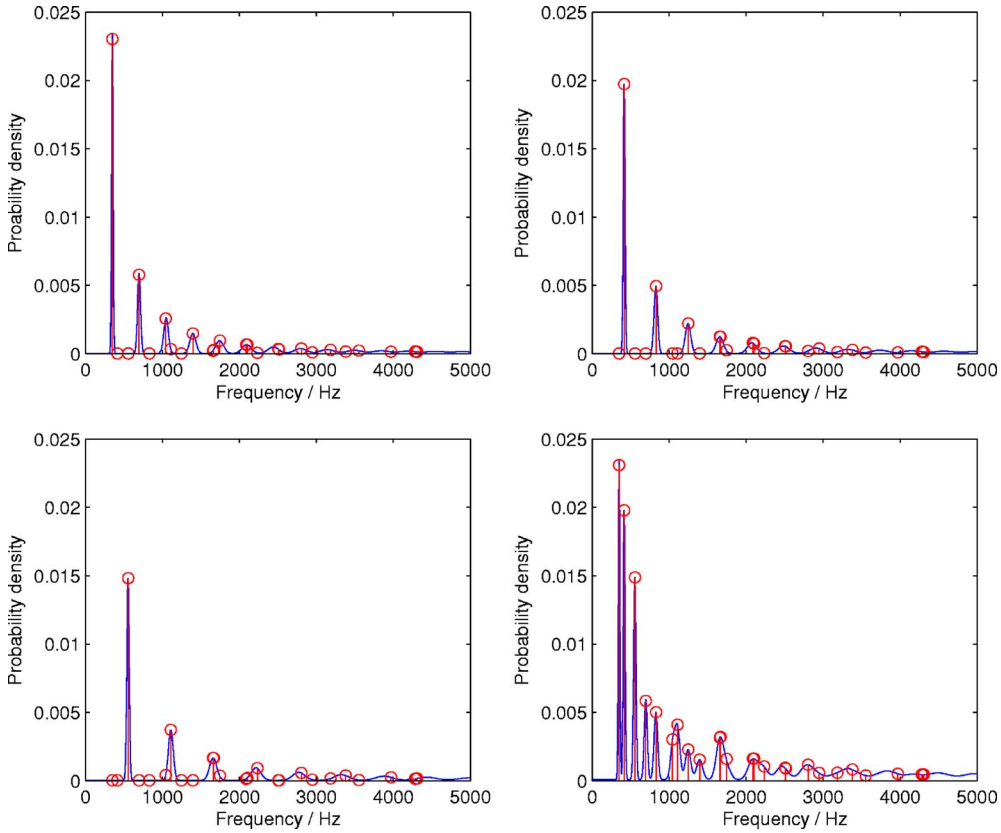


Fig. 2. (Color online) Rate function from a generic parametrized model.

Generic model. We may expect this to generalize better to a range of instruments. In this approach a Gaussian mixture model is proposed for the rate function. The mixture components are expected to be centered close to the frequency of the harmonic number h, i.e., at a frequency h_f. The general form of the rate function model for a note of fundamental frequency f is then as follows:

$$\mu_f(k) = \sum_{h=1}^{H} \frac{\beta_{f,h}}{\sqrt{2\pi\sigma_{f,h}^2}} \exp\left[-\frac{(f_k - hf)^2}{2\sigma_{f,h}^2} \right],$$

where f_k is the center frequency of bin k, $\sigma_{f,h}^2$ is the variance of that component's frequency, and $\beta_{f,h}$ are positive mixture weights (which need not sum to unity).

The variance and mixture weight components are constrained in a particular way such that $\beta_{f,h}=Ae^{-\beta h}$ and $\sigma_{f,h}^2=\kappa^2 h^2$, where A, κ, and β are parameters to be specified or fitted to the peak data. See Fig. 2 for a realization of the rate function of a note using this model.

An alternative approach investigated was to estimate the parameters of the model from labeled training data. However, we found in practice that the performance of such a scheme was generally poorer than the model suggested above.

The clutter intensity ρ_C is modeled as uniform over the frequency range of the peak detector.

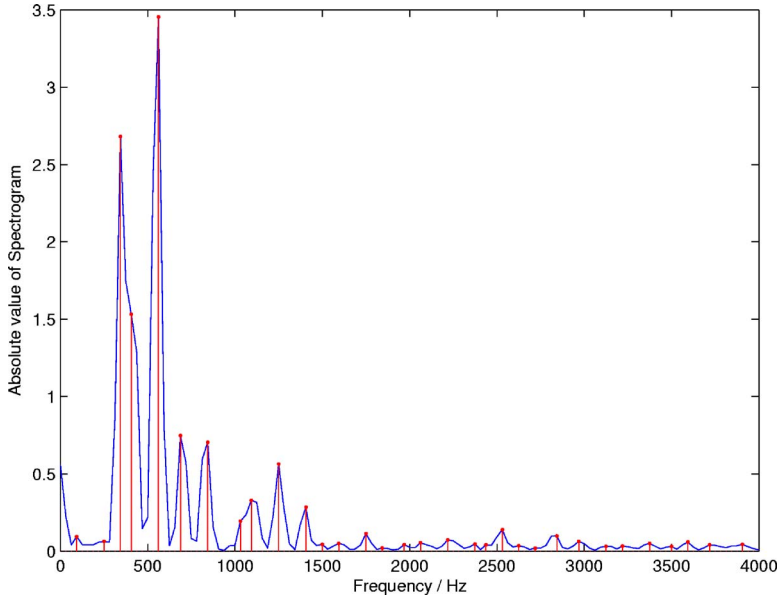


Fig. 3. (Color online) Fitting of a Poisson point process model (solid line) to peak position data (circles). The peaks correspond to three note pitches. Individual notes fitted shown in top left, top right, and bottom left panels, and the mixture of three notes fitted is shown in the lower-right panel.

4. Maximum-likelihood transcription

For this proof-of-principle implementation we perform a frame by frame maximum likelihood (ML) transcription of audio extracts. Prior to analysis, peaks are extracted as for the training data in the nonparametric approach in Sec. III. In addition a second differencing technique was found to detect some peaks that are otherwise obscured by nearby peaks with larger magnitude, detecting, for example, the peak at 400 Hz in Fig. 1 close to the peak at 350 Hz.

An exhaustive search of all the possible note combinations to find the ML solution is computationally infeasible for long extracts. Instead we iterate to find a local maximum. An effective approach was a greedy search algorithm that added at each iteration the note with greatest increase in likelihood. We verify this greedy search with a sampling procedure that takes a subset Q of m notes from the set $K=\{f_1,f_2,\ldots,f_M\}$ of notes found, and checks that the ML solution of $(m+1)$ notes given Q is still a subset of K. We found that the greedy solution consistently passes this verification, and suggest this is due to the robust behavior of the Poisson likelihood function (3) over the search space, which renders each note in the mixture reasonably independent of the others. This suggestion requires further verification in more detailed studies. See Fig. 3 for an illustration.

5. Results

We demonstrate the two models on polyphonic, classical piano music, with up to four notes playing simultaneously. Frames were grouped into single "chord" entities using a time-frequency based segmentation procedure,[13] with some manual intervention to correct for gross errors; and peaks from these grouped frames were analyzed together.

For cases where the number of simultaneous note pitches M is unknown, we estimate M by the Akaike information criterion (AIC).[14] The AIC criterion is calculated as follows:

$$\mathrm{AIC} = 2M' - 2\ln p(Y|\widehat{\mu_{M'}}),$$

where $\widehat{\mu_{M'}}$ is the rate function corresponding to maximum likelihood estimate of M' simultaneous notes. We then choose M to be the value of M' for which AIC is a minimum.

Table 1. Performance of models on a set of piano music extracts. For the generic model, we have chosen the following set of parameters: $H=10$, $A=1$, $\kappa=1$, $\beta=0$.

Extract	Recording	M	Trained	Generic
Creation	Steinway	2	100%	100%
		3	90%	84%
		4	85%	78%
Moonlight	Elena Kuschnerova[a]	3–4		78%
	MIDI Synthesized[b]	3–4		91%
Variations	Andrew Koay[c]	4		88%

[a]Available from www.elenakuschnerova.com
[b]Recorded using Winamp (www.winamp.com)
[c]Available from http://music.download.com/

The performance metric is $(N-M-E)/N$ where N is the correct number of notes from the ground truth, M is the number of notes missed from the ground truth, and E is the number of error notes not present in the ground truth.

Table 1 presents our results on the extracts (see Fig. 4) tested. Figure 5 demonstrates a transcription of the "Moonlight" extract. Results for all methods and extracts are very promising. The nonparametric trained model is observed to perform better than the generic model, but the training method used is not practical for many music transcription applications.

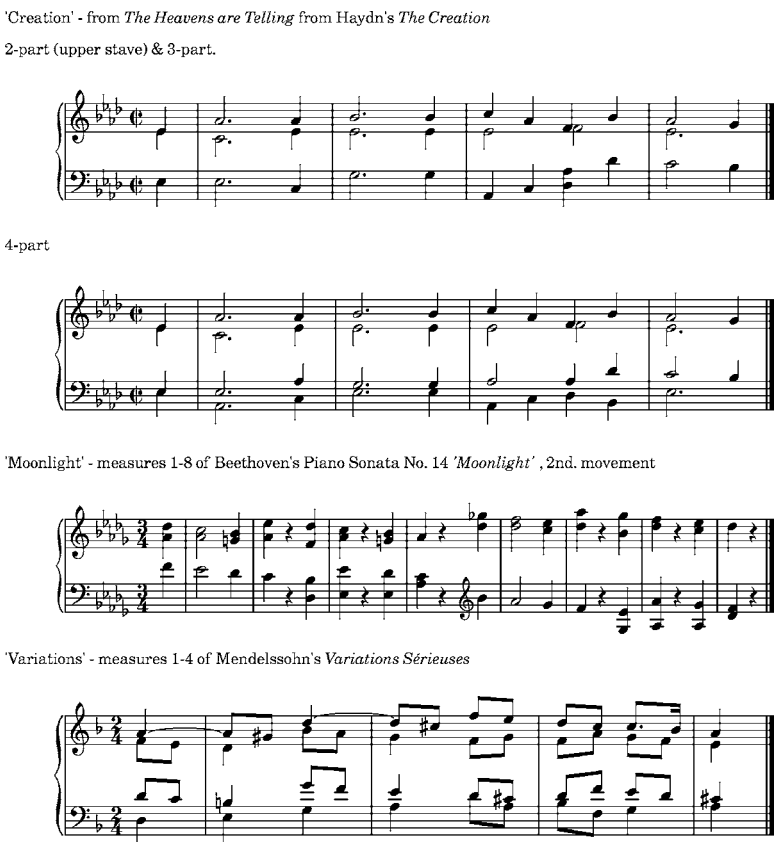


Fig. 4. Scores of extracts.

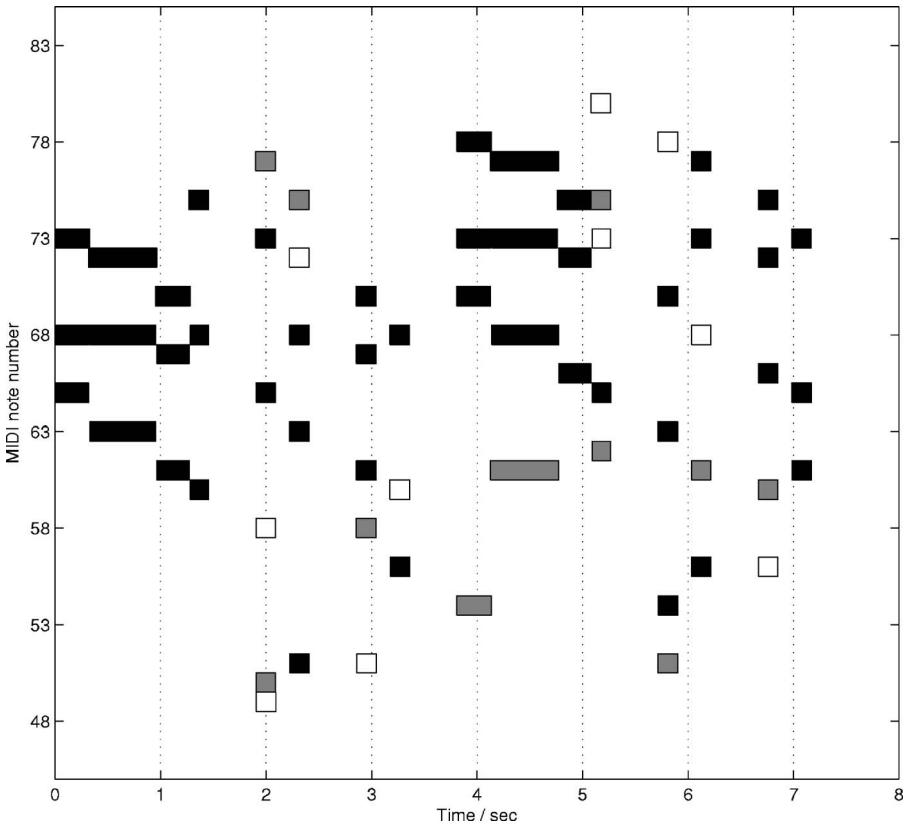


Fig. 5. Transcription of Moonlight using generic model and AIC: Black notes are correctly transcribed, white shows notes compared with the ground truth, and gray shows additional notes not present in the ground truth.

6. Conclusion and discussion

This paper has introduced Poisson point processes as a model for fundamental frequency estimation in polyphonic music. The principal advantage of such an approach is the simplicity of the resulting likelihood function, allowing many notes to be superimposed in a straightforward manner, and without performing any explicit data association task to link detected peaks with particular note harmonics. Several possible forms for the rate function were considered and example transcription results were given for polyphonic piano music.

We anticipate that performance gains can be achieved by embedding the Poisson model in a hierarchical model that links multiple frames together, thus directly modeling the evolution of pitches with time.[15] A useful starting point is a hidden Markov model for pitch transitions over time, with a Poisson observation model for individual frames.

A further unexplored area is a model for the DFT amplitude process, which could guide the transcription process to better results and also lead to inference of additional quantities such as note timbre, playing volume, or instrument identity. Here one can consider extending the point process to a *marked point process*,[8] in which both amplitudes and frequencies of peaks are modeled. Initial investigations have shown that this is a promising approach.

References and links

[1]M. Davy and S. Godsill, "Bayesian harmonic models for musical pitch estimation and analysis," Technical Report No. CUED/F-INFENG/TR 431, Engineering Department, University of Cambridge, UK (2002).
[2]M. Davy and S. Godsill, "Bayesian harmonic models for musical signal analysis," in *Bayesian Statistics VII*, edited by J. Bernardo (Oxford University Press, Oxford) (2003).
[3]A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," IEEE Trans. Speech

Audio Process. **14**, 679–694 (2006).
[4]A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," IEEE Trans. Speech Audio Process. **6**, 804–816 (2003).
[5]H. Thornburg, R. Leistikow, and J. Berger, "Melody extraction and musical onset detection from framewise STFT data," IEEE Trans. Speech Audio Process. (2006), accepted for publication.
[6]R. Maher, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," J. Acoust. Soc. Am. **95**, 2254–2263 (1994).
[7]J. Bello, L. Daudet, and M. Sandler, "Automatic piano transcription using frequency and time-domain information," IEEE Trans. Speech Audio Process. **14**, 2242–2251 (2006).
[8]D. R. Cox and V. Isham, *Point Processes* (Chapman and Hall, London, 1980).
[9]R. J. Leistikow, H. Thornburg, J. Smith III, and J. Berger, "Bayesian identification of closely-spaced chords from single-frame STFT peaks," in *Proceedings of the 7th International Conference on Digital Audio Effects*, Naples, Italy (2004).
[10]K. Gilholm and D. Salmond, "Extended object and group tracking," in *RTO SET Symposium on Target Tracking and Sensor Data Fusion for Military Observation Systems* (2003).
[11]K. Gilholm and D. Salmond, "A spatial distribution model for tracking extended objects," IEE Proc., Radar Sonar Navig. **152**, 364–371 (2005).
[12]K. Gilholm, S. Godsill, S. Maskell, and D. Salmond, "Poisson models for extended target and group tracking," in *SPIE Conference 2005: Signal and Data Processing of Small Targets* (2005).
[13]H. Nagano, K. Kashino, and H. Murase, "A fast search algorithm for background music signals based on the search for numerous small signal components," *Proc. of the 2003 International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, Vol. 5, 796–799 (2003).
[14]H. Akaike, "A new look at the statistical model identification," IEEE Trans. Comput.-Aided Des. **19**, 716–723 (1974).
[15]S. Godsill, "Computational modeling of musical signals," CHANCE Magazine **17** (2004).

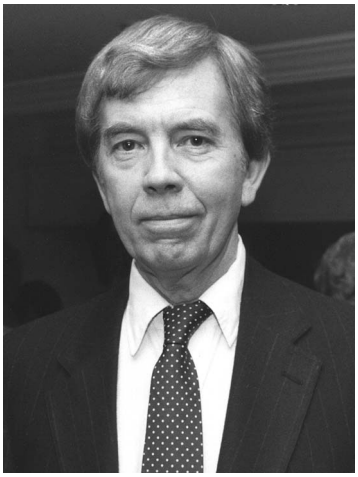
ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America

Damian J. Doria
For contributions to the acoustical design of performing arts spaces.



John M. Eargle
For contributions to electroacoustics and the recording arts.

Peter Rona
For application of acoustics to geophysical imaging.

The fourth joint meeting of the Acoustical Society of America and the Acoustical Society of Japan held in Honolulu, Hawaii

The fourth joint meeting of the Acoustical Society of America (ASA) and the Acoustical Society of Japan (ASJ) was held 28 November–2 December 2006 at the Sheraton Waikiki Hotel in Honolulu, HI. Some meeting events were also held at the Royal Hawaiian Hotel, which is adjacent to the Sheraton Waikiki. This is the fourth time that the Society has met jointly with the ASJ in this city, the previous meetings being held in 1978, 1988, and 1996.

The meeting drew a total of 1952 registrants, including 339 nonmembers and 455 students. There were 595 registrants from Japan and 1032 registrants from the USA, attesting to the joint nature of the meeting. There were 57 registrants from Canada, 35 from Germany, 33 from the U.K., 31 from Korea, 26 from Australia, 23 from France, 15 from Taiwan, 13 from the Netherlands, 9 each from Denmark, Norway, and Sweden, 8 each from Mexico and Poland, 7 each from Italy and Russia, 4 each from Belgium, Hong Kong, Israel, and Spain, 3 from Austria, 2 each from Brazil, Czech Republic, Ecuador, Singapore, Switzerland, and Turkey, and 1 each from Colombia, Egypt, India, and People's Republic of China.

A total of 1618 papers, organized into 124 sessions, covered the areas of interest of the technical committees and committees on education in acoustics from both sponsoring societies. There were also nine meetings dealing with acoustical standards.

The meeting began on Tuesday, 28 November, with an opening session that included two plenary lectures. Dr. Lawrence Crum, University of Washington, Seattle, presented "Therapeutic Ultrasound," and Professor Masayuki Morimoto, Kobe University, presented "How can auditory presence be generated and controlled?"

The ASA's 13 Technical Committees held open meetings during the Honolulu meeting where plans were made for special sessions at upcoming ASA meetings, topics of interest to the attendees were discussed, and informal socials were held after the end of the official business. These meetings are working, collegial meetings and all people attending Society meetings are encouraged to attend and to participate in the discussions. More information about Technical Committees, including minutes of meetings, can be found on the ASA Website <http://asa.aip.org/committees.html> and in the Acoustical News—USA section of the September issue of the *Journal of the Acoustical Society of America*.

Social events included a social hour held on Wednesday, an icebreaker and a reception for students, and morning coffee breaks. A banquet, which drew about 700 participants, was held on Friday, 1 December, in place of a second social hour and included a seven-course Chinese dinner. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics. The luncheon sponsored by the Committee on Women in Acoustics drew 120 participants. A program was also arranged for the 159 accompanying persons who attended the meeting including a presentation on lei making and a hula demonstration and history. These social events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions.

The awards ceremony was held during the banquet and included acknowledgment of the local organizing committee, members of the technical program organizing committee, and other award announcements. ASA presented certificates to newly elected fellows, technical area awards, and awards to meeting organizers from the ASJ. The ASJ presented awards to joint meeting organizers.

ASA President Anthony Atchley (see Fig. 1) welcomed the participants and then introduced Yôiti Suzuki, President of the Acoustical Society of Japan (see Fig. 2) who made opening remarks. Anthony Atchley introduced

 0001-4966/2007/121(4)/1811/5/$23.00

FIG. 1. Anthony A. Atchley, ASA President and ASA Technical Program Chair.

FIG. 2. Yôiti Suzuki, ASJ President and ASJ Technical Program Chair.

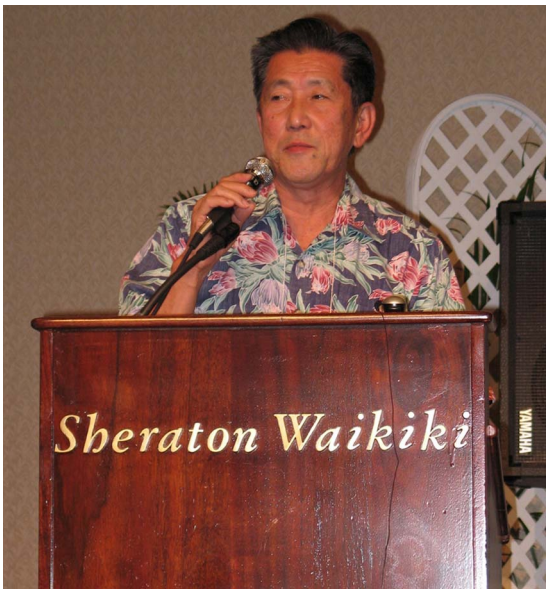


FIG. 3. Whitlow W.L. Au, ASA General Chair of the joint meeting.

FIG. 4. ASA President Anthony Atchley (l) presents Science Writing Award for Professionals in Acoustics to Taras Gorishnyy.

FIG. 5. Aaron Thode, recipient of the 2005 A. B. Wood Medal and Prize of the Institute of Acoustics (U.K.).

Whitlow Au, ASA General Chair of the meeting (see Fig. 3), who acknowledged the contributions of the members of the local committee including Anthony Atchley, ASA Technical Program Chair; Timothy F. Noonan, Audio/Visual; Dorothy E. Au, Accompanying Persons Program; Marc O. Lammers, Signs/Publicity; John S. Allen, Meeting Room Coordinator; David L. Adams, Food; Paul E. Nachtigall, Posters; Todd R. Beiler, Banquet Entertainment; Neal Frazer, Special Affairs; William Friedl, Public Relations; John C. Burgess, Consultant; and Jarleth Badham, Meeting Administrator. He also expressed thanks to the members of the joint ASA/ASJ Technical Program Organizing Committee: Anthony A. Atchley, ASA Technical Program Chair; Yôiti Suzuki, ASJ Technical Program Chair; Dezhang Chu, Acoustical Oceanography; Tomonari Akamastu and Whitlow W.L. Au, Animal Bioacoustics; David L. Adams and Hiroshi Sato, Architectural Acoustics; John S. Allen, Biomedical Ultrasound/Bioresponse to Vibration; Takayuki Arai and William A. Yost, Education in Acoustics; Shoji Makino and Timothy Leishman, Engineering Acoustics; Shigeru Yoshikawa, Musical Acoustics; Timothy F. Noonan, Kerrie Standlee, and Hiro Takinami, Noise; Henry E. Bass, Jun Kondoh, and Junichi Kushibiki, Physical Acoustics; Yoshitaka Nakajima and William A. Yost, Psychological and Physiological Acoustics; John C. Burgess and Yoshifumi Chisaki, Signal Processing in Acoustics; Victoria Anderson, Sadaoki Furui, Keikichi Hirose, and Amy

FIG. 6. New Fellows of the Acoustical Society of America with ASA President and Vice President (l to r): Philippe Blanc-Benon, Anthony Atchley, Michael Vorländer, Anthony Gummer, Anders Gade, Marehalli Prasad, David Conant, Jody Kreiman, Charles Holland, Sergio Beristain, Kevin LePage, Hiroshi Riquimaroux, and Whitlow Au.

FIG. 7. ASA President Anthony Atchley (l) presents the ASA Distinguished Service Citation to Thomas D. Rossing (r).

FIG. 8. ASA President Anthony Atchley (l) presents the Silver Medal in Noise to Alan H. Marsh.

Schafer, Speech Communication; Dean Capone and Hiro Takinami, Structural Acoustics and Vibration; and Todd R. Beiler and Martin Siderius, Underwater Acoustics.

The Science Writing Award for Professionals in Acoustics was presented to Taras Gorishnyy, Martin Maldovan, Chaitanya Ullal, and Edwin Thomas for their article "Phonic Crystals," which appeared in the December 2005 issue of *Physics World* (see Fig. 4). The Science Writing Award in Acoustics for Journalists to Radek Boschetty for the BBC Program "The Noisy Ape" was announced.

Aaron Thode, recipient of the 2005 A. B. Wood Medal and Award of the Institute of Acoustics (UK), was introduced (see Fig. 5).

Election of 17 members to Fellow grade was announced and fellowship certificates and pins were presented. New fellows are Sergio Beristain, Philippe Blanc-Benon, David A. Conant, Anders C. Gade, Anthony W. Gummer, Charles W. Holland, Jody E. Kreiman, Kevin D. LePage, James A. McAteer, David R. Palmer, Marehalli G. Prasad, Hiroshi Riquimaroux, Peter A. Rona, Mark V. Trevorrow, Michael Vorländer, Joos Vos, and Ben T. Zinn. (see Fig. 6).

The Distinguished Service Citation was presented to Thomas D. Rossing, Stanford University, "for contributions to the Society in bringing the joy of scientific discovery and knowledge of acoustics to young people, teachers, and Society members" (see Fig. 7). The Silver Medal in Noise was presented to Alan H. Marsh, DyTec Engineering, "for contributions to the reduction of aircraft noise and for improvement to the quality of acoustical standards" (see Fig. 8). The Silver Medal in Physical Acoustics was presented to Henry E. Bass, National Center for Physical Acoustics, University of Mississippi, "for leadership in physical acoustics and contributions to the understanding of atmospheric sound propagation" (see Fig. 9). The Silver

FIG. 9. ASA President Anthony Atchley (l) presents the Silver Medal in Physical Acoustics to Henry E. Bass.

FIG. 10. ASA President Anthony Atchley (l) presents the Silver Medal in Psychological and Physiological Acoustics to William A. Yost.

FIG. 11. ASA President Anthony Atchley (l) presents the Wallace Clement Sabine Medal to William J. Cavanaugh.

FIG. 12. Sadaoki Furui, ASJ General Chair of the joint meeting.

FIG. 13. ASA and ASJ award of merit recipients (l to r): Yôiti Suzuki, Anthony Atchley, Sadaoki Furui, Whitlow Au, Hiroshi Sato, Donna Neff, Charles Schmid, and Elaine Moran.

Medal in Psychological and Physiological Acoustics was presented to William A. Yost "for contributions to understanding pitch perception, sound source localization, and auditory processing of complex sounds" (see Fig. 10). The Wallace Clement Sabine Medal was presented to William J. Cavanaugh "for contributions to the practical application of architectural acoustics in building design and to education in architectural acoustics" (see Fig. 11).

Anthony Atchley announced the presentation of ASA awards of appreciation to the joint-meeting organizers from the Acoustical Society of Japan. Medals were presented to Sadaoki Furui, ASJ General Chair; Hiroshi Sato, ASJ Meeting Secretary; and Yôiti Suzuki, ASJ Technical Program Chair. Anthony Atchley then introduced ASJ President Yôiti Suzuki who together with Sadaoki Furui, ASJ General Chair of the Meeting (see Fig. 12), announced and presented awards of appreciation to ASA meeting organizers. Awards and certificates were presented to Whitlow W.L. Au, ASA Technical Program Chair; Donna L. Neff, ASA Past Vice President; Charles E. Schmid, ASA Executive Director; Elaine Moran, ASA Headquarters Office Manager; and Anthony A. Atchley, ASA Technical Program Chair (see Fig. 13).

The full technical program and award encomiums can be found in the printed meeting program or online at <scitation.aip.org/jasa> (select Volume 120, No. 5) for readers who wish to obtain further information about the meeting. We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information about future meetings can be found in the *Journal* and on the ASA Home Page at <http://asa.aip.org>.

ANTHONY A. ATCHLEY

President 2006–2007

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.

2007

4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, UT [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org].
24–26 July	"Revolutionary Aircraft for Quiet Communities", a NASA workshop hosted by the National Institute of Aerospace and co-sponsored by the Joint Planning and Development Office and the Council of European Aerospace Societies, Hampton, VA [Web: www.nianet.net].
5–8 Oct.	123rd Audio Engineering Society Convention, New York, NY [Audio Engineering Society, 60 E. 42 St., Rm. 2520, New York, NY 10165-2520, Tel.: 212-661-8528; Fax: 212-682-0477; Web: www.aes.org].

2007

22–24 Oct.	NOISE-CON 2007, Reno, NV [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.org].
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, LA (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; Web: http://asa.aip.org].

2008

29 June–4 July	Joint Meeting of the Acoustical Society of America (ASA), European Acoustical Association (EAA), and the Acoustical Society of France (SFA), Paris, France [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; Web: http://asa.aip.org/meetings.html].
27–30 July	NOISE-CON 2008, Dearborn, MI [Institute of Noise Control Engineering, INCE Business Office, 210 Marston Hall, Ames, IA 50011-2153, Tel.: (515) 294-6142; Fax: (515) 294-3528; E-mail: ibo@inceusa.org].
28 July–1 Aug	9th International Congress on Noise as a Public Health Problem Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise), Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, P. O. Box 1609, Groton, CT 06340-1609, Tel. 860-572-0680; Web: www.icben.org; E-mail icben2008@att.net].

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

- **Volumes 1–10, 1929–1938:** JASA and Contemporary Literature, 1937–1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10.
- **Volumes 11–20, 1939–1948:** JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 395. Out of Print.
- **Volumes 21–30, 1949–1958:** JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75.
- **Volumes 31–35, 1959–1963:** JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90.
- **Volumes 36–44, 1964–1968:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.
- **Volumes 36–44, 1964–1968:** Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print.
- **Volumes 45–54, 1969–1973:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
- **Volumes 55–64, 1974–1978:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
- **Volumes 65–74, 1979–1983:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound).
- **Volumes 75–84, 1984–1988:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
- **Volumes 85–94, 1989–1993:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
- **Volumes 95–104, 1994–1998:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632. Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).
- **Volumes 105–114, 1999–2003:** JASA and Patents. Classified by subject and indexed by author and inventor. Pp.616. Price: ASA members $50; Nonmembers $90 (paperbound).

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

April 2007

2–4 **International Conference on Emerging Technologies in Non-destructive Testing**, Stuttgart, Germany. (Web: www.fa.asso.fr/secretariat/etech-flyerStuttgart.pdf)

9–12 **International Congress on Ultrasonics (2007 ICU)**, Vienna, Austria. (Fax: +43 158801 13499; Web: www.icultrasonics.org)

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

16–18 **29th International Symposium on Acoustical Imaging**, Shonan Village Center, Kanagawa Pref., Japan. (Web: publicweb.shonan-it.ac.jp/ai29/AI29.html)

24–25 **Institute of Acoustics (UK) Spring Conference**, Cambridge, UK. (Web: www.ioa.org.uk/viewupcoming.asp)

May 2007

15–18 *__6th International Symposium on Hydroacoustics joint with domestic XXIV Symposium on Hydroacoustics (SHA2007)__, Leba, Poland. (Web: www.hydro.eti.pg.gda.pl/sha2007)

June 2007

1–3 **Second International Symposium on Advanced Technology of Vibration and Sound**, Lanzhou, China. (Web: www.jsme.or.jp/dmc/Meeting/VSTech2007.pdf)

3–6 *__14th International Conference on Noise Control (noise control'07)__, Elblag, Poland. (Web: www.ciop.pl/noise_07)

3–7 **11th International Conference on Hand-Arm Vibration**, Bologna, Italy. (Web: associazioneitalianadiacustica.it/HAV2007/index.htm)

4–6 **Japan-China Joint Conference on Acoustics**, Sendai, Japan. (Fax: +81 3 5256 1022; Web: www.asj.gr.jp/eng/index.html)

18–21 **Oceans07 Conference**, Aberdeen, Scotland, UK. (Web: www.oceans07ieeeaberdeen.org)

25–27 *__31st International AES Conference - New Directions in High Resolution Audio__, London, UK. (Web: www.aes.org/events/31/)

25–29 **2nd International Conference on Underwater Acoustic Measurements: Technologies and Results**, Heraklion, Crete, Greece. (Web: www.uam2007.gr)

July 2007

2–6 **8th International Conference on Theoretical and Computational Acoustics**, Heraklion, Crete, Greece. (Web: www.iacm.forth.gr/~ictca07)

3–5 **First European Forum on Effective Solutions for Managing Occupational Noise Risks**, Lille, France. (Web: www.noiseatwork.eu)

4–7 **International Clarinet Association Clarinetfest**, Vancouver, British Columbia, Canada. (e-mail: john.cipolla@wku.edu; phone: 1 270 745 7093)

9–12 **14th International Congress on Sound and Vibration (ICSV14)**, Cairns, Australia. (Web: www.icsv14.com)

16–21 **12th International Conference on Phonon Scattering in Condensed Matter**, Paris, France. (Web: www.isen.fr/phonons2007)

August 2007

6–10 **16th International Congress of Phonetic Sciences (ICPhS2007)**, Saarbrücken, Germany. (Web: www.icphs2007.de)

27–31 **Interspeech 2007**, Antwerp, Belgium. (Web: www.interspeech2007.org)

28–31 **Inter-noise 2007**, Istanbul, Turkey. (Web: www.internoise2007.org.tr)

September 2007

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

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

9–12 **ICA Satellite Symposium on Room Acoustics (ISRA2007)**, Sevilla, Spain (Web: www.ica2007madrid.org)

10–13 *__54th Open Seminar on Acoustics (OSA2007)__, Przemysl, Poland. (Web: www.univ.rzeszow.pl/osa2007/)

17–19 **3rd International Symposium on Fan Noise**, Lyon, France. (Web: www.fannoise.org)

18–19 **International Conference on Detection and Classification of Underwater Targets**, Edinburgh, Scotland, UK. (Web: ioa.org.uk)

19–21 **Autumn Meeting of the Acoustical Society of Japan**, Kofu, 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)

20–22 **Wind Turbine Noise 2007**, Lyon, France. (Web: www.windturbinenoise2007.org)

24–28 **XIX Session of the Russian Acoustical Society**, Nizhny Novgorod, Russia. (Web: www.akin.ru)

October 2007

3–5 *__Pacific Rim Underwater Acoustics Conference 2007__, Vancouver, BC, Canada. (Web: PRUAC.apl.washington.edu)

9–12 **2007 Canadian Acoustic Conference**, Montréal, Québec, Canada. (Web: caa-aca.ca)

25–26 **Autumn Meeting of the Swiss Acoustical Society**, Bern, Switzerland. (Web: www.sga-ssa.ch)

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 (Web: www.congrex.com/18th_isna)

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

August 2008

25–29 *__10th International Conference on Music Perception and Cognition (ICMPC 10)__, Sapporo, Japan. (Web: icmpc10.typepad.jp)

September 2008

22–26 **INTERSPEECH 2008 - 10th ICSLP**, Brisbane, Australia. (Web: www.interspeech2008.org)

 0001-4966/2007/121(4)/1817/2/$23.00

October 2008

26–29 *__inter-noise 2008__, Shanghai, China. (Web: www.internoise2008.org)

November 2008

2–5 *__IEEE International Ultrasonics Symposium__, Beijing, China. (Web: www.ieee-uffc.org/ulmain.asp?page=symposia)

September 2009

6–10 __InterSpeech 2009__, Brighton, UK. (Web: www.interspeech2009.org)

August 2010

23–27 __20th International Congress on Acoustics (ICA2010)__, Sydney, Australia. (Web: www.ica2010sydney.org/)

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.

Archaeoacoustics

Chris Scarre and Graeme Lawson

McDonald Institute for Archaeological Research and Oxbow Books, 2006, 118 pp. $50.00 (hardcover), ISBN:-10:1-902937-35-X

This book evolved from a workshop on archaeoacoustics held at Cambridge University by the McDonald Institute in June, 2003. The Preface states that the book is intended to give archaeoacoustics "the prominence in archaeology that it most certainly deserves." As an avocational archaeoacoustician, I warmly endorse that purpose.

Thirteen practitioners of archaeoacoustics contribute 12 chapters spanning a broad range of subjects that include: the history and prehistory of human awareness of sound, the creation of sound instruments, and, most interesting to me, the acoustics of built and unbuilt spaces, and soundscape.

Archaeology seems to be undergoing another paradigm shift. Supposedly the shift is not just about sound. The aim is a sensual archaeology that includes "touch, smell, and hearing." At this time, however, archaeoacoustics seems to be generating the most excitement.

The evidence for prehistoric human sound awareness discussed in this book is no surprise. Other animals are highly sound aware. Would we expect less for humans? More intriguing is the evidence for deliberate manipulation of sound, not only by ancient humans, but possibly even by Neanderthals. Some evidence suggests that by the late Paleolithic, humans were conferring unique acoustic properties to fabricated sound instruments and marking resonant niches in caves. There is also evidence that Neolithic people built unique acoustic features into the architecture of temples and burial chambers, although it remains to be proven that these features were intended. The vexing question of intentionality is the common theme connecting these chapters.

Francesco d'Errico and Graeme Lawson devote their chapter to the question of intentionality in considering Paleolithic and Medieval bone pipes, flutes, and tuning pegs. Ezra B. W. Zubrow and E. C. Blake discuss the Mousterian bone flute dated 43,000 years before the present (BP) in their chapter on the Origins of Music and Rhythm. Were those almost perfectly conical holes fabricated for their sound by humans? Or are they bite marks of a cave bear with exceptional orthodontia? Or are they taphonomical effects (changes after death from a variety of causes).

Graeme Lawson devotes most of his chapter to the acoustic features of medieval stone buildings, and of earlier medieval timber halls. He believes they may have been intended for playing the lyre.

Several chapters are devoted to music archaeology—a field recognized by archaeologists even before the aforementioned paradigm shift. Peter Holmes considers sound intentionality in Scandinavian lurs of the late bronze age. Their sounds critically depend on instrument morphology, and especially on the conicity of their brass tubes. A study opportunity arose because these lip reed instruments are often made in left-right pairs by the same makers using the same tools and methods. It was hoped that comparing their sounds might reveal the acoustical intentions of their makers. It did not reveal intentions, but did provide strong, if unsurprising, hints. Holmes found that each instrument closely mirrored its mate in morphology and sound. He also found that lurs were fabricated so that "a player can sound a series of notes that accurately conform to the harmonic series of notes." Clearly, that would serve musical purposes in "the modern western sense." But Holmes concludes that he can only assume, does not know, if that was the fabricator's intention.

The acoustics of Paleolithic cave art and rock art are prominent in this book. A chapter is contributed by Iegor Resnikoff, perhaps the preeminent pioneer of archaeoacoustics, on the evidence for human use of sound resonance from Paleolithic to Medieval times. Resnikoff writes from the perspective of music and architectural history more than physics. His 1987 study found relationships between sound and art in Paleolithic caves. It attracted little interest at first, perhaps because it was published in a relatively obscure journal. This changed dramatically when Chris Scarre reviewed Resnikoff and M. Dauvois' work in a higher profile journal ("Painting by Resonance," [Nature, 338, 1989, 382]). The review conveyed an image of painted caves as Paleolithic cathedrals in which cave artists chanted incantations before cave art, and where rituals of dance, song, drums, flutes, and whistles were conducted in flickering torchlight.

Paul Devereux is the author of the first-ever book on archaeoacoustics ("Stone Age Soundtracks," 2001). His chapter reports my early findings at Chichen Itza in Mexico. He also reports discoveries at other archaeological sites, including the Treasury of Atreus in Mycenae, and Ireland's Newgrange, as well as the Wayland Smithy passage grave in England and a Neolithic chambered tomb in Cornwall. Devereux sought evidence of intentionality by comparing the acoustic resonance frequencies of chambers with brain response to tones (e.g., 95–120 Hz) and with the formant frequencies of the male voice. Devereux seems more open to belief in the influence of acoustical design on hallucinatory experience than some other contributors.

A chapter by British acoustician Aaron Watson reports sound measurements at high profile Neolithic monuments, including Stonehenge and Maeshowe. Watson shows his acoustical training explicitly as co-author (with Ian Cross) of another chapter on "socially organized sound." It describes conventional architectural acoustic measurement methodologies, and rightly questions their applicability to societies with different esthetics. The reports of methodical sound measurements at Stonehenge are interesting, but inconclusive, in part because so little is known about the culture. The effects may be chance. Still it is argued, sonic effects may have been noticed by users and incorporated into rites there. But we are given little evidence for that. What seems more promising are reports of eerie acoustic effects at other sites. We might not think to listen for these effects because of our western bias. Watson's ideas about flutter echoes parallel my own, but I think he has not yet found an ancient site where they might be important.

Rock art acoustics is ably represented in this book by Steven J. Waller, probably its preeminent practitioner. Like Resnikoff, Waller believes that sound was a motivator for the placement of cave art. He also believes that echoes motivated the placement of some outdoor rock art and in some cases, the art itself. He has tirelessly collected ethnographic field reports and studies of rock art acoustics. He documents an impressive number of echo myths of tribes and civilizations throughout the world and human history. Those myths reveal a long history of human interest in sound that seems deeply embedded in the human psyche. If that history has become unintelligible to us, it is because of noise and the learned disregard for sound that chronic noise engenders. If that is true, noise pollution may have another adverse effect. It may be cutting humans off from their ancient sonic heritage, just as smoke and light pollution renders the once-impressive night sky almost meaningless to moderns. Many rock art experts promote rock art conservation, but Waller encourages preservation of the soundscape at these sites.

Ethnographic studies of sound appear in other chapters as well. Ian Morley considers hunter-gatherer music among Native Americans, African Pigmies, Australian Aborigines, and Eskimos, drawing implications about their uses of acoustic space. Ezra Zubrow and Elizabeth Blake explore the origin of music and rhythm, suggesting that rhythm originated with the heartbeats of flint knappers.

At several places in this book we encounter sacred sites and other ancient spaces that possess compelling acoustic qualities. Is it a coincidence or a clue that both ancient temples and modern worship spaces often have special acoustics? Many people (including me) believe there are deep connections between sound, magic, and numinous experience. Individuals who

 0001-4966/2007/121(4)/1/0/$23.00

report visions or supernatural experiences find they are often mediated by unusual sound qualities of the space. For that reason, archaeoacoustics often crosses over to the study of religious experience.

DAVID LUBMAN
David Lubman and Associates
Westminster, California

Ultrasound Imaging: Waves, Signals, and Signal Processing

Bjorn A. J. Angelsen

Emantec, Norway, 2000. 1416 pp. $490.00. ISB: 82-995811-0-9, ISBN: 82-995811-1-7

This text is the result of research and teaching by the author and his colleagues in the area of ultrasound imaging. Although the main author is Bjorn A. J. Angelsen, Dr.Techn., Professor of Biomedical Engineering, Department of Physiology and Biomedical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, Chapter 10 was contributed by Hans G. Torp, Dr. Techn., Professor of Medical Informatics, Department of Physiology and Biomedical Engineering, Norwegian University of Science and Technology, Trondheim, Norway, and Section 6.4 was contributed by Sverre Holm, Dr. Ing., Professor of Signal Processing, Department of Informatics, University of Oslo, Oslo, Norway. The book reflects Professor Angelsen's professional life in the field of medical ultrasound imaging, beginning with Doppler ultrasound in the '70s, development of annular array colorflow imaging in the '80s, and the propagation of ultrasound in inhomogeneous tissue in the '90s. Much of the material in the book, e.g., particularly about wave propagation in inhomogeneous tissue and correction of phase aberration as well as reverberation, has not previously been published.

The book is comprised of two volumes pedagogically organized in 12 chapters that span elementary introductory material through current research topics and in a 115-page appendix that gives an overview of relevant mathematical methods. The introductory material is an easy-to-read overview of methods and instrumentation used in ultrasound imaging. Intermediate-level material introduces mathematical concepts of wave propagation, beamformation, and signal processing. Advanced material is at the current frontier of research in nonlinear wave propagation through inhomogeneous tissue, in aberration correction, and in scattering from ultrasound contrast agents. A set of problems is at the end of each chapter. Navigation in the book is facilitated by a 17-page alphabetical index and by a list of chapter contents at the beginning of each chapter.

This is a well-organized, comprehensive, and informative text that covers all aspects of ultrasound wave propagation, signal modeling, and signal processing in medical ultrasound imaging.

The first volume starts with an overview of methods and instrumentation used for imaging tissue and blood flow. This is followed by an analysis of one-dimensional vibrations in plates and ultrasound transducers. The analysis is then expanded to waves in three-dimensional space for the description of propagation and beamformation in lossy homogeneous media. A thorough analysis of absorption mechanisms in tissue is included. A detailed analysis of beamformation with a single-element transducer and with arrays completes the first volume.

The second volume describes propagation and scattering in lossy inhomogeneous media. Scattering and received signals are modeled for b-scan imaging and Doppler measurements and imaging of blood flow. Signal processing to image amplitude and Doppler shift is described. Scattering is first analyzed using the Born approximation and linear propagation. In the final two chapters, Chapter 11 and Chapter 12, second-order scattering is treated. Chapter 11 models propagation-induced aberration and reverberation in tissue. Chapter 12 describes nonlinear propagation of ultrasonic waves and ends with an analysis of nonlinear scattering from ultrasound contrast agents. Topics in these last chapters are currently active areas of research in medical ultrasound and, therefore, can be used to bring student researchers up to speed quickly. A good list of references is provided at the end of each chapter. These lists serve as a guide to important papers and research in the field.

The development of the book started with Professor Angelsen's Ph.D. thesis in the area of Doppler signal analysis in the '70s. In the '80s, Professor Angelsen's attention turned to transducers, beamforming, and tissue characterization. This led to Chapter 4 about propagation in lossy homogeneous media, Chapter 5 about single-element transducers, Chapter 6 about arrays, and Chapter 7 about scattering. Professor Angelsen's recent interests in nonlinear propagation, imaging contrast agents, and aberration of ultrasonic beams by tissue in inhomogeneities resulted in Chapter 11 and Chapter 12.

As noted in the Preface, the text forms the basis for two four-hour courses given on acoustics and signal processing in medical ultrasound imaging at the Norwegian University of Science and Technology. The first course is taught in the fourth year of a five-year engineering curriculum. The second course is taught at the Ph.D. level. Since the text contains more material than can reasonably be covered in two courses, the topics taught in one or two courses must be chosen selectively. No matter what the selection of topics, a student in ultrasound would be well prepared to study independently other topics in the text by having experienced the presentation of selected topics in one or two formal courses based on the text.

A version of the first chapter that covers basic material is now available in a 130-page volume with seven color illustrations. This is a useful collection of material for students entering the field and company employees who are new to the area of ultrasound. The main difference between this material and introductory material found elsewhere is a view from someone who is expert in the signal processing aspects of ultrasound imaging as well as the physical acoustic principles on which the processing is based.

The view of ultrasound imaging from such an individual is also evident throughout the two-volume text. This perspective distinguishes the text from other books such as the various well-known editions of the book first authored by Kinsler and Frey, the book by Pierce, the text by Morse and Ingard, and the more recent books by Blackstock, by Szabo, and by Shung and Thieme. Professor Angelsen presents his material at a level that is generally higher than that of all the Kinsler and Frey editions. The mathematical emphasis on signal processing as it is driven by phenomenon associated with wave propagation differentiates the book from texts by Pierce, Morse and Ingard, and Blackstock. Angelsen's book contains more derivations and is, therefore, more mathematically comprehensive than Szabo's book that is also limited to ultrasound in medicine written primarily at the graduate level and is well referenced to the literature of current research. The book that Shung and Thieme edited treats only a limited number of topics.

A strength of Professor Angelsen's two-volume book is the strong emphasis on the mathematical treatment of physical phenomenon. A corresponding weakness is the paucity of text that explains and interprets equations running throughout the text. On balance, however, this unique and comprehensive treatment of ultrasound imaging in medicine is a valuable reference for experienced researchers in the field and also a useful text for students, particularly those who are interested in signal processing, to learn about medical ultrasound.

ROBERT C. WAAG
Yates Professor of Engineering
Department of Electrical and Computer Engineering and
Department of Imaging Sciences
University of Rochester
Rochester, New York 14642

Spaces Speak, Are You Listening? Experiencing Aural Architecture

Barry Blesser and Linda-Ruth Salter

The MIT Press, Cambridge, MA, 2007, 437 pp, Price: $39.95 (cloth), ISBN 13: 978-0-262-02605-5

Human beings have an awareness of the spaces in which they are living. This experience results from input to all the sensory modalities and is heavily based on cognition. Thus, it is always elucidating to look at multimodal phenomena by primarily focusing on a single modality as, for example, Patrick Sueskind has dramatically demonstrated for the olfactory sense in his novel *Perfume* (Penguin Books, 1985). Now we have a book that looks at the world from an aural point of view, focusing on the phenomenon of "auditory spatial awareness" and on the people that design and engineer it: "aural architects."

However, this is neither a novel nor a stringently scientific work. It is a nonfiction monograph, written by an experienced and worldwide recog-

nized senior engineer, consultant, and academic teacher in audio technology, co-authored by a scholar in the arts and social sciences. The authors themselves call their work an "intellectual mosaic," attempting to fuse disparate knowledge into a common framework whereby they elegantly mix facts and opinions. Their approach concentrates on aural spatial phenomena, but they treat these from a broad, interdisciplinary viewpoint.

The volume presents no easy reading, but once you have worked your way into it, you will be swept away by its scientific and philosophical richness. In fact, the work "contains a lot of truth" to quote a German saying. This reviewer, once started, could not stop reading for two days and a night.

There are nine chapters plus some acknowledgments and personal statements by the first author. Chapter 1 is the introduction to the main topic, "aural architecture." It deals with the essence of architecture in general and specifies its subdiscipline "aural architecture," specifically, as the meaningful manipulation of the aural properties of space.

Chapter 2 tackles auditory spatial awareness as an experience, featuring subtopics such as soundscapes, unusual spaces, the social component of spatial awareness, and the components of spatial experience. The concept "acoustic arena" is introduced as a section of space within which a given group of people can share sonic events aurally. Another topic is navigation by listening, an art which some blind people master excellently. Other treated items are aural enrichments, spatial distortions, illusions of expanded space, aural textures, and the enveloping effect of reverberation.

Chapter 3 deals with aural spaces from prehistory to the present, with the notion of understanding the aural experience of space as a cultural filter. The discussion moves from caves through modern concert halls to radio broadcasting, including topics like auditory icons, religion, industrialization, and modern audio technology—always considering related social forces.

Chapter 4 regards the aural arts and musical spaces. In a philosophical introduction, musical spaces are looked at as artistic abstractions. This is followed by an analysis of what actually happens in these spaces perceptually, e.g., temporal and spatial spreading of the auditory events. The authors explain that, being elements of the sound-production process, spaces transform proto-instruments into metainstruments. The artistic value of auditory spreading is commented on in detail. Finally, spatial rules are analyzed concerning their applicability to music making.

Chapter 5 is labeled "Inventing Virtual Spaces for Music." It provides insights into the invention process itself and presents a number of realized examples. To start with, the artistic dimensions of space and location are identified, and the way in which controlled auditory location is used in music is explained. As documented in this chapter, natural spaces have been used to add spatial attributes to music; this has led directly to the creation of virtual aural spaces by means of electroacoustics. As an early example, the dome of loudspeakers which has been provided for Karlheinz Stockhausen at the Osaka World's Fair 1970 is discussed. The two basic acoustic approaches of presenting spatial sound to listeners—namely, head-related with headphones or transaural systems, and room-related with multichannel arrangements of loudspeakers—are treated in detail, specifically regarding typical application scenarios such as automobiles, virtual spaces, live auditoria, and movie theatres.

The title of Chapter 6 is "Scientific Perspectives of Spatial Acoustics." Here, the authors explicitly refer to science. In fact, this is the chapter where the first author draws heavily on his knowledge and experience as a technical expert in audio engineering. An interesting fact, however, is that there is not a single mathematical formula in this chapter—nor in the whole book. Yet, this does not in any way impair the exactness of the description of scientific and technical details. Actually, this reviewer feels confirmed in his teaching approach that thoroughly formulated prose can be as exact as mathematical symbolism. If it is not possible to describe the contents of a formula in words, something must be wrong with the formula. The scientific and technical details critically presented in this chapter extend from perceptual evaluation of concert-hall acoustics, such as psychoacoustic-measurement methods, preference judgments, identification of audible acoustic defects, and auditory source width and envelopment, to mathematical issues such as models of enclosed spaces and reverberation as a random process. A major section concerns generators for artificial reverberation. This is the nucleus of the first author's technical competence, having built the first commercially available reverberator based on digital signals processing as early as 1974—together with Karl O. Bäder. Besides various technical details being disclosed in this chapter, it is especially interesting that a complete history of reverberators is presented and commented on here by a first-hand eyewitness. The author is obviously capable of self-critical reflection about his own role and the social context in which he was acting.

Chapter 7 deals with the people who actually prompt spatial innovation, and how this activity fits into their private agendas. This chapter has a heavy psychological and sociological touch. It discusses issues relating to social values such as goals, rewards and careers, resources, decision-making, knowledge, political power, conservatism as a cultural tradition, and career management. Fundamental concepts like subjectivity, personality, and cognitive judgment are touched upon, and more "global" issues like the life cycle of disciplines are broadly discussed. The intellectual framework is identified that underlies assumptions and paradigms, and the process of fusing intellectual fragments. The work of different categories of observers such as expert perceivers using formal science versus folk-science is examined, and epistemological backgrounds are analyzed. This reviewer read this chapter with commitment. He got the impression that the individual viewpoint of the authors stands out pronouncedly here. In other words, in a way this monograph is also an intellectual autobiography of the authors.

Chapter 8 treats spatial awareness as an evolutionary artifact. This is a compelling view, although the authors state correctly that Darwin's hypothesis of environmental pressure as a causal reason for evolution cannot be proven. Nevertheless, readers will agree with the authors that evolution as an empirical fact exist. In any case, to build a story around evolution is tempting as it allows for including ample speculations. Consequently, Chapter 8 is very entertaining, maybe because it is meant as a narrative exposition rather than a scholarly based proof on reliable knowledge. Still and all, many of the statements made by the authors seem quite plausible.

The final chapter, Chapter 9, concludes and sums up the material that has come before. Actually, readers may be advised to start their browsing of the book from this point. As mentioned earlier, the book is extremely interesting reading, specifically for the following reasons. First, it provides the first comprehensive coverage of auditory spatial awareness and aural architecture, although it is not a textbook for aural architects. Second, it presents relevant technical details of artificial reverberation and auditory virtual-reality generators. Third—and this makes the book so fascinating—it allows a glimpse into how an experienced audio-technology expert and his wife, an interdisciplinary social scientist, both educated in and residing on the east coast of the United States in the 20th century, experience the world of science and technology at large. This makes the work an invaluable document of our time.

JENS BLAUERT
Institute of Communication Acoustics
Ruhr-Universität Bochum
Bochum, Germany

 0001-4966/2007/121(4)/1/0/$23.00

REPORTS OF RELATED MEETINGS

This Journal department provides concise reports of meetings that have been held by other organizations concerned with acoustical subjects; and of meetings co-sponsored by the Acoustical Society but planned primarily by other co-sponsors.

17th International Symposium on Nonlinear Acoustics (ISNA17) including the International Sonic Boom Forum

The 17th International Symposium on Nonlinear Acoustics (ISNA) was held 18–22 July 2005 at the Penn Stater Conference Center Hotel on the campus of The Pennsylvania State University in State College, PA, USA, with 178 participants. Following in the footsteps of the previous ISNAs, the scope of the symposium covered nonlinear acoustical phenomena in solids, liquids, and gases. Nineteen technical sessions were held. Sessions devoted to special topics and consisting of invited and contributed papers included Nonclassical Nonlinear Acoustics of Solids and Nondestructive Evaluation (NDE) Applications (organized by P. A. Johnson, L. A. Ostrovsky, and I. Solodov), Elastic Wave Effects on Fluids in Porous Media (P. M. Roberts and I. B. Esipov), the Science and Application of High Intensity Focused Ultrasound in Medicine and Biology (R. A. Roy), Shock Wave Therapy (M. R. Bailey), Infrasound (K. A. Naugolnykh and A. J. Bedard), Harmonic Imaging in Diagnostic Ultrasound (R. O. Cleveland), and Thermoacoustics (R. M. Keolian). General sessions consisting of contributed papers included Sound Beams, Resonators, and Streaming (chaired by B. O. Enflo), Bubbles, Particles, and Flows (P. L. Marston), Propagation, Shocks, and Noise (H. Hobæk) Nonlinear Acoustics in Solids (V. Espinosa), and Nonlinear Acoustics in Medicine and Biology (Yu. A. Ilinskii and E. A. Zabolotskaya), as well as a poster session.

In addition, ISNA17 included the International Sonic Boom Forum (ISBF) co-organized by V. W. Sparrow and F. Coulouvrat. This set of special sessions encapsulated increasing international interest in building and operating small supersonic jets with low-amplitude, shaped boom signatures. The timing of the ISBF occurred when many research programs were beginning to address the technical feasibility and design for such aircraft. The purpose of holding the ISBF was to foster technical communication and exchange between university, industry, and government scientists, engineers, and executives interested in sonic booms. This Forum provided a timely and unique opportunity to communicate on sonic boom and to obtain information on the latest research advances and progress toward possible community acceptance of shaped sonic boom. Although not represented in these proceedings, the participation of Peter Coen of NASA Langley Research Center as one of the ISBF Plenary speakers and of all the ISBF panelists is greatly appreciated. The panelists included Akira Murakami, Institute of Space Technology and Aeronautics (JAXA); Laurette Fisher, Federal Aviation Administration; Kenneth Orth, Consultant to Gulfstream; Gerard Duval, retired Concorde pilot for Air France; Thierry Auger, Airbus France; Sam Bruner, Raytheon; Nicolas Heron, Dassault Aviation; Tom Hartmann, Lockheed-Martin Aeronautics; and Richard G. Smith, NetJets Inc. The ISBF organizers also appreciate Gulfstream Aerospace Corporation temporarily locating their portable sonic boom simulator, Supersonic Acoustic Signature Simulator Generation II, at ISNA17 for demonstration to interested ISBF and ISNA17 attendees.

The proceedings of ISNA17, *Innovations in Nonlinear Acoustics*, are published as number 838 in the AIP Conference Proceedings Series. Information can be found at http://proceedings.aip.org/proceedings/

Conferences and symposia play a vital role in the advancement of science and engineering. They provide a forum for the exchange of ideas and cultures and provide a sense of continuity and community. Faced with the ever increasing number of conferences that one could or should attend, it is fair to ask whether a symposium series such as the International Symposium on Nonlinear Acoustics is necessary. Does it play a unique and important enough role to justify the resources required to sustain it? Based on the experience gained through organizing and hosting ISNA17, we believe that the answer to this question is yes. While it is true that topics presented at ISNA17 are found at other conferences, no single venue captures the diversity of the field of nonlinear acoustics as does ISNA. It brings together researchers from disparate fields as no other forum does. It provides a place to renew old friendships and make new ones. And it provides an opportunity to collect the latest results into a single volume, a snapshot in time from which the breadth, the depth, and the interconnectivity of the field of nonlinear acoustics can be appreciated.

ISNA17 and ISBF gratefully acknowledge the financial support of the Penn State's Graduate Program in Acoustics, the Acoustical Society of America (ASA), and the International Commission for Acoustics (ICA). We are particularly indebted to the Biomedical Ultrasound/Bioresponse to Vibration and the Physical Acoustics Technical Committees of the ASA for being strong advocates for ISNA17 and ISBF within the ASA. The support from the ICA allowed us to defray the costs of participation for some of the international participants. Finally, we express our gratitude to the ISNA International Organizing Committee for giving us the opportunity to host the international nonlinear acoustics community for a brief time.

ANTHONY A. ATCHLEY

 0001-4966/2007/121(4)/1822/1/$23.00

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:

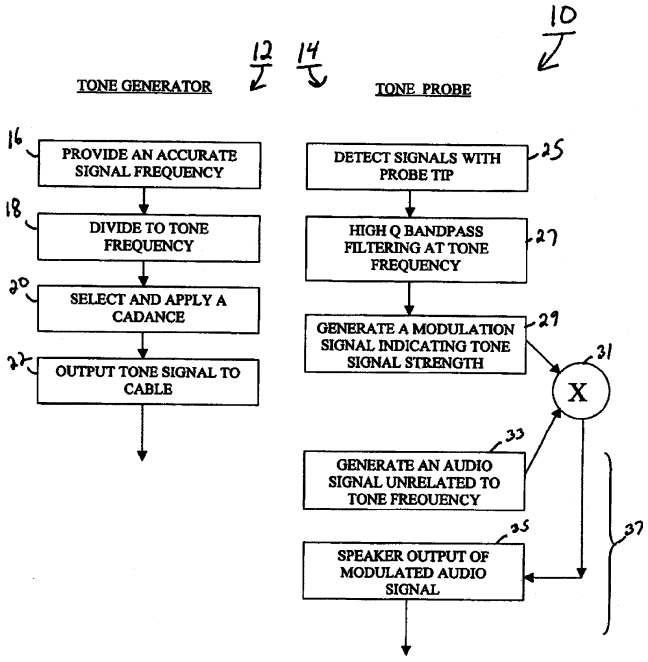
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
JOHN M. EARGLE, *JME Consulting 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*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,116,093

43.20.Ye TONE GENERATOR AND PROBE

Darrell J. Johnson and John C. McCosh, assignors to Psibor Data Systems, Incorporated
3 October 2006 (Class 324/76.28); filed 30 January 2004

A cable testing system **10** (''tracer'' or ''ringer'') is claimed where the emitted trace signal **22** is at an exact frequency and with unique modulation **29**. The probe receiver processor **14** has a narrow-band filter **27** and may emit an audio sound **25** at a frequency different from the probe frequency to reduce feedback to probe **12**.—AJC

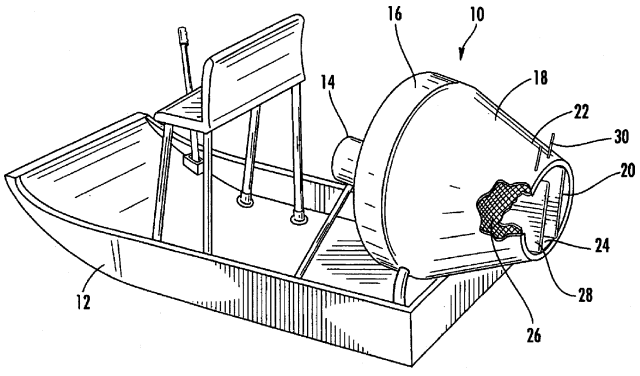


7,101,235

43.30.Nb AIR-BOAT SOUND SUPPRESSOR AND DIRECTIONAL CONTROL SYSTEM

Charles Baldwin, Palm Beach Gardens, Florida
5 September 2006 (Class 440/37); filed 26 July 2004

The propeller of an air-boat **12** is surrounded by a solid shroud **10**, which is internally lined with a layer of sound absorbing material **26**. Vanes **28** are controllable by the boat operator for steering purposes. In addition, doors on the aft portion of the shroud (not shown) are controllable by the boat operator for additional maneuverability and stoppage functions. Within the shroud may be mounted a baffle device, perhaps a cylindrical tube, that extends slightly forward of the point where the shroud diameter begins to decrease, and is said to aid in the sound reduction efficacy of the shroud without reducing its maneuverability functions.—WT



 0001-4966/2007/121(4)/1825/23/$23.00

7,106,656

43.30.Vh SONAR SYSTEM AND PROCESS

Donald Lerro *et al.*, assignors to AC Capital Management, Incorporated
12 September 2006 (Class 367/99); filed 29 September 2004

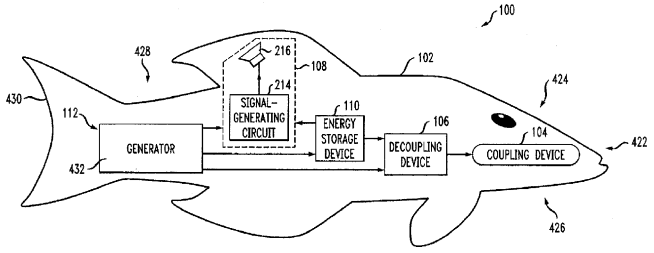
A signal processing procedure is described that allows an active and passive sonar system to operate in both modes simultaneously, even in a situation where the direct acoustic signal level is 60 dB or more higher than the echo signal level.—WT

7,113,097

43.30.Vh UNDERWATER DEVICE WITH TRANSMITTER

John Wallace Kline and Frederick R. Dental, assignors to Lockheed Martin Corporation
26 September 2006 (Class 340/573.1); filed 3 November 2003

An apparatus is discussed that can attach itself to a movable, man-made, fishlike object **100**. The fishlike object is a housing that can be coupled or decoupled from some other moving underwater object of interest, such as another vessel. In addition to coupling **104** and decoupling **106**



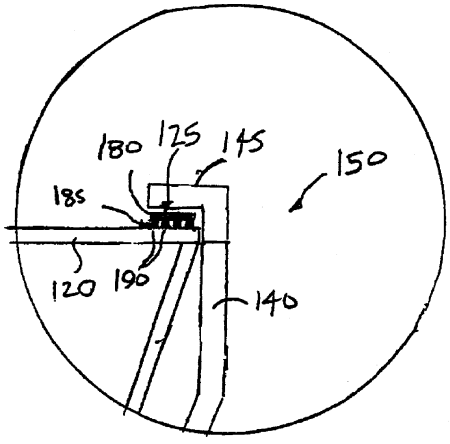
devices, which are often an electromagnet, the fishlike object houses a generator **432**, energy storage device **110**, and acoustic signal transmitting device **108**. The generator could be a water wheel activated by the movement of the fishlike object or a portion of the fishlike housing **102** may be movable, which motion then stresses a piezoelectric polymer film, thereby generating electricity.—WT

7,119,800

43.35.Pt ACOUSTIC TOUCH SENSOR WITH LOW-PROFILE DIFFRACTIVE GRATING TRANSDUCER ASSEMBLY

Joel C. Kent *et al.*, assignors to Tyco Electronics Corporation
10 October 2006 (Class 345/177); filed 24 June 2003

Display face touch screen **150** is implemented by a transducer **125** having a grating **185** formed either in the transducer face **180**, in the



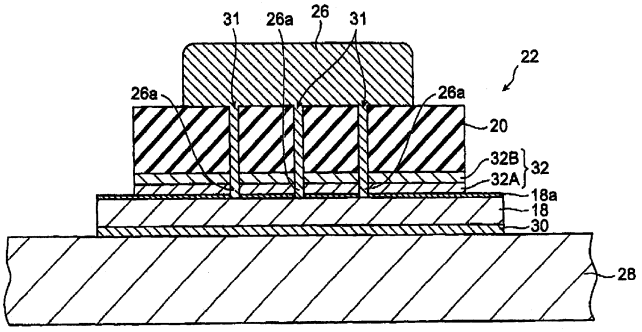
substrate **120**, or by an intermediate material **190**. The spacing between grating ridges is preferably 0.1–5 mm to launch suitable surface waves across the display face.—AJC

7,102,461

43.35.Zc SURFACE ACOUSTIC WAVE ELEMENT, SURFACE ACOUSTIC WAVE DEVICE, SURFACE ACOUSTIC WAVE DUPLEXER, AND METHOD OF MANUFACTURING SURFACE ACOUSTIC WAVE ELEMENT

Masahiro Nakano *et al.*, assignors to TDK Corporation
5 September 2006 (Class 333/133); filed in Japan 28 July 2003

Stability under repeated high-power operation of SAW duplexers depends on avoiding diffusion migration of aluminum-finger grain boundaries into those of the substrate. Single crystals of both are formed to minimize grain boundaries at the finger mounts by fast epitaxial deposition of insulating film **30** at low (80 °C) temperature followed by slow evaporation deposition of aluminum **18** to form single-crystal finger contact there. Buffer



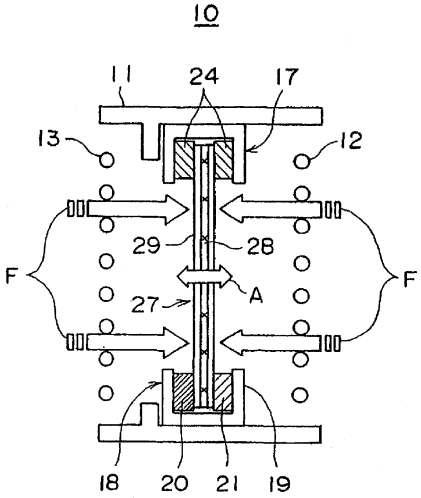
layer **32a** and adhesion layer **32b** are applied after finger-pattern photo etching, followed by aluminum conductor **20**. Good current conduction is achieved with gold bump material **26**, which is forced through galleys **31** and layers **32b**, **32a**, and incidental aluminum oxidation **18a** by compression bonding using force on and ultrasonic vibration of gold bumps **26** to achieve thermodiffusion bonds at points **26a**.—AJC

7,114,395

43.38.Ar VIBRATION DETECTOR

Hiroshi Miyazawa and Yoshikazu Oka, assignors to Kabushiki Kaisha Kenwood
3 October 2006 (Class 73/655); filed in Japan 9 February 2001

Most so-called ''optical'' microphones use a mirrored diaphragm that deflects a beam of light aimed at a receptor. This patent covers a new approach. As shown in the figure, optical waveguide **28** is fed by light source **20**. As the compound diaphragm/waveguide structure bends under



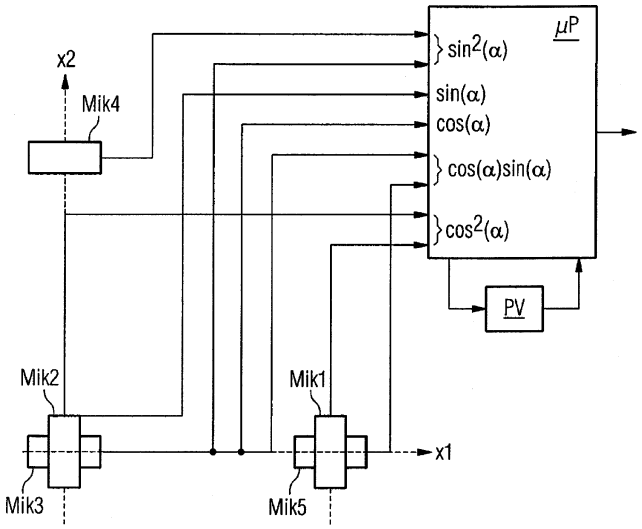
the influence of sound waves, the amount of light reaching the receptor **24** varies and the signal is then amplified. As shown, the microphone is configured as a gradient type with a figure-eight pattern.—JME

7,120,262

43.38.Ar DIRECTIONAL-MICROPHONE AND METHOD FOR SIGNAL PROCESSING IN SAME

Stefano Ambrosius Klinke, assignor to Siemens Aktiengesellschaft
10 October 2006 (Class 381/92); filed in Germany 25 May 2000

The patent deals with the synthesis of second-order directional patterns through the use of multiple monopole (omnidirectional) elements suitably



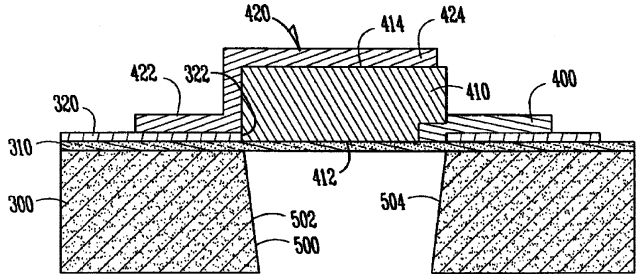
positioned and their outputs combined. Nothing really new to report. There is only one figure, a circuit diagram; how helpful it would have been had the authors shown some relevant measurements.—JME

7,116,034

43.38.Ar STRUCTURE TO ACHIEVE HIGH-Q AND LOW INSERTION LOSS FILM BULK ACOUSTIC RESONATORS

Li-Peng Wang *et al.*, assignors to Intel Corporation
3 October 2006 (Class 310/320); filed 23 February 2005

This patent describes an undercut-style film bulk acoustic resonator that supposedly minimizes the surface roughness and attendant losses associated with sputtering ZnO, PZT, or other piezoelectric materials directly on a metal electrode. The figure shows a cross section through the proposed



configuration of a vertically oriented bulk acoustic wave resonator. The crux of the invention is to use a sacrificial seed layer of smooth, crystalline material (Si or SiC) to make the top and bottom, pushing the base electrode off into a corner. This is not a particularly new idea, but perhaps the inventors have found a twist on it to refresh the concept.—JAH

7,120,261

43.38.Ar VEHICLE ACCESSORY MICROPHONE

Robert R. Turnbull *et al.*, assignors to Gentex Corporation
10 October 2006 (Class 381/86); filed 19 November 1999

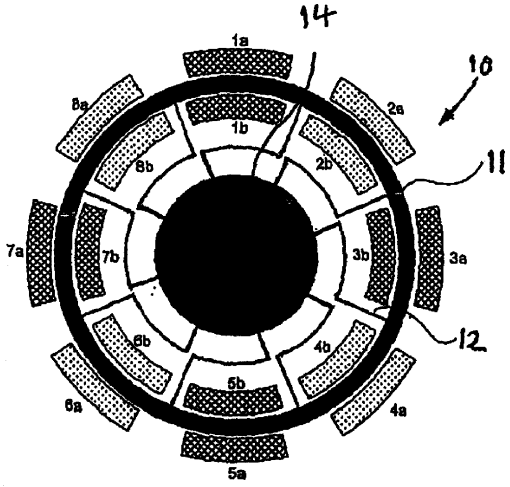
The patent describes in detail the construction and mounting of a bidirectional (figure-eight) pattern microphone into the rear view mirror assembly in an automobile. Improvements in directionality and freedom from feedback in two-way communications are cited.—JME

7,123,111

43.38.Bs MICRO-ELECTROMECHANICAL SYSTEMS

Kevin M. Brunson *et al.*, assignors to QinetiQ Limited
17 October 2006 (Class 331/116 M); filed in United Kingdom 20 March 2002

This vaguely worded patent describes in generalities the operation of an accelerometer or gyro (unspecified) that is based on a resonant ring of some unspecified material that is endowed with high mechanical Q simply



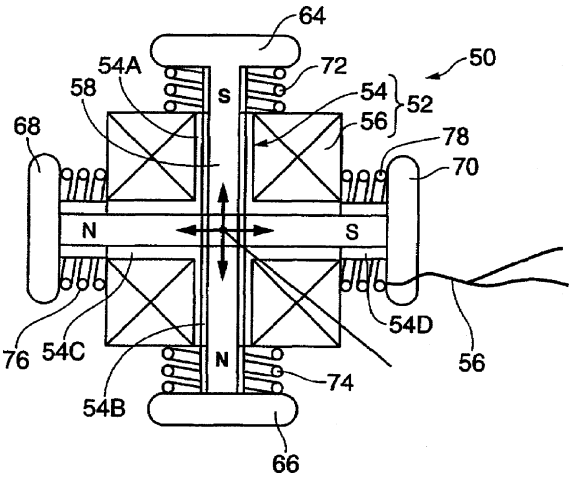
through the use of positive electrical feedback on the electrodes. It is not clear how the ring is suspended, but who cares when all of the losses can be overcome with feedback? Whatever is novel about this system is lost on this reviewer.—JAH

7,009,315

43.38.Dv APPARATUS FOR CONVERTING VIBRATION ENERGY INTO ELECTRIC POWER

Kesatoshi Takeuchi, assignor to Seiko Epson Corporation
7 March 2006 (Class 310/15); filed in Japan 20 April 2001

This is a fairly straightforward application of an electromagnetic generator in which permanent magnets are made to oscillate within a set of windings. A full-wave rectifier converts the resulting alternating current into



dc power. A pair of bar magnets crossing at a 90° angle reduces the need for the unit to be oriented with respect to the plane of the vibration. What is it here that is not obvious?—DLR

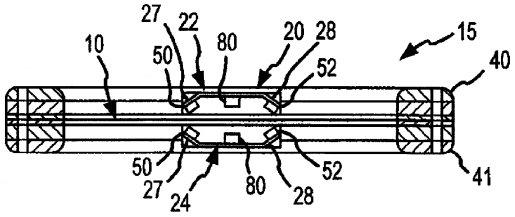
7,088,837

43.38.Dv HIGH EFFICIENCY PLANAR MAGNETIC TRANSDUCER WITH ANGLED MAGNET STRUCTURE

Chris Von Hellermann, Penang, Kedah, Malaysia and Dragoslav Colich, Costa Mesa, California

8 August 2006 (Class 381/191); filed 14 August 2003

A device is disclosed that uses angled magnet structures **22** and **24**, between which lies diaphragm **10**. The magnets can be disposed just on the angled sides of structures **22** and **24** as well as in the center of these



structures. The magnets can be stacked to increase the flux density. Among the listed benefits of this topology are operation "at a much lower frequency with suitable efficiency," although this reviewer would be concerned about the limited excursion of the diaphragm between structures **22** and **24**.—NAS

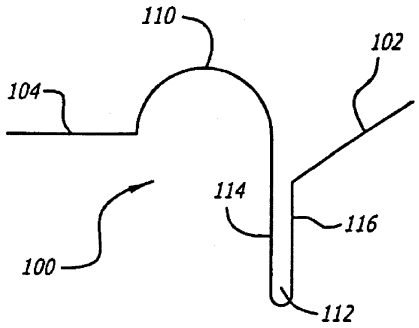
7,095,869

43.38.Dv LOUDSPEAKER COIL SUSPENSION SYSTEM

Douglas J. Button, assignor to Harman International Industries, Incorporated

22 August 2006 (Class 381/423); filed 31 August 2004

A unitary-suspension pocket attachment (USPA) **112** composed of a continuous layer of polymer can perform several functions in a compression driver—suspension, voice-coil former, a means of attachment to the driver



or cone, and voice-coil overdrive protection. The voice coil sits in pocket **112** between walls **114** and **116**, the cone can attach to **102**, and suspension **110** can be varied to change several operating parameters.—NAS

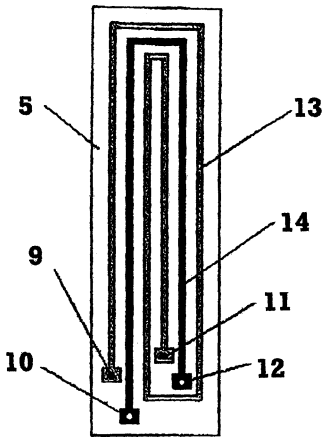
7,099,488

43.38.Dv PLANAR SPEAKER WIRING LAYOUT

Jack T. Bohlender, assignor to Wisdom Audio Corporation

29 August 2006 (Class 381/408); filed 3 May 2001

Many planar line-source transducers have multiple traces on the film diaphragm, but these may occupy different portions of the diaphragm **5**. By nestling traces **13** and **14** on the same area of the film **5** and applying signals



with different spectral content, the high-frequency response is said to improve, the "line source nature" of the driver is preserved, and the amplitude peaking that can occur in the midaudio frequencies can be better controlled.—NAS

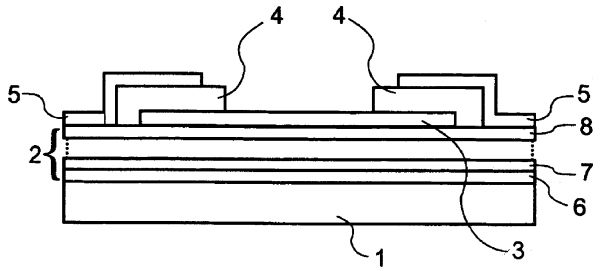
7,119,637

43.38.Fx FILTER SYSTEM COMPRISING A BULK ACOUSTIC WAVE RESONATOR

Hans Peter Loebl *et al.*, assignors to Koninklijke Philips Electronics N.V.

10 October 2006 (Class 333/187); filed in the European Patent Office 14 August 2001

The authors of this patent describe a way of isolating a bulk acoustic wave resonator from its mounting substrate by using a quarter-wave stack made of insulating films commonly used in semiconductor processing. They



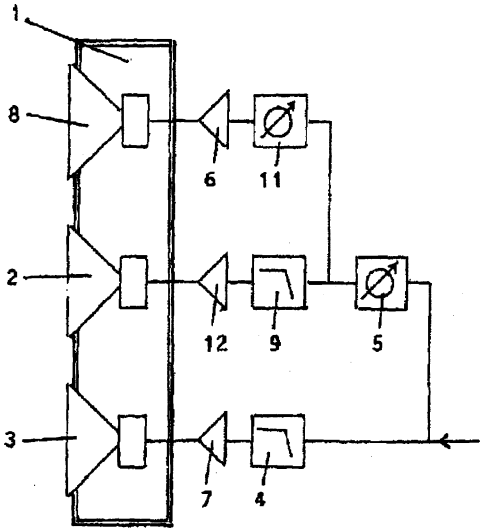
claim improved out-of-band performance of the quarter-wave stack through the use of composition-modulated mixtures of, e.g., TaO2 and SiO2. In this way, the band-rejection characteristics of the stack can be narrowed and (so it is claimed) the out-of-band attenuation is improved.—JAH

7,088,833

43.38.Hz MULTIPLE-SPEAKER

Martin Kling, Hannover, Germany
8 August 2006 (Class 381/97); filed in Germany 1 October 1999

The ancients, like Olson at RCA and many of Fletcher's colleagues at Bell Labs, would be interested to see that some of the techniques and



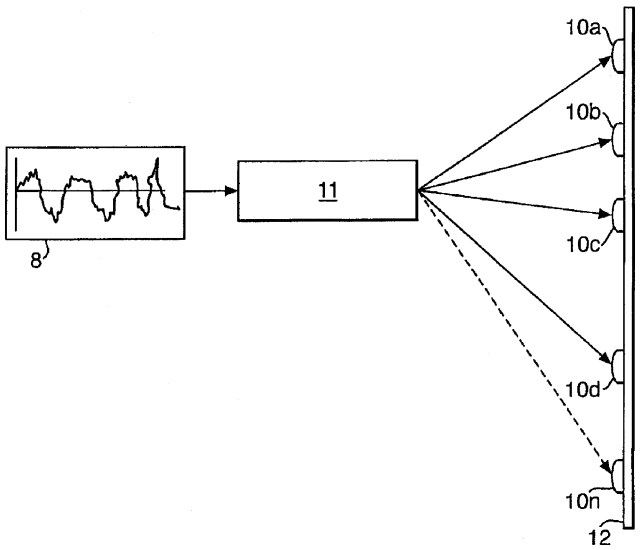
methods that they pioneered and implemented, i.e., phase array via **5** and **11**, frequency shaping via **4** and **9**, and amplitude shaping via **6**, **7**, and **12**, are now covered by a patent.—NAS

7,116,790

43.38.Hz LOUDSPEAKER

Kenneth H. Heron *et al.*, assignors to QinetiQ Limited
3 October 2006 (Class 381/59); filed in United Kingdom 6 February 2001

This panel-form loudspeaker is driven by multiple analog transducers **10a**, **10b**, etc. The goal is to provide greater dynamic range than conventional transducers can deliver. In any given frequency band, drive levels are controlled such that "[t]he number of drivers in operation at any one time is



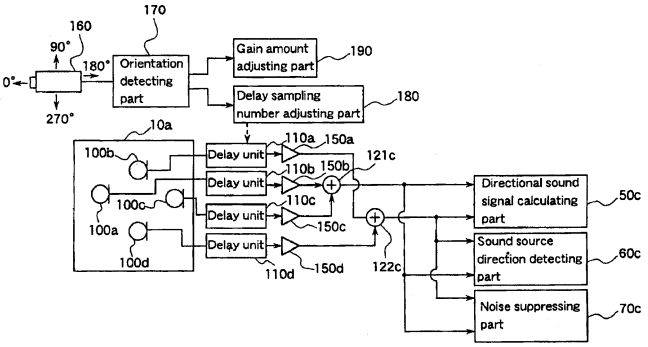
determined by the amplitude of the input signal to the control processor." According to the patent, "Prior art devices have suggested the use of more than one driver for a single loudspeaker, but none of them have recognized the need to control how these drivers interact to obtain the benefits of the present invention."—GLA

7,116,791

43.38.Hz MICROPHONE ARRAY SYSTEM

Naoshi Matsuo, assignor to Fujitsu Limited
3 October 2006 (Class 381/92); filed in Japan 2 July 1999

The patent describes a multi-element, first-order microphone array intended for use with a video camera, among other applications. Through signal delay, linear combination, and application of some degree of noise



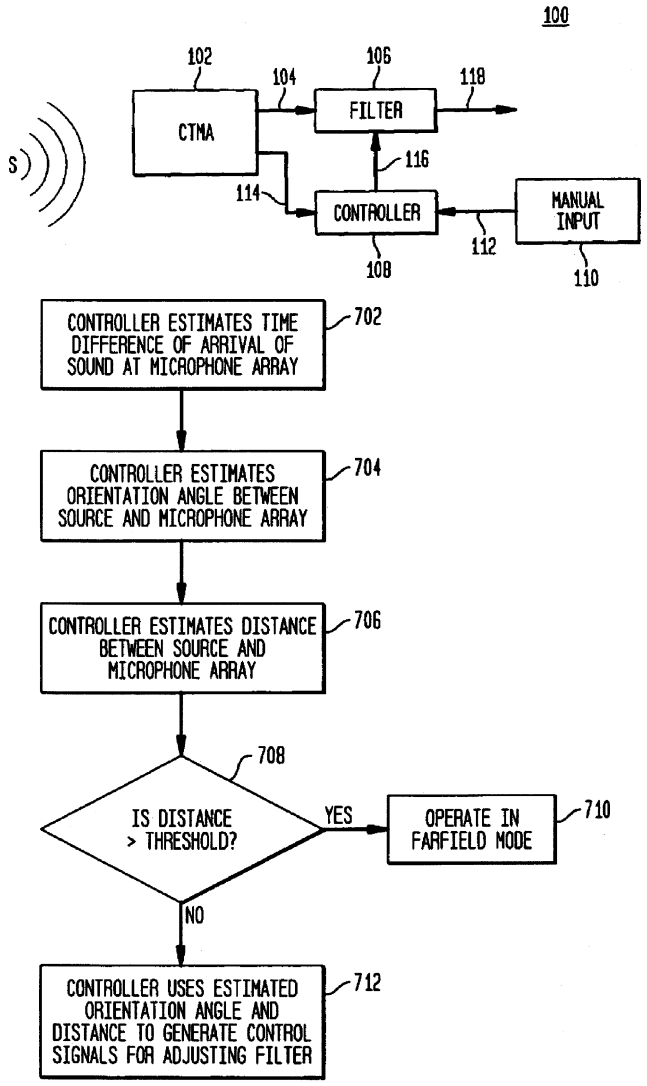
suppression, the microphone array senses the direction of sound sources in the median plane and adjusts the effective directional pattern to maximize sound pickup in those directions.—JME

7,123,727

43.38.Hz ADAPTIVE CLOSE-TALKING DIFFERENTIAL MICROPHONE ARRAY

Gary W. Elko and Heinz Teutsch, assignors to Agere Systems Incorporated
17 October 2006 (Class 381/92); filed 30 October 2001

One major use for gradient (differential array) microphones is in close-talking applications, where distant interfering sounds are substantially attenuated. The higher the order of the microphone, the greater will be low-frequency level fluctuations, which are highly dependent on the distance between the talker and the microphone. In some cases, a variation of a



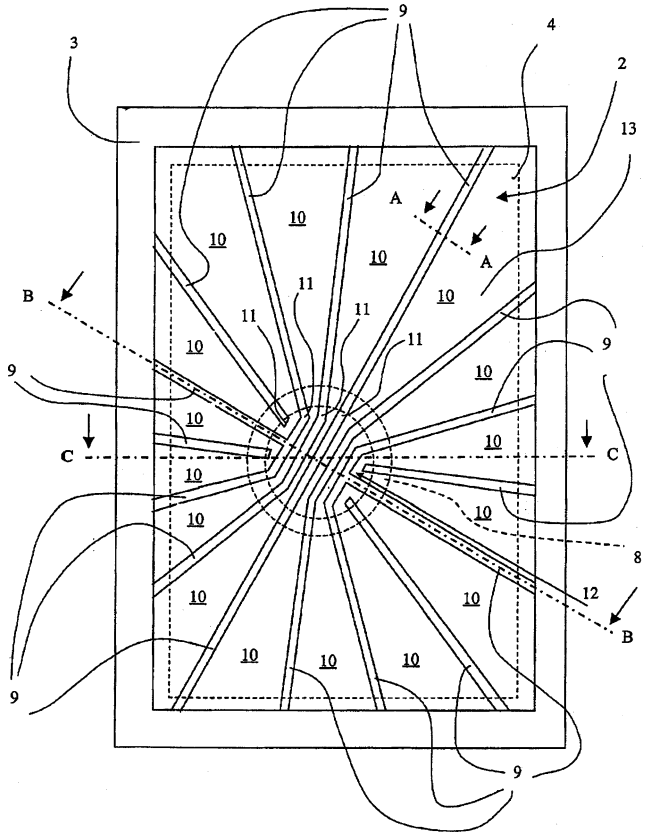
centimeter or two can produce a wide fluctuation in output level. The authors of this patent have come up with a first-order adaptive filter that can track the effect of the talker's change in position and/or bearing angle, and make the necessary level and equalization changes to ensure proper response.—JME

7,120,263

43.38Ja BENDING WAVE ACOUSTIC RADIATOR

Henry Azima *et al.*, assignors to New Transducers Limited
10 October 2006 (Class 381/152); filed in United Kingdom 23 March 2001

In this panel-form loudspeaker, a thin diaphragm is pleated to form integral stiffening ribs **9**. Two dissimilar pleated panels can be combined to form a more complex diaphragm. The process of designing and fabricating



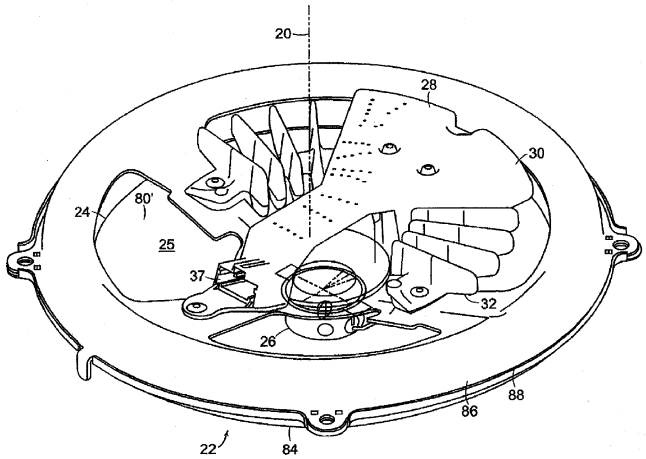
such a loudspeaker is described in considerable detail. The measured frequency response of one embodiment is reasonably smooth from about 150 Hz to 15 kHz.—GLA

7,120,270

43.38.Ja AUDIO DEVICE HEAT TRANSFERRING

Scott H. Aronson *et al.*, assignors to Bose Corporation
10 October 2006 (Class 381/397); filed 18 September 2002

The language employed in most Bose patents defies translation into simple English, and this patent is no exception. However, the invention itself is interesting. Prior art includes several examples of inside-out loudspeaker geometry in which the entire magnetic structure is positioned within the cone rather than behind it. It is also possible to utilize a conventional



magnetic structure behind the cone, but position the supporting frame in front of the cone for improved heat dissipation. The design shown here has a conventional basket-type frame behind the cone, but inverts the magnetic structure. Heat is transferred from the magnetic structure to a separate finned

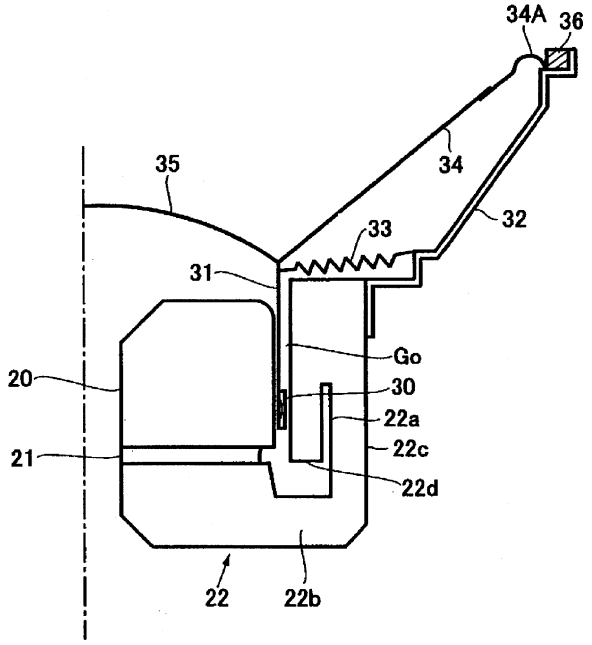
radiator **32**. The heat sink can also accommodate a built-in power amplifier.—GLA

7,120,271

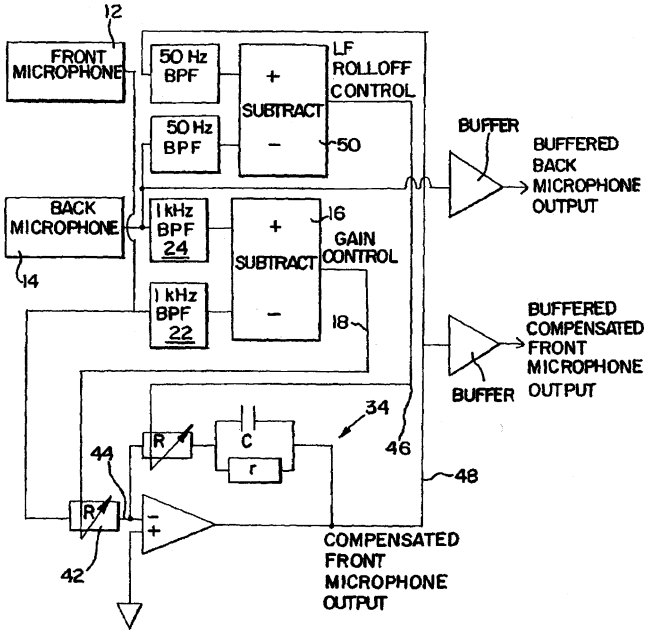
43.38.Ja INTERNAL MAGNETIC CIRCUIT AND LOUDSPEAKER SYSTEM INCORPORATING THE SAME

Hiroyuki Kobayashi *et al.*, assignors to Pioneer Corporation
10 October 2006 (Class 381/414); filed in Japan 19 June 2002

The magnetic gap of this long-excursion loudspeaker design is extended by an air space below the gap and an additional reverse slot **22a**.



Finite element modeling is used to show that the resulting magnetic field is uniform over the entire length of the gap.—GLA



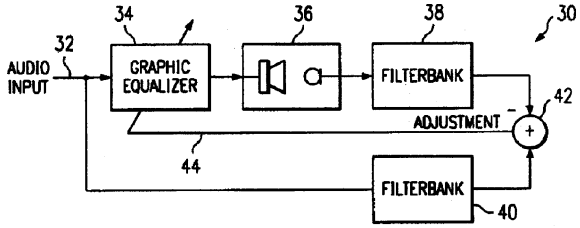
tion of both elements) is continuous and the output parameters of one of the elements can be adjusted to match the other so that accurate directional characteristics can be maintained over a wide frequency range. Useful measurements and design data are presented.—JME

7,092,537

43.38.Lc DIGITAL SELF-ADAPTING GRAPHIC EQUALIZER AND METHOD

Rustin W. Allred *et al.*, assignors to Texas Instruments Incorporated
15 August 2006 (Class 381/103); filed 28 September 2000

Graphic equalizer **34** is adjusted by signal **44** that is derived from the input **32** modified by filter bank **40** and the signal from the loudspeaker, microphone, and room that make up block **36**. Known nonlinearities in the



speaker and microphone can be addressed in filter bank **38**. Blocks **34**, **38**, and **40** all divide the spectrum of interest into N subbands, with the center frequencies being equidistant from one another.—NAS

7,113,604

43.38.Kb APPARATUS AND METHOD FOR MATCHING THE RESPONSE OF MICROPHONES IN MAGNITUDE AND PHASE

Stephen C. Thompson, assignor to Knowles Electronics, LLC
26 September 2006 (Class 381/92); filed 29 April 2003

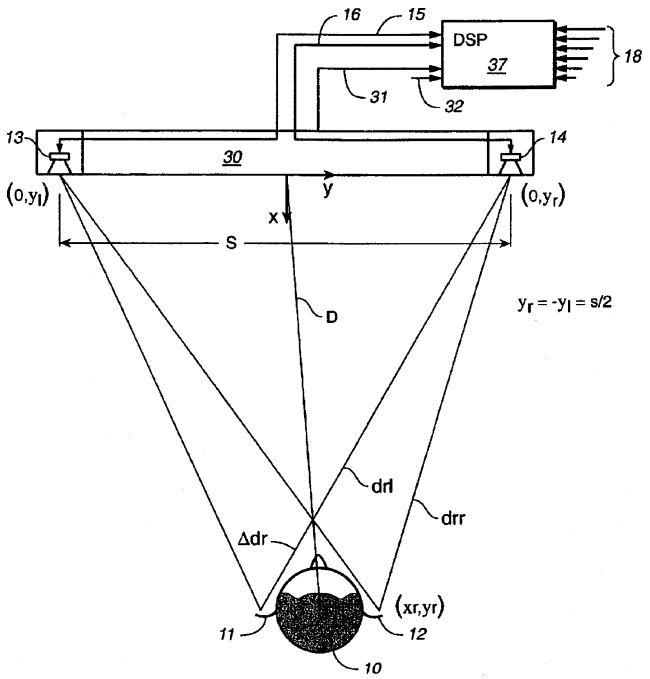
The patent relates primarily to hearing aid applications where fore-aft directionality is achieved by combining the outputs of a pair of omnidirectional elements operating as a dipole. The transition from omni (only a single element in operation) to maximum first-order directivity (a combina-

7,113,609

43.38.Lc VIRTUAL MULTICHANNEL SPEAKER SYSTEM

Michael I. Neidich *et al.*, assignors to Zoran Corporation
26 September 2006 (Class 381/305); filed 4 June 1999

Most synthetic surround-sound schemes generate virtual sound sources from only two loudspeakers by introducing head-related transfer functions and partially cancelling interaural crosstalk. The algorithms used depend on the angular locations of the loudspeakers in relation to the listener's ears. If one assumes that some installations will require less than optimum loud-



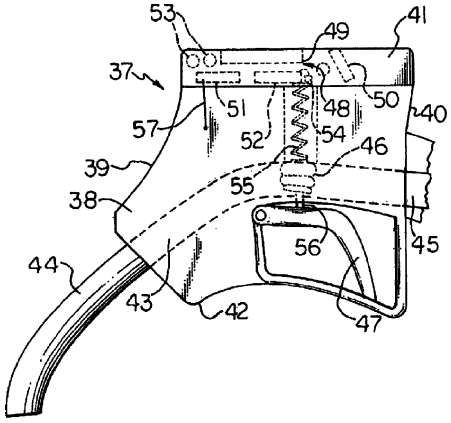
speaker separation, then some means should be provided to adjust the signal processing accordingly. (In fact, best results are obtained with narrow angular separation.) Well, suppose that left and right loudspeakers **13**, **14** are mounted in a single, telescoping cabinet **30**. When the cabinet is extended or compressed, an internal sensor relays the information to signal processor **37**. Stranger ideas have been patented and marketed.—GLA

6,957,673

43.38.Md FUEL DISPENSING DEVICE EQUIPPED WITH A SOUND AND/OR VIDEO SYSTEM

Alan S. Ambrose *et al.*, assignors to Advanced Information Systems, LLC
25 October 2005 (Class 141/94); filed 22 July 2003

Once again, entertainment during gasoline refueling is the concern. As shown, a speaker **50** in conjunction with a video screeen **49** can thrill the



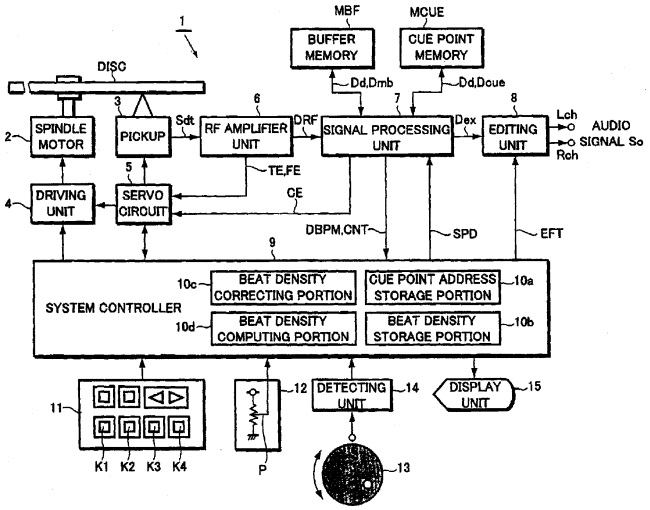
pump operator with the latest in exciting advertisements for bigger gas guzzling vehicles.—MK

6,967,905

43.38.Md INFORMATION PLAYBACK APPARATUS

Masahiko Miyashita *et al.*, assignors to Pioneer Corporation
22 November 2005 (Class 369/30.11); filed in Japan 19 October 2001

Magnetic tape recorder users will recall the ''jog dial'' for finding edit points. This disclosure illustrates how to implement the equivalent with a



DVD. More significantly, beat density computation can be measured and used to split sudden discontinuities.—MK

6,969,797

43.38.Md INTERFACE DEVICE TO COUPLE A MUSICAL INSTRUMENT TO A COMPUTING DEVICE TO ALLOW A USER TO PLAY A MUSICAL INSTRUMENT IN CONJUNCTION WITH A MULTIMEDIA PRESENTATION

John Brinkman *et al.*, assignors to Line 6, Incorporated
29 November 2005 (Class 84/625); filed 21 November 2001

Yet another web music-server delivery system; this time, this one includes MIDI, encryption, tube-amplifier simulation, and a graphical user interface. Individually, none of these is novel—does the combination make it novel and patent worthy? Caveat lector.—MK

6,990,453

43.38.Md SYSTEM AND METHODS FOR RECOGNIZING SOUND AND MUSIC SIGNALS IN HIGH NOISE AND DISTORTION

Avery Li-Chun Wang and Julius O. Smith III, assignors to Landmark Digital Services LLC
24 January 2006 (Class 704/270); filed 20 April 2001

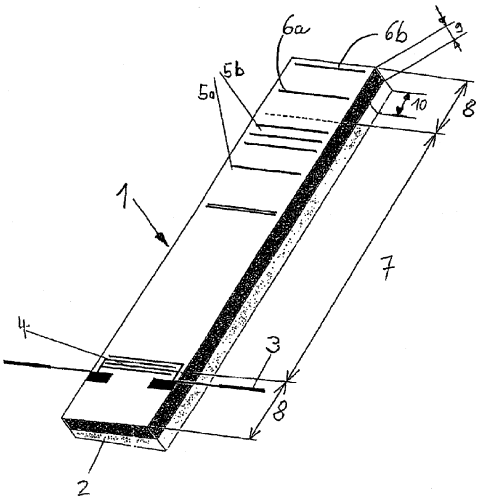
The human auditory system does a remarkable job of recognizing sounds under noisy and distorted conditions. How then can a digital method identify sounds found in these circumstances? The inventors propose using spectrally significant ''landmarks'' to identify sounds. Examples of landmarks include ceptral coefficients, LPC coefficents, and ''slice fingerprints.'' They propose using five to ten landmarks per second. These landmarks can be hashed together for efficient searching as well.—MK

7,109,632

43.38.Rh SURFACE WAVE SENSOR

Henri van Knokke, assignor to FAG Kugelfischer AG
19 September 2006 (Class 310/313 D); filed in Germany 14 March 2002

SAW device **7** is cemented **2** to a body whose expansion or contraction is to be measured (e.g., for a strain gage). Body temperature, needed for



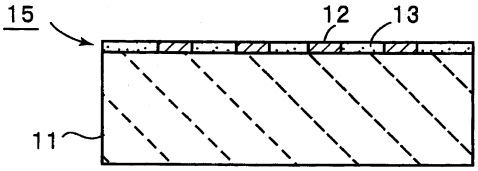
accurate expansion measurement, is sensed by additional SAW length **8** with electrodes **6a** and **6b** over noncemented region **10**.—AJC

7,109,634

43.38.Rh END SURFACE REFLECTION TYPE SURFACE ACOUSTIC WAVE DEVICE

Takeshi Nakao *et al.*, assignors to Murata Manufacturing Company, Limited
19 September 2006 (Class 310/313 R); filed in Japan 20 January 2003

To minimize the temperature change of insulated SAW filter properties, the density of insulation film **13** is chosen to be about 1.5 times that of electrode material **12**. The insulating film is placed between electrodes and



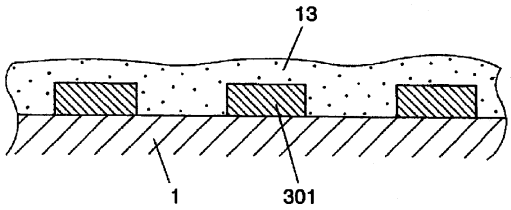
its thickness is adjusted to be about 0.15 to 0.4 that of the SAW wavelength. The temperature coefficient then becomes about half or less of that without film **13**.—AJC

7,109,828

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE, AND MOBILE COMMUNICATION DEVICE AND SENSOR BOTH USING SAME

Ryoichi Takayama *et al.*, assignors to Matsushita Electric Industrial Company, Limited
19 September 2006 (Class 333/193); filed in Japan 15 April 2002

A leaky wave mode is used to reduce the SAW device size and to



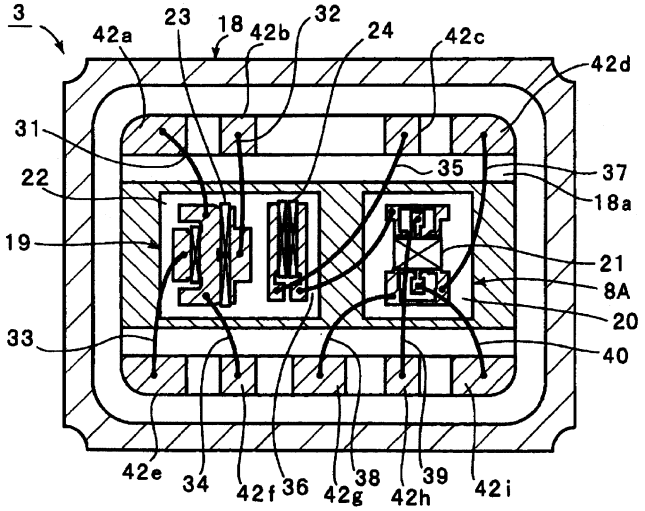
allow operation at higher frequencies. This is accomplished with a dielectric film **13** over the substrate **1** and electrodes **301**.—AJC

7,119,634

43.38.Rh SURFACE ACOUSTIC WAVE DEMULTIPLEXER USING DIFFERENT PIEZOELECTRIC MATERIALS FOR SUBSTRATES OF TWO SAW FILTERS

Kiwamu Sakano and Ryoichi Omote, assignors to Murata Manufacturing Company, Limited
10 October 2006 (Class 333/133); filed in Japan 25 July 2003

Multiplexers for cell phones require simultaneous operation at two different SAW frequencies. The optimum material for filtering differs with rf



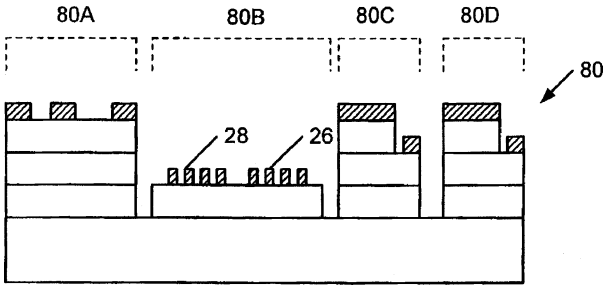
frequency. A packaging system is claimed that accommodates both lithium niobate **22** and lithium tantalate **20** substrates.—AJC

7,112,860

43.38.Rh INTEGRATED NITRIDE-BASED ACOUSTIC WAVE DEVICES AND METHODS OF FABRICATING INTEGRATED NITRIDE-BASED ACOUSTIC WAVE DEVICES

Adam William Saxler, assignor to Cree, Incorporated
26 September 2006 (Class 257/416); filed 3 March 2003

A variety of semiconductor devices can be mounted in a single package using substrate materials that are nitrides of periodic-table group III. These devices include HEMT, HFET, MSFET, JFET, MOSFET, photo



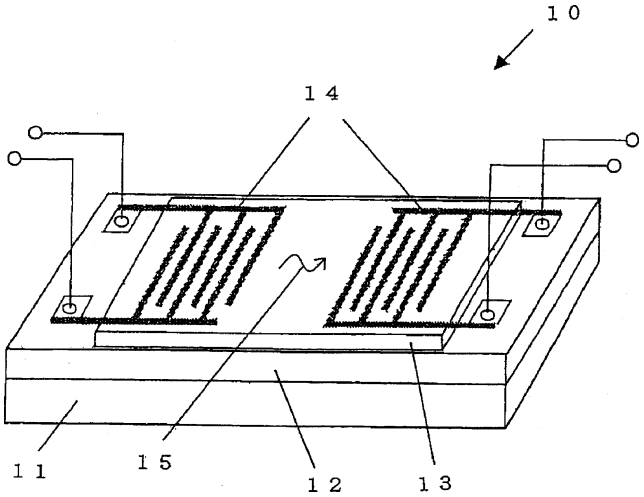
diodes, and LED devices, along with SAW devices and filters. The author claims layering, insulation, and trenching means to allow a plurality of these devices to operate on a single substrate.—AJC

7,122,938

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Hitoshi Noguchi and Yoshihiro Kubota, assignors to Shin-Etsu Chemical Company, Limited
17 October 2006 (Class 310/313 A); filed in Japan 12 May 2003

Formation of a diamond surface acoustic wave (SAW) substrate **12** (wave speed over 10 000 m/s) up to 20 μm thick is feasible by chemical vapor deposition (CVD). Piezoelectric film **13** and interdigital fingers **14** form a common SAW device. But diamond has a very high resistivity,



allowing charge buildup that can result in arc-over that fractures the diamond film **12**, causing loss in production. The author claims a procedure to add a boron or phosphorous dopant during CVD that maintains the diamond resistivity to less than 1014 Ω-cm, thereby avoiding arc-over.—AJC

7,126,251

43.38.Rh INTERFACE ACOUSTIC WAVE DEVICE MADE OF LITHIUM TANTALATE

Marc Solal *et al.*, assignors to Thales
24 October 2006 (Class 310/313 R); filed in France 19 March 2002

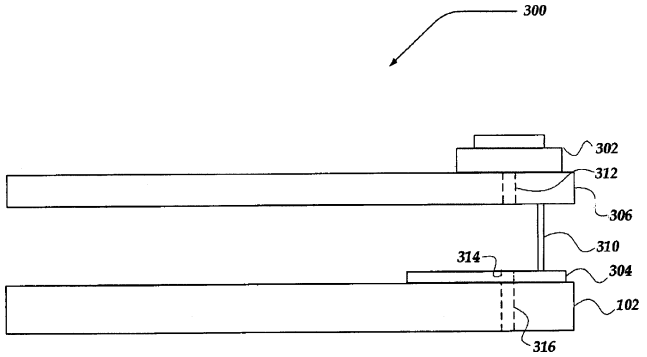
This patent discloses the use of Stoneley waves in a surface acoustic wave (SAW)-type device that can be used for filters or resonators. The authors assert the benefits to be the lack of need for hermetic encapsulation and the precise control of the surface acoustic wave velocity and attenuation resulting from the variations in cap material. Their patent focuses on the use of lithium tantalate and/or lithium niobate as piezoelectric materials. The writing is clear and concise.—JAH

7,123,734

43.38.Si ANTENNA AND SPEAKER CONFIGURATION FOR A MOBILE DEVICE

David William Voth and Dina Christine Taylor, assignors to Microsoft Corporation
17 October 2006 (Class 381/334); filed 11 April 2003

In a miniature wireless communication device such as a cellular telephone, the stray-magnetic field from the loudspeaker can interfere with the antenna if the two are located in close proximity. This patent describes an improved geometry in which sound from speaker **302** passes through



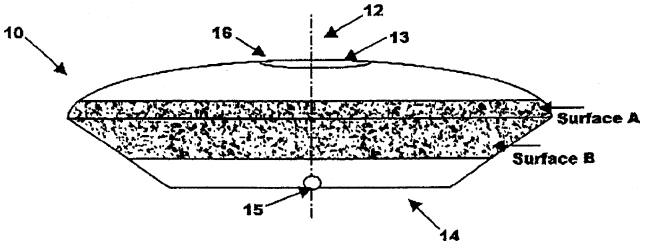
perforations in circuit board **306** and may also pass through perforations in antenna **304**, thus achieving maximum spacing between the two devices. The goal is to achieve acceptable audio quality with minimal rf interference.—GLA

7,123,735

43.38.Si METHOD AND APPARATUS TO INCREASE ACOUSTIC SEPARATION

James G. Ryan and Michael R. Stinson, assignors to National Research Council of Canada
17 October 2006 (Class 381/346); filed 12 September 2001

A portable, self-contained audio teleconferencing system is often housed in a saucer-shaped enclosure that can be placed on a conference table. Loudspeaker **13** is recessed into the top of the enclosure and one or more microphones **15** are located near the bottom. For a device of moderate size, acoustical isolation between the loudspeaker and microphone is poor at



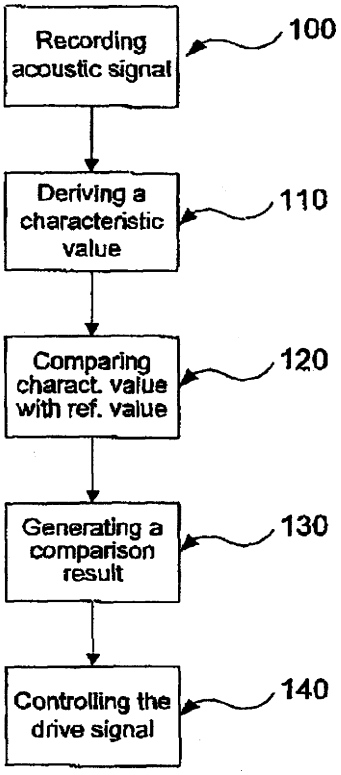
voice frequencies because sound simply diffracts around the edge of the saucer. This patent suggests that the introduction of damped reactive acoustical impedances—designated as **Surface A** and **Surface B**—can greatly improve sound isolation over a substantial frequency range. The idea is clever and is clearly explained in the patent.—GLA

7,123,948

43.38.Si MICROPHONE AIDED VIBRATOR TUNING

Claus Peter Nielsen, assignor to Nokia Corporation
17 October 2006 (Class 455/567); filed 16 July 2002

In many cellular telephones the tactile alert signal is provided by a motorized vibrator. It is important to maintain an optimum vibrator frequency, yet the speed of the vibrator motor is established by the drive level.



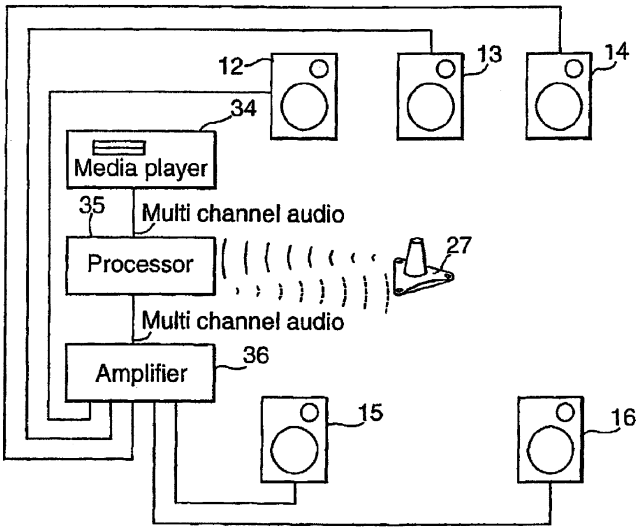
Practical manufacturing tolerances plus drift over time can result in unacceptable variations. Various feedback controls have been proposed but most of these involve additional complexity and cost. This patent sets forth a calibration routine that uses the existing microphone as a sensor. No additional space is required and the cost of the software is minimal.—GLA

7,123,731

43.38.Vk SYSTEM AND METHOD FOR OPTIMIZATION OF THREE-DIMENSIONAL AUDIO

Yuval Cohen *et al.*, assignors to BE4 Limited
17 October 2006 (Class 381/303); filed in Israel 9 March 2000

There have been numerous reviews in these pages of patents that deal with automatic calibration of five-channel surround systems for listeners not on-axis. Most of them have been similar in technique, but this one takes the task to new heights (literally and figuratively). The analysis system uses an array of four microphones and, as such, can sense direction in three dimensions. This enables the system to make adjustments for loudspeakers that may not all be positioned in the same horizontal plane. Otherwise, the system appears to make all requisite adjustments in playback level, equalization, and relative timing.—JME

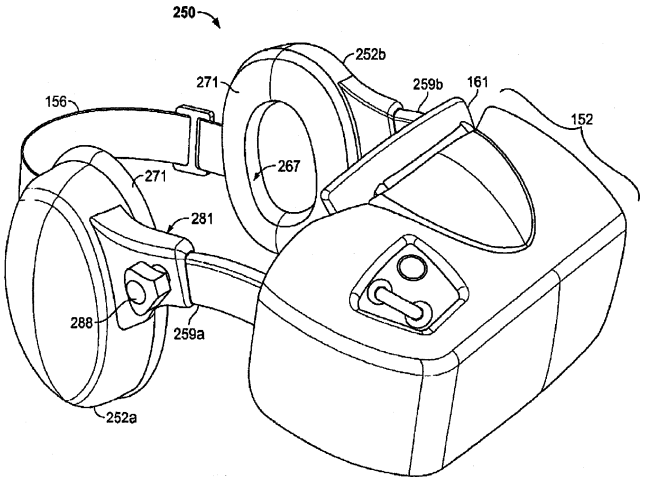


7,124,425

43.38.Vk AUDIO/VIDEO SYSTEM AND METHOD UTILIZING A HEAD MOUNTED APPARATUS WITH NOISE ATTENUATION

Tazwell L. Anderson, Jr. and Mark A. Wood, assignors to Immersion Entertainment, L.L.C.
17 October 2006 (Class 725/68); filed 31 August 1999

The inventor has previously described a head-mounted audio-video package that would allow spectators at sports events to act as their own television directors, making moment-by-moment selections from a variety of camera/narration feeds. Since the presence of surrounding sports fans might be distracting, the device's headphones have now been replaced by larger, sealed-muff units to attenuate external noise. This change appears to be the sole novel feature of the new patent. Exactly why the sports fan in question is sitting in the bleachers rather than in front of his home television set is not explained.—GLA

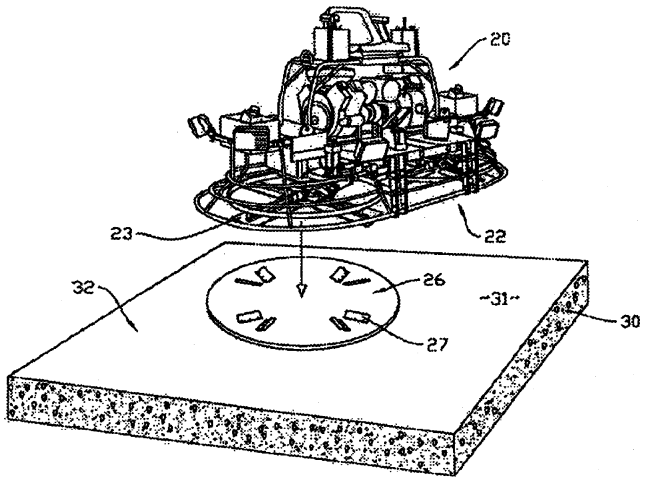


7,108,449

43.40.Ga METHOD AND APPARATUS FOR ACOUSTICALLY MATCHED SLIP FORM CONCRETE APPLICATION

J. Dewayne Allen and Richard P. Bishop, assignors to Allen Engineering Corporation
19 September 2006 (Class 404/75); filed 21 December 2004

In finishing the surface of poured concrete, floats of wood or magnesium bring up water and fines. But steel floats seal the surface and trap water and fines, making the concrete vulnerable to delamination in service. In this patent, the rotating blades of power troweler **20** rotate pan **26** to flatten the



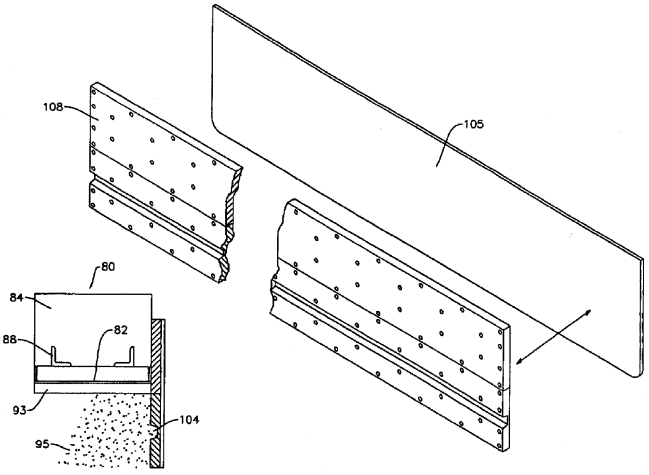
surface of **30** and to produce vibrations that purge free water and air bubbles entrained in **30**. When tilted slightly, pan **26** produces translation motions of **20**. The author argues that improved purging of water and fines results when a polyethylene disc is placed on the bottom of pan **26** to provide an acoustical impedance match to the green concrete value of 2.5 μNs/m^3.—AJC

7,114,876

43.40.Ga ACOUSTICALLY MATCHED CONCRETE FINISHING PANS

J. Dewayne Allen and Richard P. Bishop, assignors to Allen Engineering Corporation
3 October 2006 (Class 404/112); filed 21 December 2004

This patent relates to United States Patent 7,108,449, reviewed above, regarding power finishing of freshly poured concrete. Here, the vibration-



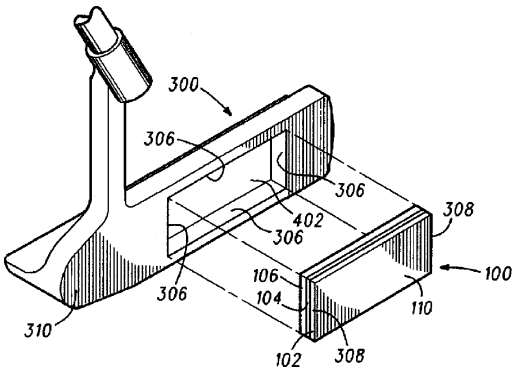
impedance matching material **108** is improved to be ultra-high molecular weight (UHMW) polyethylene, to also be applied on power-trowel vertical surfaces **105**.—AJC

7,086,961

43.40.Kd METHODS AND APPARATUS FOR USING A FREQUENCY-SELECTABLE INSERT IN A GOLF CLUB HEAD

David E. Wright and Eric V. Cole, assignors to Karsten Manufacturing Corporation
8 August 2006 (Class 473/329); filed 20 May 2002

Insert **100**, which consists of a plate **102** connected to a mass **106** via a damper **104**, is set in a ''coplanar'' way in club head **300**, which appears to



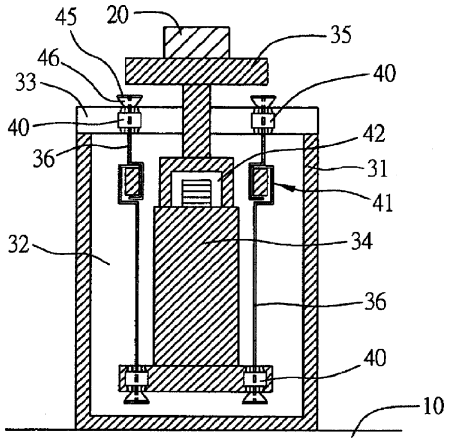
be a putter. The insert can be configured to have different responses so that the player can subjectively select the insert that provides the most pleasant sound when the ball is struck by striking surface **110** or the best ''feel'' of the vibrations from said striking, among other bio-mechanical responses.—NAS

7,114,692

43.40.Tm VIBRATION ISOLATION DEVICE

Yann-Shuoh Sun *et al.*, assignors to Industrial Technology Research Institute
3 October 2006 (Class 248/550); filed in Taiwan, 22 December 2004

This device, intended for isolation of high-technology equipment, employs pendulum action for horizontal isolation in conjunction with active control for vertical isolation. A mass **34** is suspended in a cylindrical housing **31** by a series of circumferentially spaced rods **36** that are connected to a cover **33** and to the bottom of the mass via ball bearings **40**. Piezoelectric



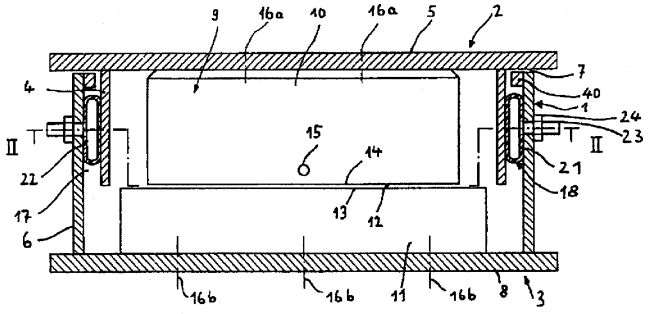
actuators **41** that are built into the support rods are driven via a controller in response to signals from a sensor **42**. The space between the mass and the housing may be filled with oil or the like to provide damping. A platform **35** that is rigidly attached to the top of mass **34** supports the isolated item **20**.—EEU

7,114,710

43.40.Tm PNEUMATIC VIBRATION ISOLATOR

Ulf Jörgen Motz, assignor to Bilz Schwingungstechnix GmbH
3 October 2006 (Class 267/123); filed 3 September 2004

The natural frequency provided by this isolator for sensitive equipment, such as that used in the semiconductor industry, may be adjusted remotely while the isolated item's position in the horizontal plane is restricted. The figure shows a diametral section through such an isolator, which consists of upper and lower cylindrical elements. The upper part **2**, which carries the isolated item, is rigidly attached to the upper element **9** of an air bearing. The lower part **11** of the air bearing rests on base plate **8** and



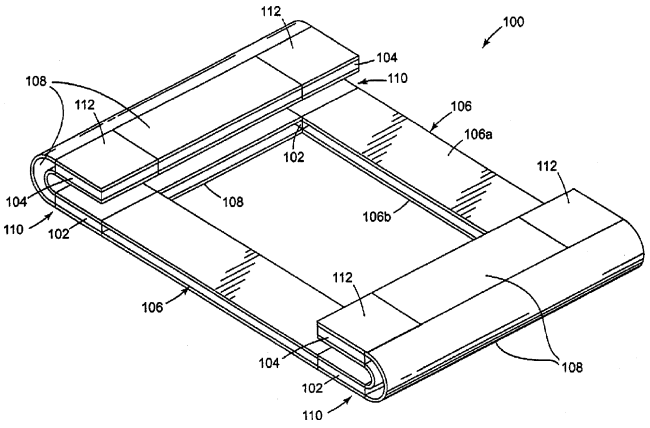
has a smooth upper surface **13**. The lower surface **14** of air-bearing element **9** is provided with channels through which air is fed from a pneumatic connection **15** into the gap **12**. Pneumatic tubes **18**, each extending around roughly one-quarter of the circumference, provide radial isolation and position control, with a ring **40** limiting the radial excursion. Control of the pressures via external valves permits adjustment of the axial and radial natural frequencies.—EEU

7,114,711

43.40.Tm SMART ISOLATION MOUNTS WITH CONTINUOUS STRUCTURAL ELEMENTS FEATURING VIBRATION ENERGY MANAGEMENT

Daryoush Allaei, assignor to Quality Research, Development & Consulting, Incorporated
3 October 2006 (Class 267/136); filed 20 January 2004

The concept underlying this patent involves diverting vibrations from an input to elements that confine the vibrations and prevent these from reaching the item to be protected. This concept is described in United States Patent 6,032,552, ''Vibration control by confinement of vibration energy'' and in United States Patent 6,116,389, ''Apparatus and method for confinement and damping of vibration energy.'' In the embodiment shown in the figure, the item to be protected is supported on the upper four corner regions



112 and the whole arrangement is attached to a vibration-generating structure at the four lower corner regions **110**. The four vibration diverters **102** and **104** (one in each corner) couple the vibration-generating structure and the item to be protected and are connected to each other by vibration confiners **106** and **108**. Each of confiners **106** consists of parallel plates with viscoelastic material between them. Confiners **108** are similar, but essentially C-shaped. The diverters **102** and **104** may be passive elements including plates, ribs, notched plates, etc., selected to divert energy in specific frequency ranges, or they may consist of active elements.—EEU

7,126,257

43.40.Tm PIEZOELECTRIC CERAMIC-REINFORCED METAL MATRIX COMPOSITES

Stephen L. Kampe *et al.*, assignors to Virginia Tech Intellectual Properties, Incorporated
24 October 2006 (Class 310/327); filed 21 May 2004

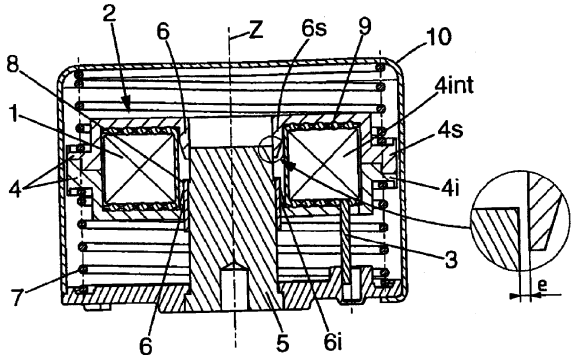
Composite materials capable of providing vibration damping are formed by dispersing piezoelectric particulates in a metal matrix. The particulates generate a voltage when they are subjected to strain as the composite is deformed. The resulting electrical energy is converted into heat in the surrounding metal matrix, thus dissipating mechanical energy and providing damping. The damping increases with the volume fraction of the particulates. Data presented in the patent document show that loss factors approaching 0.1 may be obtained at low frequencies at temperatures up to the Curie temperature of the piezoelectric material.—EEU

7,113,064

43.40.Vn ACTIVE DEVICE FOR DAMPING THE VIBRATIONS OF A VIBRATING ELEMENT

Patrice Loubat and Jérôme Joly, assignors to Hutchinson
26 September 2006 (Class 335/220); filed in France 1 October 2003

The authors claim a ''beater'' mass **1** for active vibration control that is driven along axis **Z** by an electric current via coil **9**. This beater does not



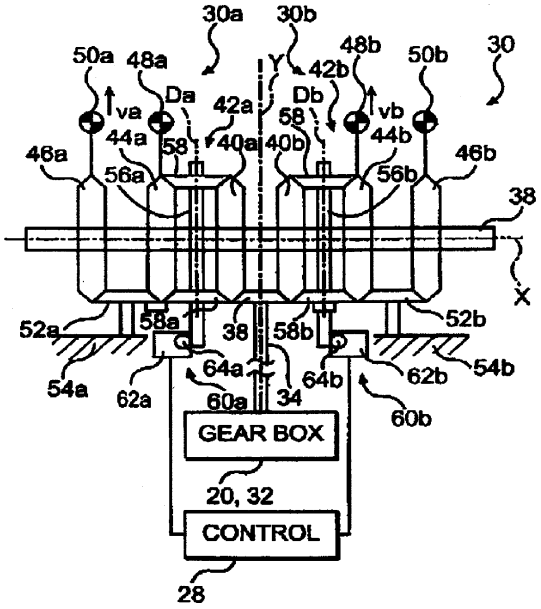
require a permanent magnet that could be affected by high heat near an engine.—AJC

7,118,328

43.40.Vn GEARBOX MOUNTED FORCE GENERATOR

William A. Welsh *et al.*, assignors to Sikorsky Aircraft Corporation
10 October 2006 (Class 415/170 R); filed 12 March 2004

An active vibration control (AVC) system is claimed where one or more gearbox mounted force generators (GMFGs) **30** are attached at suitable locations **54** on motor transmission **20** of an operating helicopter. Vibration sensors mounted at vibration sensitive positions (cabin, etc.) provide an error signal to control **28** for AVC. CFMG **30** comprises two rotating



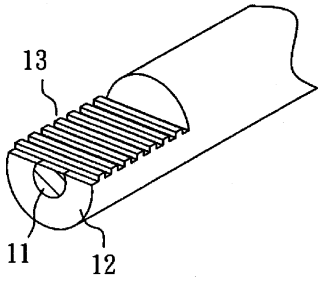
mirror image masses **30a** and **30b**. Control **28** operates motors **62a,b** and worms **64a,b** that turn differential shafts **58** and **56**. That is, differential assembly **42a,b** selectively advances or retards **44a,b** and **46a,b** relative to output shaft **34**, and likewise for **42** and **44**, proving a continuous variation of control-mass vibration amplitude and phase to effect AVC.—AJC

7,110,626

43.40.Yq FIBER OPTIC VIBRATION SENSOR

Woo-Hu Tsai, assignor to Tatung Company, Limited
19 September 2006 (Class 385/12); filed in Taiwan 21 November 2003

A very small vibration sensor is claimed in the form of optical fiber **11** whose cladding **12** is ground away on one side and roughened into a slotted plane **13** where effervescent light waves are exposed. An object vibrating



near this plane will modulate the amplitude of the waves back-scattered from this roughened area. The rate and amplitude of modulation of this back-scattered wave are a good measure of the vibration of the proximate body.—AJC

7,116,035

43.40.Yq SOUND/VIBRATION RESONANCE SEPARATING DEVICE

Yo Sugawara, Hokkaido, Japan
3 October 2006 (Class 310/322); filed 1 February 2002

This device is intended for visualization of the frequency components present in a sound or vibration. It consists in essence of a series of cantilever elements that are attached to a vibrating base. In one embodiment the base consists of a rod that is driven vertically and the cantilever elements consist of piano wires of various lengths that extend horizontally from the rod and may be distributed around the rod in esthetically pleasing arrangements. The rod may be supported on a vibrator that is driven in response to a signal received by a microphone.—EEU

7,116,426

43.40.Yq MULTI-BEAM HETERODYNE LASER DOPPLER VIBROMETER

Amit K. Lai *et al.*, assignors to MetroLaser
3 October 2006 (Class 356/486); filed 31 October 2005

A beam of coherent light is split into an object beam and a reference beam. The object beam is divided into a number of separate beams that are directed at locations on the object under investigation. The reference beam is frequency-shifted and split into a like number of frequency-shifted reference beams. The object beams reflected from the object are mixed with the reference beams, resulting in multiple beam pairs, each of which is focused onto a photodetector whose output may be processed to determine the vibration characteristics of the object. The system enables simultaneous measurements at multiple points down to low frequencies, with a wide dynamic range and high signal-to-noise ratio.—EEU

7,124,637

43.40.Yq DETERMINING AMPLITUDE LIMITS FOR VIBRATION SPECTRA

Ashish Singhal *et al.*, assignors to Johnson Controls Technology Company
24 October 2006 (Class 73/659); filed 22 March 2004

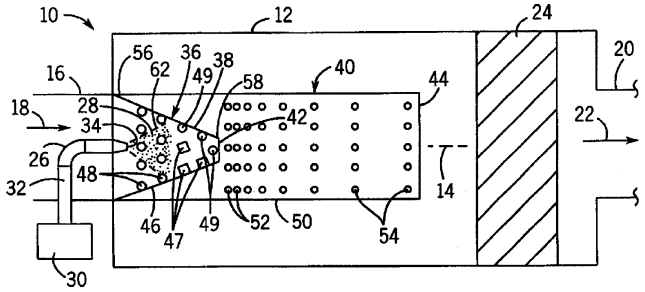
This patent relates generally to vibration monitoring of mechanical equipment for the purpose of detecting faults. Many monitoring approaches involve measurement of frequency spectra and noting when spectrum components exceed prescribed limits, limits that generally are determined from experience and rules-of-thumb. The present patent delineates means for determining suitable limits from historical data and statistical probability distributions determined from measurements.—EEU

6,722,123

43.50.Gf EXHAUST AFTERTREATMENT DEVICE, INCLUDING CHEMICAL MIXING AND ACOUSTIC EFFECTS

Z. Gerald Liu *et al.*, assignors to Fleet guard, Incorporated
20 April 2004 (Class 60/286); filed 27 February 2002

Reduction of chemical emissions from internal combustion engines may require the introduction of additives into the exhaust stream. An exhaust muffler that incorporates aftertreatment devices for chemical emissions reduction consists of chemical injector **26**, which introduces chemical species (e.g., aqueous urea solution) into the exhaust flow. A conical "tur-



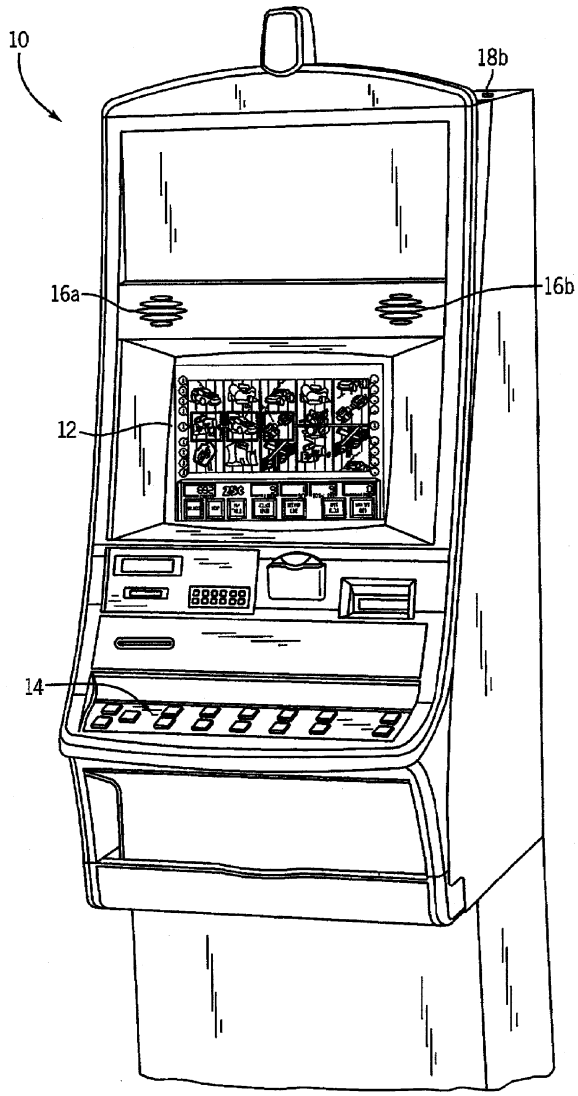
bulator" **36** is designed to enhance the mixing of the chemical with the exhaust gases. Acoustical attenuation is provided by perforated tube **40**. Performance of this design is enhanced by ensuring that the perforation hole sizes are greater than 1/4 in. in the turbulator and smaller than 1/4 in. in the acoustic element.—KPS

7,112,139

43.50.Ki GAMING MACHINE WITH AMBIENT NOISE ATTENUATION

Francisco Jose Paz Barahona and Timothy C. Loose, assignors to WMS Gaming Incorporated
26 September 2006 (Cluss 463/35);, filed 19 December 2001

The author claims ambient noise reduction for gaming machine **10** where the player head position is relatively near the display, the touch screen **12**, and entertainment/active-noise-cancellation (ANC) speakers **16a-16b**.



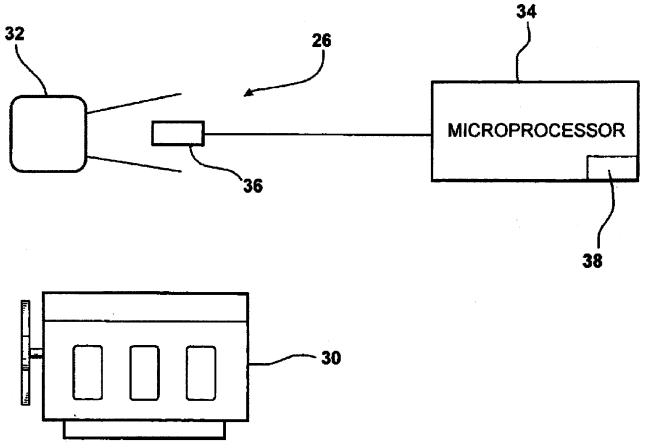
Entertainment and intrigue are enhanced by ANC rejection of ambient sounds from other nearby gaming machines while effecting full volume of the player's own game action sounds.—AJC

7,106,867

43.50.Ki ACTIVE NOISE CANCELLATION FOR A VEHICLE INDUCTION SYSTEM HAVING SELECTABLE ENGINE NOISE PROFILE

Paul D. Daly, assignor to Siemens VDO Automotive Incorporated
12 September 2006 (Class 381/71.4); filed 6 March 2002

An active noise canceller (ANC) for induction noise **26** comprises speaker/microphone **36** and processor **34**. The vehicle driver can select ANC



sound parameters for engine **30**, air induction system **32**, or controller **34** via communications **38** (driver controls or cell phone) to suit his/her individual taste. Communications **38** may be expanded to record engine sound profiles preferred by the driver and driver demographics, to adjust subsequent products to meet customer demands.—AJC

7,113,850

43.50.Ki METHOD AND APPARATUS FOR ACTIVE ACOUSTIC DAMPING MOTOR CONTROL

Robert J. Atmur, assignor to The Boeing Company
26 September 2006 (Class 701/1); filed 3 December 2003

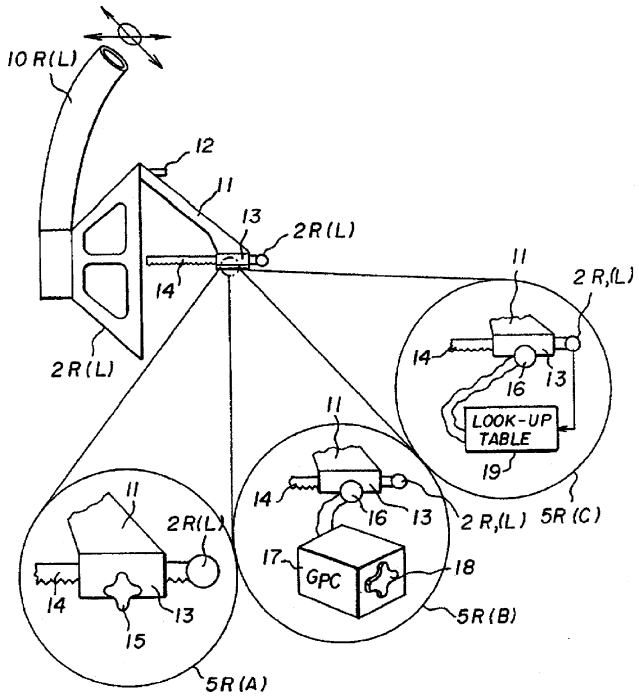
Active control of the noise of an electric motor, such as a multiphase brushless dc motor, is obtained without the addition of actuators. A transfer function between motor operating modes and noise is measured and used to develop control signals that are applied to one or more phases of the motor so as to produce actuating forces that act directly on the motor. ''Precorrections'' may be applied to the motor phase currents to reduce the noise signature of a vehicle in which the motor is mounted.—EEU

7,110,551

43.50.Ki ADAPTIVE PERSONAL ACTIVE NOISE REDUCTION SYSTEM

William Richard Saunders and Michael Allen Vaudrey, assignors to Adaptive Technologies, Incorporated
19 September 2006 (Class 381/71.6); filed 27 March 2000

A wide-frequency-range active noise reduction scheme for headsets with a greater degree of cancellation is claimed. The position of error-sensing microphone **2R(L)** (left side shown) on support **10-11-13** is adjusted



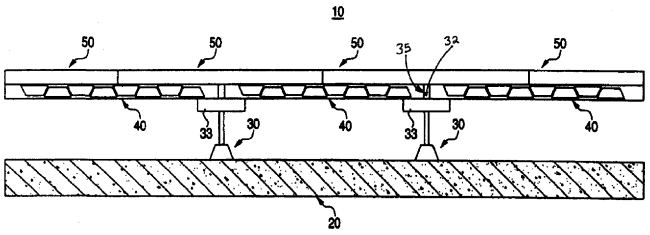
by rack **14** and pinion **15** to be very near the ear canal opening. Adjustment is either manual via knob **15** or automatic with a look-up table for optimal position and feed-forward control.—AJC

7,114,302

43.55.Ti FLOOR STRUCTURE AND FLOOR BASE PANEL

Rento Tanase *et al.*, assignors to Yamaha Corporation
3 October 2006 (Class 52/220.1); filed in Japan 6 March 2002

The floor base panel for a raised floor system has sections of different rigidity: more rigid at the point of contact with the supporting leg and less



rigid (by virtue of cavities in the panel) between the supports. The system claims to reduce the transfer of vibration energy (footfall?) into the structure.—CJR

7,003,093

43.60.Bf TONE DETECTION FOR INTEGRATED TELECOMMUNICATIONS PROCESSING

Raghavendra S. Prabhu *et al.*, assignors to Intel Corporation
21 February 2006 (Class 379/390.02); filed 23 August 2001

This fairly elaborate tone/voice discriminator would detect signal amplitudes at many discrete frequencies and then consult a tone dictionary to distinguish silence or any of a variety of tones, such as Touch Tone™ (DTMF), ring tones, fast or slow busy tones, etc. Tone presence is detected

Exemplary Filter Coefficients for Goertze Filter

frequency	cos(2* pi*f1/fs)	frequency index
350	31536	0
400	31163	1
425	30958	2
440	30829	3
480	30465	4
540	29863	5
600	29195	6
620	28958	7
660	28462	8
697	27978	9
700	27938	10
770	26955	11
780	26808	12
852	25700	13
900	24916	14
941	24218	15
1020	22802	16
1100	21280	17
1140	20487	18
1209	19072	19
1300	17120	20
1336	16324	21
1380	15332	22
1477	13084	23
1500	12539	24
1620	9634	25
1633	9314	26
1700	1649	27
1740	6644	28
1860	3595	29
1980	514	30
2040	-1029	31
2100	-2570	32
2280	-7147	33
2400	-10125	34
2600	-14875	35
3825	-32457	36

over short intervals and the process then yields the length of time a given tone pattern was present. Much of the patent is taken up by a description of a custom digital-signal-processor chip, but the related group of claims refers only to a ''machine-readable medium.''—DLR

7,103,540

43.60.Cg METHOD OF PATTERN RECOGNITION USING NOISE REDUCTION UNCERTAINTY

James G. Droppo *et al.*, assignors to Microsoft Corporation
5 September 2006 (Class 704/226); filed 20 May 2002

This patent improves extant methods for cleaning noise from a speech signal in a speech recognition application. The patent notes that prior art methods of noise removal do not exploit the statistics of the signal that are computed for other purposes, such as recognition. A method is then described that uses the measured uncertainty of noise reduction to good advantage within a hidden Markov modeling framework.—SAF

7,120,246

43.60.Dh COMPUTER APPARATUS HAVING A MULTIWAY ACOUSTIC OUTPUT AND INPUT SYSTEM

Stefano Ambrosius Klinke and Karl-Heinz Pflaum, assignors to Siemens Aktiengesellschait
10 October 2006 (Class 379/406.04); filed in Germany 16 March 2000

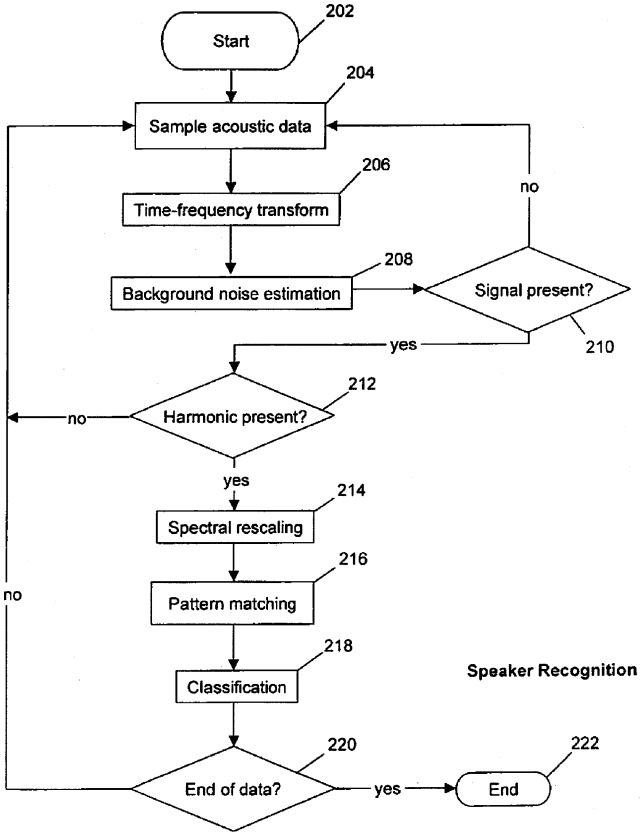
The patent deals with an add-on computer device that enables computer microphone and loudspeaker to operate in an echo-cancelling mode, regardless of the specific locations of the microphone, loudspeaker, and/or operator.—JME

7,117,149

43.60.Np SOUND SOURCE CLASSIFICATION

Pierre Zakarauskas, assignor to Harman Becker Automotive Systems-Wavemakers, Incorporated
3 October 2006 (Class 704/233); filed 30 August 1999

This concise but weighty patent deals with the complexities of identifying specific sound sources within a larger environment of other sound sources. The principal analysis method is to narrow down the field in terms of classes of sound (transients, harmonics, signal spectra, and other salient elements). These are compared with a vast library of specific identifiers. Eventually, through adaptive and various ''learning'' processes, a template can be matched and a source identified. Aside from the obvious uses in



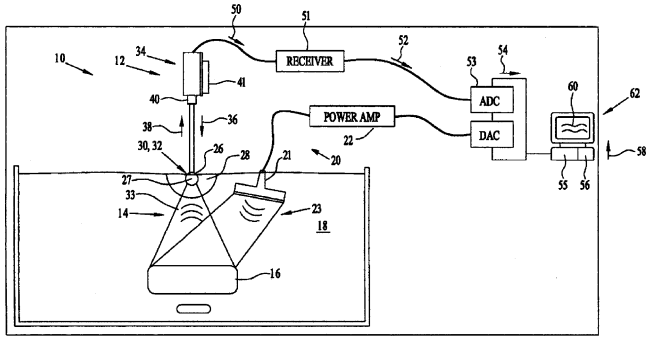
speech identification, other sounds such as musical sounds, sounds of nature, or mechanical sounds can also be identified. As the abstract states: ''If an existing template is found that resembles the input pattern, the template is averaged with the pattern in such a way that the resulting template is the average of all the spectra that matched that template in the past.'' The figure shows a typical flow chart for the system operating in a voice recognition mode.—JME

7,113,447

43.60.Rw LASER PUMPED COMPACT ACOUSTIC SENSOR SYSTEM

Anthony D. Matthews and Victor Johnson, assignors to The United States of America as represented by the Secretary of the Navy
26 September 2006 (Class 367/7); filed 13 September 2004

An underwater sound sensing system is claimed where the instantaneous motion of an immersed hollow sphere **26** (a ping-pong ball) is sensed by a laser velocimeter **34**. A target **16** is illuminated by sound **23** from transducer **21** or is self-excited by internal noise, such as from machinery. The sound from the target reaching **26** drives the position of **26**. Reflective spots **30**, **32**, ... on **26** reflect the incident laser light, said reflected laser light



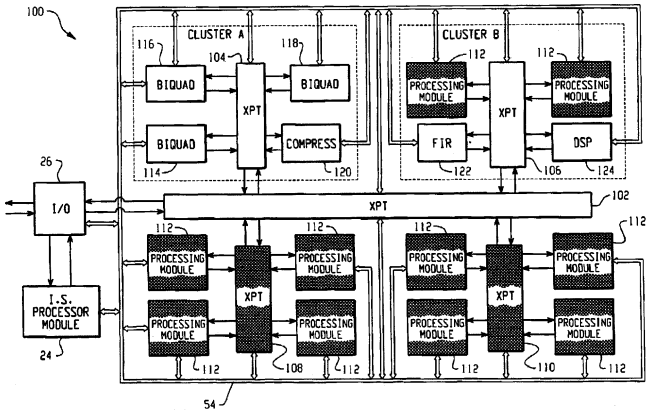
38 being frequency and phase modulated via the motion and displacement of **26**. The positions of the reflective spots are chosen to be sensitive to various radial directions and motions of sphere **26**. Each spot is addressed separately by translation of carriage **41** and consequently beams **36** and **38**. Signal processing **50-62** reduces data to provide target direction, velocity, and other characteristics. Sphere **26** resonance, 18.33 kHz in the case of a ping-pong ball, enhances the sensitivity of this sensor system at that frequency.—AJC

7,113,589

43.66.Ts LOW-POWER RECONFIGURABLE HEARING INSTRUMENT

Dennis Wayne Mitchler, assignor to Gennum Corporation
26 September 2006 (Class 379/399.01); filed 14 August 2002

An instruction-set (IS) processor module receives a hearing aid configuration and is configured to process audio signals. Audio signals from at least one processing module are coupled via a crosspoint-switch matrix to



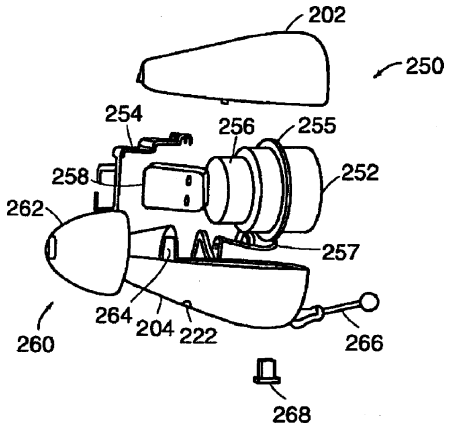
other processing modules, compressors, and filters. The hearing aid configuration programs the crosspoint-switch matrix configuration to control how audio signals are combined. A feedback signal generated from at least one processing module is monitored by the IS processor to determine an optimal configuration.—DAP

7,113,611

43.66.Ts DISPOSABLE MODULAR HEARING AID

Marvin A. Leedom *et al.*, assignors to Sarnoff Corporation
26 September 2006 (Class 381/322); filed 13 March 2001

A standard-fit disposable hearing aid designed for mass production consists of two half shells that interlock together to form a housing that is said to conform to the ear canal shape. An aperture in the housing is used to



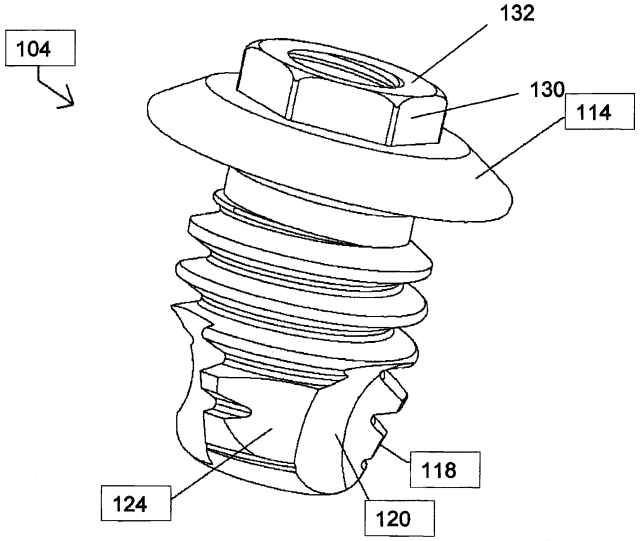
insert potting material into the hearing aid. One end of a flexible tip attached to the housing is mushroom-shaped to provide an acoustic seal in the ear canal and on the other end isolates the hearing aid receiver from the housing.—DAP

7,116,794

43.66.Ts HEARING-AID ANCHORING ELEMENT

Patrik Westerkull, Gothenburg, Sweden
3 October 2006 (Class 381/326); filed 4 November 2004

A device is described for anchoring a bone-conduction hearing aid directly to the skull bone. The device consists of a threaded portion having



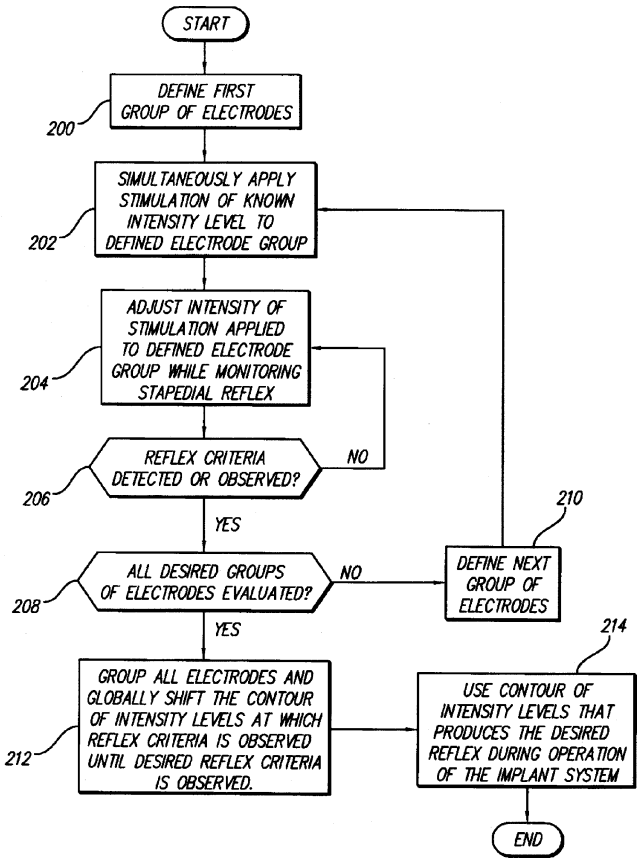
a depth related to its outer and inner diameters. The threaded portion also has a cutting edge with an adjacent cavity having a specified internal volume to collect bone fragments.—DAP

7,117,038

43.66.Ts METHOD AND SYSTEM FOR OBTAINING STAPEDIAL REFLEXES IN COCHLEAR IMPLANT USERS USING MULTIBAND STIMULI

Edward H. Overstreet, assignor to Advanced Bionics Corporation
3 October 2006 (Class 607/57); filed 15 September 2003

A cochlear implant system is fitted to patients using amplitude-modulated electrical stimuli that are applied to groups of multiple electrodes



at pulse rates typically>2 kHz. The resulting stapedial reflexes of the patient are monitored to predict comfort levels for speech and to adjust the intensity levels of the electrical stimuli.—DAP

7,123,732

43.66.Ts PROCESS TO ADAPT THE SIGNAL AMPLIFICATION IN A HEARING DEVICE AS WELL AS A HEARING DEVICE

Hans-Ueli Roeck, assignor to Phonak AG
17 October 2006 (Class 381/321); filed 10 September 2002

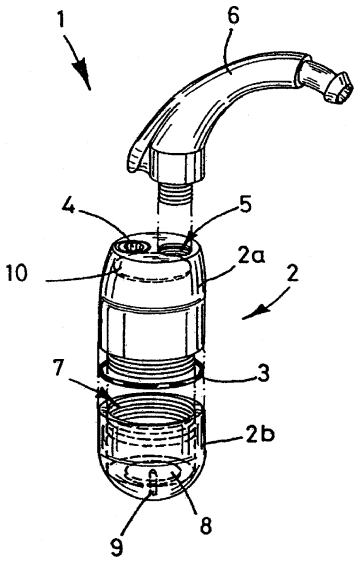
The amount of amplification is determined by first determining a reference level for the input signal using statistical methods, and then defining adaptive minimum and maximum input signal levels by subtracting or adding, respectively, predefined difference values from/to the input signal reference level. Different amplification functions, for example, squelch or limiting, are applied to the input signal depending on whether it is above or below the minimum- or maximum-level values. A first time constant is used to determine the reference level and a second smaller time constant is used to set the amplification function.—DAP

7,123,733

43.66.Ts AUDITORY TREATMENT DEVICE

Hans-Dieter Borowsky *et al.*, assignors to Auric Horsysteme GmbH & Company KG
17 October 2006 (Class 381/322); filed in Germany 27 January 1999

A behind-the-ear hearing aid assembly has a cylindrical-shaped metal housing that is detachable from both a battery compartment and a sound exit opening covered by an acoustically transparent water-tight film. The metal



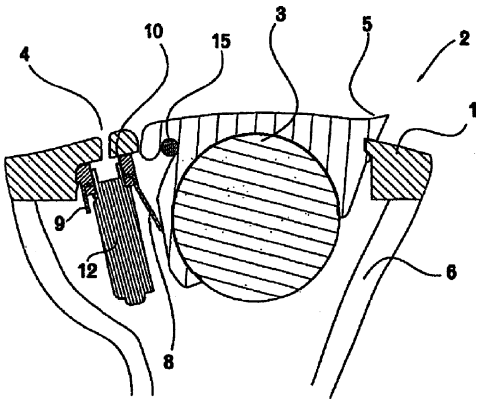
housing incorporates no mechanical user controls and shields the circuitry from electromagnetic interference. The battery compartment employs a water-tight O-ring seal and includes a ring magnet for retaining batteries.—DAP

7,127,077

43.66.Ts ITE HEARING AID AND CONTACT MODULE FOR USE IN AN ITE HEARING AID

Matthew Hall and Jesper Kock, assignors to Oticon A/S
24 October 2006 (Class 381/312); filed in Denmark 25 April 2001

The contacts for programming the signal processor of a custom in-the-ear or completely-in-the-canal hearing aid are contained, together with the



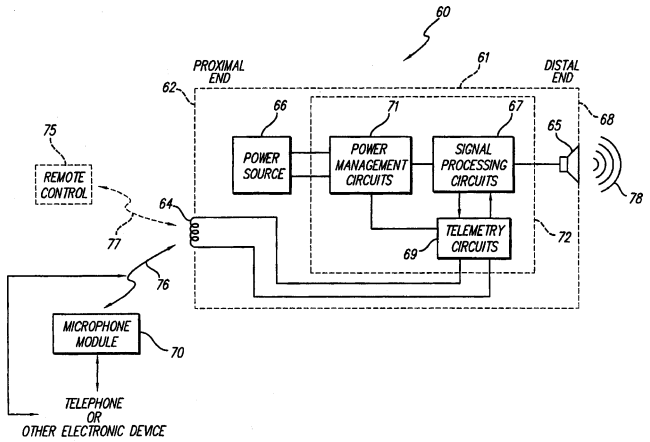
microphone, in a separate module that inserts into the faceplate. Each contact has a first and second leg, which may be on opposite sides of the microphone and connected with an intermediate part.—DAP

7,127,078

43.66.Ts IMPLANTED OUTER EAR CANAL HEARING AID

Alfred E. Mann *et al.*, assignors to Advanced Bionics Corporation
24 October 2006 (Class 381/326); filed 5 November 2003

One portion of a hearing aid system consists of an implanted housing containing electronics and power source placed under the skin and adjacent to the ear canal and to the retro-auricular space behind the pinna. An acoustic output transducer drives processed sound energy into the ear canal from



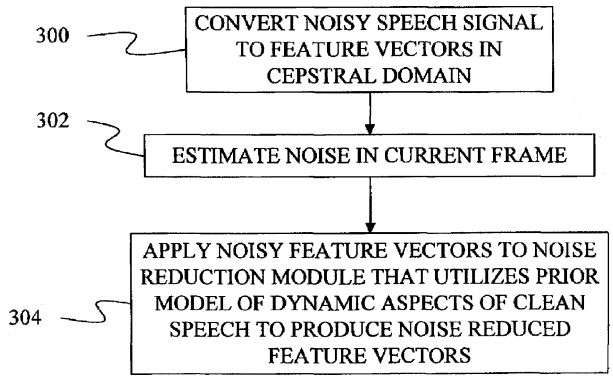
the electronics in the implanted housing. A remote microphone, which can be hand-carried by the patient, transmits sounds to the implanted signal processing electronics via wireless telemetry.—DAP

7,107,210

43.72.Dv METHOD OF NOISE REDUCTION BASED ON DYNAMIC ASPECTS OF SPEECH

Li Deng *et al.*, assignors to Microsoft Corporation
12 September 2006 (Class 704/226); filed 20 May 2002

A complicated approach to speech noise reduction is described that seems to represent a significant advance. The authors note that prior art methods estimate noise-reduced feature vectors using static noise models



and static models of clean speech. As shown in the figure, a method is put forth for utilizing a prior dynamic model of clean speech in a new sort of acoustic model that relates clean speech, noisy speech, and noise. Sufficient detail is provided.—SAF

7,107,214

43.72.Dv MODEL ADAPTATION APPARATUS, MODEL ADAPTATION METHOD, STORAGE MEDIUM, AND PATTERN RECOGNITION APPARATUS

Hironaga Nakatsuka, assignor to Sony Corporation
12 September 2006 (Class 704/244); filed 8 July 2005

Within the context of word recognition in a noisy environment, the general class of stochastic feature extraction methods can be degraded due to a failure to jointly model the actual noise. The patent puts forth ''a technique in which an acoustic model is corrected using information of ambient noise...'' The poorly translated document then offers that a speech acoustic model can be adapted to the noise detected during the prespeech

interval "by means of one of the most likelihood method, the complex statistic method, and the minimum distance-maximum separation theorem.—SAF

7,110,944

43.72.Dv METHOD AND APPARATUS FOR NOISE FILTERING

Radu Victor Balan and Justinian Rosca, assignors to Siemens Corporate Research, Incorporated
19 September 2006 (Class 704/226); filed 27 July 2005

This patent tersely describes a technique for noise filtration that relies on some kind of microphone array. The relative impulse response between adjacent microphones is used in an apparently novel formulation to compute the short-time spectral amplitude and short-time complex exponential, and from these the noise can be eliminated under certain stationarity and independence assumptions.—SAF

7,107,213

43.72.Fx DEVICE FOR NORMALIZING VOICE PITCH FOR VOICE RECOGNITION

Mikio Oda and Tomoe Kawane, assignors to Matsushita Electric Industrial Company, Limited
12 September 2006 (Class 704/234); filed in Japan 29 October 1999

This patent puts forth a simple pitch adjustment to incoming speech signals, to better align the pitch input to the recognizer with the system's optimal pitch. The input voice has its pitch modified by a probabilistic procedure until the likelihood of the received signal (given some plurality of possible recognized words) reaches a threshold probability.—SAF

7,117,053

43.72.Gy MULTI-PRECISION TECHNIQUE FOR DIGITAL AUDIO ENCODER

Mohammed Javed Absar *et al.*, assignors to STMicroelectronics Asia Pacific Pte. Limited
3 October 2006 (Class 700/94); filed 26 October 1998

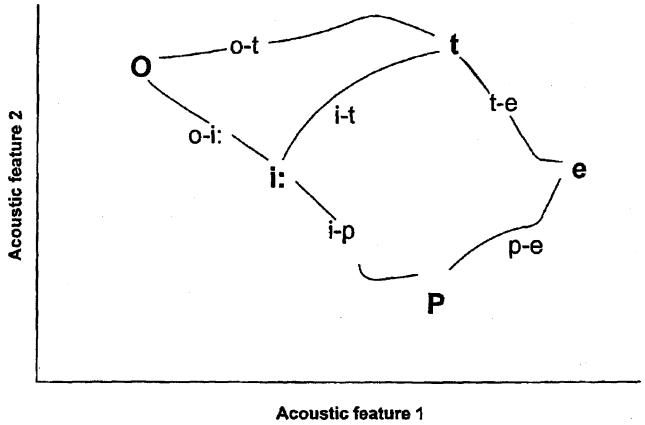
A method is proposed for coding digital audio data on a 16-bit, fixed-point digital signal processor having several stages with multiple levels of computational precision. The encoding process adheres to the AC-3 Digital Audio Compression standard and is said to produce the equivalent of 24-bit encoding without significant additional computational complexity.—DAP

7,010,488

43.72.Ja SYSTEM AND METHOD FOR COMPRESSING CONCATENATIVE ACOUSTIC INVENTORIES FOR SPEECH SYNTHESIS

Jan P. H. van Santen and Alexander Kain, assignors to Oregon Health & Science University
7 March 2006 (Class 704/258); filed 9 May 2002

According to this patent, speech synthesis systems that work by concatenating short segments of speech typically use vector coding to store the generally large inventory of segments. Each stored segment includes all of



the acoustic parameters needed to insert that segment into the generated speech signal. The method patented here would use some of the acoustic parameters themselves as the lookup keys. This has the added benefit that those parameters no longer need be stored with the segment.—DLR

7,010,489

43.72.Ja METHOD FOR GUIDING TEXT-TO-SPEECH OUTPUT TIMING USING SPEECH RECOGNITION MARKERS

James R. Lewis *et al.*, assignors to International Business Machines Corporation
7 March 2006 (Class 704/260); filed 9 March 2000

Speech synthesis system designers, in general, have yet to solve the problems of producing natural-sounding fluent output. The approach described here would be used in conjunction with a recognition system, such as a dictation system. During the process of analyzing the input speech signal, special markers are stored that represent the prosodic structure of the spoken material. These markers would include sufficient phonetic, syntactic, and/or semantic cues such that prosodic information can be recalled at the appropriate places and times during synthesis. In the simple case, as covered in the patent text, the synthesizer is simply playing back a portion of text that has just been dictated. The markers can then be more or less assumed to be relevant. Whether a more general case can be handled is left as an open question.—DLR

7,113,909

43.72.Ja VOICE SYNTHESIZING METHOD AND VOICE SYNTHESIZER PERFORMING THE SAME

Nobuo Nukaga *et al.*, assignors to Hitachi, Limited
26 September 2006 (Class 704/258); filed 31 July 2001

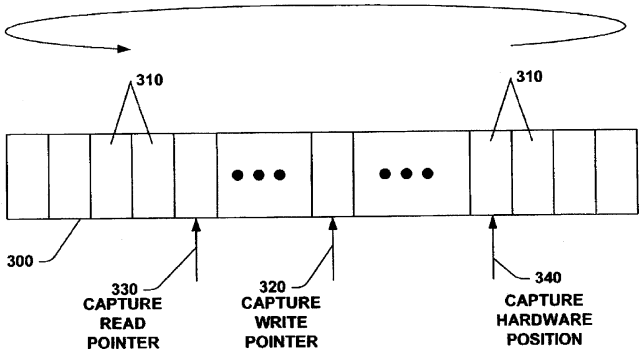
This patent purports to describe a method for synthesizing speech in any number of different "speech styles" by invoking a particular speech-style dictionary. The basic considerations given do not go very far in providing any actual methods for specifying a speech style in a way that would allow uniform and adequate modification of a segmental synthesizer. There are some half-baked notions of changing the prosody and the segmental timing.—SAF

7,120,259

43.72.Kb ADAPTIVE ESTIMATION AND COMPENSATION OF CLOCK DRIFT IN ACOUSTIC ECHO CANCELLERS

Joseph Cox Ballantyne *et al.*, assignors to Microsoft Corporation
10 October 2006 (Class 381/71.12); filed 31 May 2002

Echoes forming in a full-duplex audio telephone system are attenuated,



where voice data are captured in buffer **300**, a write pointer **320** and a capture delay are computed, then the read pointer **330** is adjusted.—AJC

7,003,457

43.72.Ne METHOD AND SYSTEM FOR TEXT EDITING IN HAND-HELD ELECTRONIC DEVICE

Katriina Halonen and Sailesh Sathish, assignors to Nokia Corporation
21 February 2006 (Class 704/235); filed 29 October 2002

This speech recognizer is intended for dictation or text editing in a cell phone or other hand-held unit. A small command-word grammar is contained in the unit's local memory, but the device also has access to a larger grammar, typically located at a remote site. As content words are recognized using the large grammar, those words are added to the small grammar. After the editing session, the newly added words are deleted.—DLR

7,003,459

43.72.Ne METHOD AND SYSTEM FOR PREDICTING UNDERSTANDING ERRORS IN AUTOMATED DIALOG SYSTEMS

Allen Louis Gorin *et al.*, assignors to AT&T Corporation
21 February 2006 (Class 704/240); filed 22 January 2001

This speech recognition system uses a dialog manager to estimate the probabilty that the user is being correctly understood. As the dialog proceeds, the set of recognized items and the state of a natural-language-understanding monitor unit are evaluated according to the current dialog state and a dialog history database. If a predetermined level of accuracy is not maintained, the system will switch to a failure-mode dialog, requesting help from any of several sources and informing the user of that change. The patent reports that the system was evaluated in a test using recorded dialogs involving nearly 12 000 utterances.—DLR

7,010,427

43.72.Ne ROUTE GUIDANCE SYSTEM HAVING VOICE GUIDANCE CAPABILITY

Masaki Ebi, assignor to Denso Corporation
7 March 2006 (Class 701/210); filed in Japan 8 April 2003

The vehicle navigation system described here would use voice commands to set up and step through landmarks as they are passed or missed while following a preset route, apparently under the assumption that a GPS receiver is not available to track the vehicle's location. The recognizer has no provision to detect speech input—a push-to-talk switch is specified. Voice commands may be used to add or delete landmarks or to update the state of progress through waypoints, such as intersections, detours, or milemarkers.—DLR

7,010,486

43.72.Ne SPEECH RECOGNITION SYSTEM, TRAINING ARRANGEMENT AND METHOD OF CALCULATING ITERATION VALUES FOR FREE PARAMETERS OF A MAXIMUM-ENTROPY SPEECH MODEL

Jochen Peters, assignor to Koninklijke Philips Electronics, N.V.
7 March 2006 (Class 704/255); filed in Germany 13 February 2001

Disclosed is an n-gram-based speech recognizer in which the probability of each specific n-word sequence is based on a maximum entropy computation, as is well-known in prior-art recognition systems. In this system, formulas are given for updating the probabilities, including a method of setting boundary values, which may be applied to narrow the search range for finding the most probable n-gram. Both the maximum entropy and the boundary computation methods are described in cited publications, but not in the patent.—DLR

7,103,543

43.72.Ne SYSTEM AND METHOD FOR SPEECH VERIFICATION USING A ROBUST CONFIDENCE MEASURE

Gustavo Hernandez-Abrego and Xavier Menendez-Pidal, assignors to Sony Corporation
5 September 2006 (Class 704/240); filed 13 August 2002

A very simplistic method is described for using a speech recognition confidence measure as a score for whether speech is in fact present in the signal or, alternatively, for determining whether a received signal contains any recognizable vocabulary item.—SAF

7,103,544

43.72.Ne METHOD AND APPARATUS FOR PREDICTING WORD ERROR RATES FROM TEXT

Milind Mahajan *et al.*, assignors to Microsoft Corporation
5 September 2006 (Class 704/240); filed 6 June 2005

This patent describes a method of doing the opposite of speech recognition; namely, given a text, the methods compute phonetic confusion metrics to predict the word-error rate that would be expected with the speech recognition model at hand. The stated purpose of this seemingly strange calculation is to be able to determine how well a particular acoustic- and language-model combination would work in a given text domain, without the need for actual testing.—SAF

7,113,908

43.72.Ne METHOD FOR RECOGNIZING SPEECH USING EIGENPRONUNCIATIONS

Silke Goronzy and Ralf Kompe, assignors to Sony Deutschland GmbH
26 September 2006 (Class 704/244); filed in the European Patent Office 7 March 2001

Chiefly addressing the problems of speech recognition when a user has pronunciations that deviate considerably from the most common (e.g., when

the user is not a native speaker), a method is proposed for deriving a set of pronunciation rules for the target speaker and situating the current pronunciation in a ''pronunciation space.'' The pronunciation rules can then be used to perform a speaker-specific adaptation of a recognizer.—SAF

7,117,150

43.72.Ne VOICE DETECTING METHOD AND APPARATUS USING A LONG-TIME AVERAGE OF THE TIME VARIATION OF SPEECH FEATURES, AND MEDIUM THEREOF

Atsushi Murashima, assignor to NEC Corporation
3 October 2006 (Class 704/233); filed in Japan 2 June 2000

This patent introduces a method and circuit for detecting voiced/unvoiced segments of a speech signal based on changes in the long-time average of audio parameters. The method calculates four parameters from frames of the input signal and compares them with their corresponding long-time average values. These parameters are spectral energy, signal energy, low-band energy, and zero-crossing number. If the difference between long-time and short-time average values is above the threshold, then the frame is labeled as voiced.—AAD

7,127,046

43.72.Ne VOICE-ACTIVATED CALL PLACEMENT SYSTEMS AND METHODS

Robert C. Smith and George Demetrios Karis, assignors to Verizon Laboratories Incorporated
24 October 2006 (Class 379/88.03); filed 22 March 2002

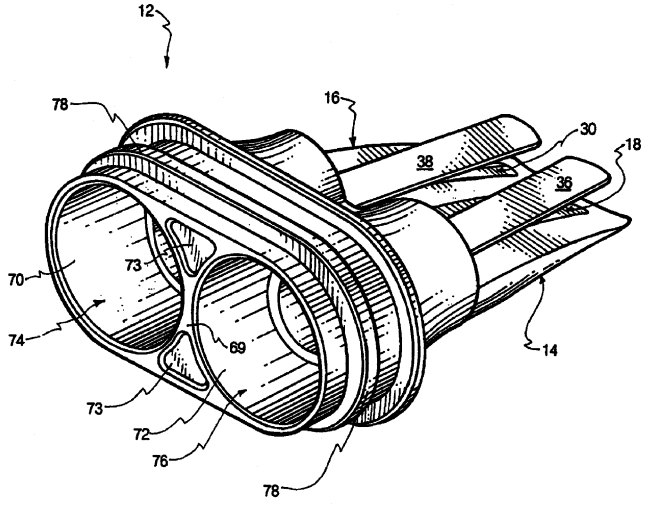
A voice recognition system receives utterances and nonspeech data such as a telephone number or zip code from a calling party and recognizes the entity type of the called party. The called-party name is identified from a database corresponding to the identified type of entity of the called party. The calling party is connected to the called party that was identified.—DAP

6,926,578

43.75.Ef DOUBLE INLET GAME CALL APPARATUS AND METHOD

Mark A. Casias and Wilbur R. Primos, assignors to Primos, Incorporated
9 August 2005 (Class 446/202); filed 15 April 2002

In the never ending struggle between man and beast (elk in particular),



a dual reed **36**, **38** game call is described. Two reeds are used to create two (competitive) animal sounds.—MK

6,925,880

43.75.Hi APPARATUS AND METHOD FOR MEASURING THE ACOUSTIC PROPERTIES OF A MEMBRANOPHONE

John H. Roberts, Hickory, Mississippi
9 August 2005 (Class 73/587); filed 17 November 2003

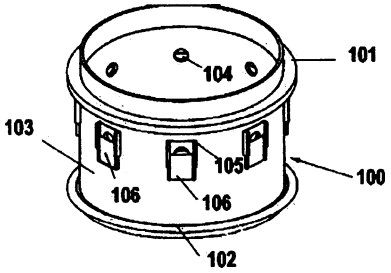
Tuning drums is a trial and error process. By exciting the drum head by a pair of loudspeakers, a microphone can measure the phase difference between the excitation and the received signal. This can be used to look for nodes, which are well described by Bessel functions. However, the patent does not describe where to place the transducers on a head, leaving it as a matter of trial and error.—MK

6,927,330

43.75.Hi DRUM WITH MODULATED ACOUSTIC AIR VENT

Randall L. May, Irvine, California
9 August 2005 (Class 84/411 R); filed 24 June 2003

Simply put, the inventor proposes the use of slide valves **106** that can



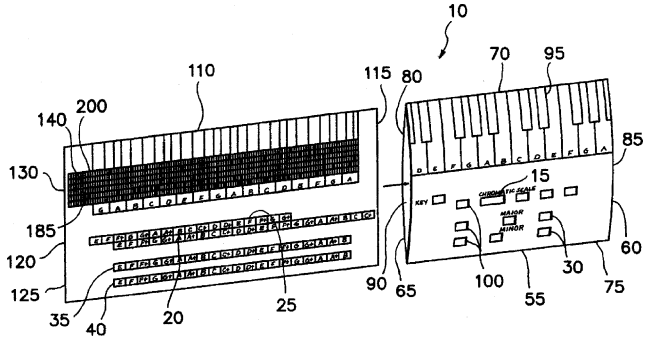
be used to tune the drum. Numerous variations are given.—MK

6,969,793

43.75.St PIANO KEY FINDER AND CHORD INDICATOR

Dean Kerkhoff, Baldwin Park, California
29 November 2005 (Class 84/478); filed 15 July 2003

As shown, this is a kind of musical slide rule (an early visual analog computer for those too young to know). The outer sleeve **55** slides over the



inner part **110** to give the student major and melodic minor scales. It is no substitute for rote learning of intervals, however.—MK

6,958,442

43.75.Wx DYNAMIC MICROTUNABLE MIDI INTERFACE PROCESS AND DEVICE

Ahmed Shawky Moussa, assignor to Florida State University Research Foundation
25 October 2005 (Class 84/645); filed 6 February 2003

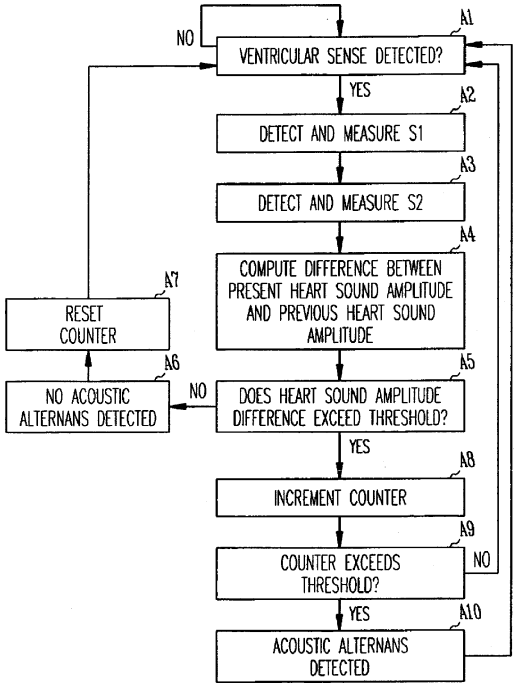
The inventor is concerned with MIDI's inability to represent nonwestern tuning systems. The solution as given is to use ''aftertouch'' to bend the pitch. A pity that this is blatantly obvious to anyone who has used MIDI.—MK

7,113,825

43.80.Qf METHOD AND APPARATUS FOR DETECTING ACOUSTIC OSCILLATIONS IN CARDIAC RHYTHM

Joseph M. Pastore *et al.*, assignors to Cardiac Pacemakers, Incorporated
26 September 2006 (Class 607/14); filed 3 May 2002

This is an implantable cardiac rhythm management device configured to detect oscillations in cardiac rhythm by measuring the heart sound amplitude during successive heart beats. Upon detection of acoustic alternations, the device may adjust its operating behavior to compensate for the deleterious effects of the condition. Acoustic alternation is indicative of the alternations in the strength of systolic contractions and pulse pressure, a condition known as *pulsus alternans*, generally taken by clinicians to indicate systolic dysfunction and possible heart failure.—DRR



7,115,092

43.80.Vj TUBULAR COMPLIANT MECHANISMS FOR ULTRASONIC IMAGING SYSTEMS AND INTRAVASCULAR INTERVENTIONAL DEVICES

Byong-Ho Park *et al.*, assignors to The Board of Trustees of the Leland Stanford Junior University
3 October 2006 (Class 600/143); filed 18 September 2003

These mechanisms are formed from a tube of an elastic or superelastic material. They are fabricated by laser machining and have no mechanical joints. The mechanisms can be manipulated into various shapes without permanent deformation.—RCW

7,115,093

43.80.Vj METHOD AND SYSTEM FOR PDA-BASED ULTRASOUND SYSTEM

Menachem Halmann and Shinichi Amemiya, assignors to GE Medical Systems Global Technology Company, LLC
3 October 2006 (Class 600/437); filed 12 April 2002

A personal digital assistant (PDA) is interfaced to a hand-held probe assembly through a digital interface. The probe assembly consists of a beamforming module and a detachable transducer head. The PDA runs Windows™ applications, displays menus and images to a user, and runs ultrasound data processing software that supports different imaging modes.—RCW

7,115,094

43.80.Vj ULTRASONIC PROBE, ULTRASONIC IMAGING APPARATUS AND ULTRASONIC IMAGING METHOD

Takashi Azuma and Shinichiro Umemura, assignors to Hitachi, Limited
3 October 2006 (Class 600/459); filed in Japan 30 October 2001

The design of the ultrasonic probe in this system reduces electromagnetic leakage to a magnetic resonance imaging (MRI) system when the probe is inside the MRI gantry. Electronics in the system form transmit and receive beams. The transmit-receive electronics can be electrically disconnected from the probe by a switch.—RCW

LETTERS TO THE EDITOR

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

Temporal limits of level dominance in a sample discrimination task (L)

Matthew D. Turner and Bruce G. Berg[a)]
Department of Cognitive Sciences, 3151 Social Sciences Plaza, University of California, Irvine, California 92697-5100

(Received 4 September 2006; revised 29 January 2007; accepted 30 January 2007)

Level dominance refers to the effect where attention is automatically directed to the loudest part of an auditory display. In a sample discrimination task, the frequencies of five 50 ms tones were sampled from normal distributions with means of 1000 and 1100 Hz and presented sequentially, with the tones alternating in intensity. Observers decide from which distribution the sample was drawn. The informativeness of the even numbered tones ($d'=2$) was greater than the informativeness of the odd numbered tones ($d'=1$). Estimates of decision weights and performance levels (d') show that when the more informative tones were less intense, observers attended to the louder tones rather than the more informative tones. This effect extends well beyond the temporal limits expected from forward masking studies. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2710345]

PACS number(s): 43.66.Mk, 43.66.Fe, 43.66.Cb [RAL] Pages: 1848–1851

I. INTRODUCTION

This report concerns a property of auditory attention we refer to as level dominance, a straightforward notion that attention is automatically directed to the loudest component in an auditory display. Berg (1990) estimated decision weights in a sample discrimination task which uses sequences of tone bursts with frequencies sampled from normal distributions. The odd numbered and even numbered tones had respective sensitivity indices of $d'=1$ and $d'=2$, so the even numbered tones were more reliable (i.e., more informative). The intensity of the tones was sequentially alternated with level differences between odd and even tones of either 20 or 40 dB. The ideal observer gives greater weight to the more reliable tones without regard to their intensity. Real observers give greater weight to the more intense tones without regard to their reliability, implying that observers cannot selectively attend to the quieter tones in a sequential presentation.

There must be limits to the existence region of level dominance. Lutfi and Jesteadt (in press) investigated the effect of intensity differences in a multiple tone increment detection task. Five tones were presented sequentially, with odd and even tones alternating in intensity. The signal was an intensity increment of 6 dB for the quiet tones and 3 dB for the loud tones. For alternating intensity differences between 10 and 40 dB they found that observers give greatest weight to the louder tones, even though the quieter tones had the greater intensity increment. The effect diminished only when the intensity difference was reduced below 10 dB. They also found that the effect is diminished substantially when dissimilar stimuli were alternated (i.e., noise and tone).

The temporal limits of level dominance are considered here. Berg (1990) found that the effect was undiminished when the inter-tone interval (ITI) was increased from 0 to 200 ms, suggesting that the effect is not attributable to forward masking (Jesteadt *et al.*, 1982). Lutfi and Jesteadt (in press) found no differences between conditions with a 0 or 100 ms ITI. In this experiment, the ITI is extended to durations where level dominance diminishes.

II. METHOD

A. Basic procedure and stimuli

The experiment was a sample discrimination task (Berg, 1990; Lufti, 1989). On each trial, observers were presented with a sequence of five tones and decided whether the frequencies of the tones were sampled from normal distributions with a mean frequency of 1100 Hz ("high" frequency distributions) or from others with mean frequency of 1000 Hz ("low" frequency distributions). All five tones on a given trial came from distributions with the same mean frequency. The frequencies of the odd tones (first, third, and fifth) were sampled from a distribution with standard devia-

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

 0001-4966/2007/121(4)/1848/4/$23.00

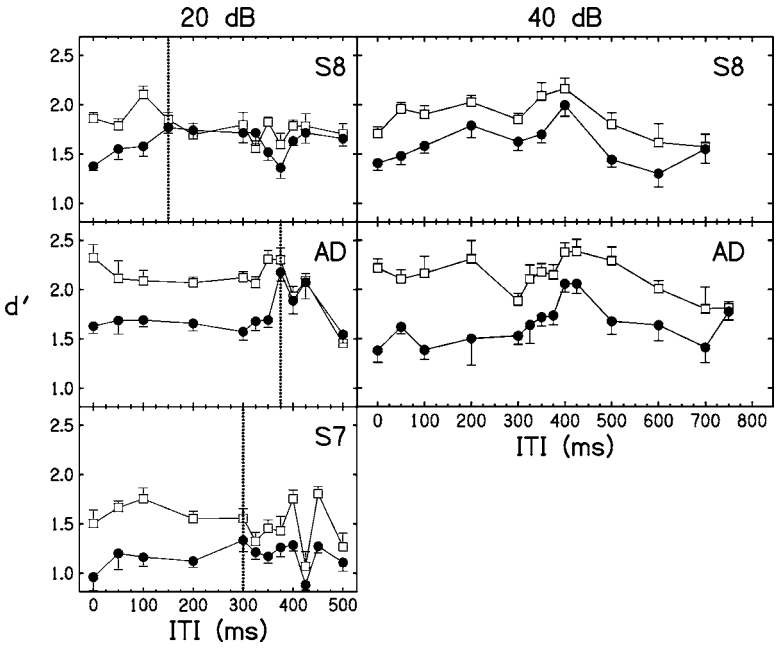


FIG. 1. d' as a function of ITI (inter-tone interval). Open squares represent the loud-reliable condition, filled circles represent quiet-reliable condition. Each row represents a single observer. The left column contains conditions with a 20 dB level difference; the right column is 40 dB. Vertical dotted lines in the 20 dB difference condition represent the point of d' convergence as defined in the text. Error bars represent standard error of the d' estimates.

tion of 100 Hz (greater variance/lower reliability) while the even tones (second and fourth) came from a Normal distribution with standard deviation of 50 Hz (lesser variance/higher reliability). After presentation of each sequence, observers indicated whether the sequence was sampled from the high or low distributions. They were given immediate feedback on their accuracy after each trial.

B. Conditions

There were three factors manipulated in the experiment: alternating intensity difference, the correlation of intensity with reliability, and inter-tone interval (ITI).

There were two intensity differences. In the 20 dB difference condition the quieter tones were presented at 45 dB SPL and the louder at 65 dB SPL. In the 40 dB difference condition the levels were 45 and 85 dB SPL, respectively. For each intensity difference there were two reliability configurations: louder tones reliable (loud-reliable) or quieter tones reliable (quiet-reliable).

Within each condition, ITI was set to values between 0 and 750 ms. The values of ITI were selected as the experiment progressed, giving additional ITI values in regions of particular interest. For each value of ITI a minimum of 500 trials (five blocks of 105 trials, with the first five trials of each block discarded) were collected. For ITIs below 500 ms an average of approximately 1000 trials (ten blocks) were collected. For each block the d' was computed, and the standard error of these d' values is based on the collection of estimates, one from each block.

C. Apparatus and observers

Stimuli were generated digitally on a microcomputer, played out with a 16 bit digital-to-analog converter at a sampling rate of 20 kHz, low pass filtered at either 6 or 10 kHz, and presented to the observers diotically through Sennheiser HD414SL headphones. Responses were collected by computer keyboard and feedback was provided by cathode ray tube. Observers were seated individually in single walled acoustic chambers. The three observers were paid volunteers, aged 18–22 years, with a minimum of two months of experience in various psychoacoustic tasks. All three had normal pure tone thresholds.

III. RESULTS

Figure 1 displays d' as a function of ITI. Each column represents a different alternating level difference, 20 or 40 dB, respectively, and each row represents a different observer. For brief ITIs, performance in loud-reliable conditions (open symbols) is always better than performance in quiet-reliable conditions (closed symbols). Level dominance in the quiet reliable condition incurs a cost in overall performance (d') because the observers are presumably unable to attend selectively to the quiet tones. Of interest is the ITI at which performance is the same between the two conditions. The upper limit of level dominance is arbitrarily defined as the first ITI where the standard error of the d' values for the quiet-reliable and loud-reliable conditions overlaps, but only if this overlap is maintained at the next consecutive ITI value. We refer to this as "convergence."

If the effect results from forward masking, then one would expect that the d' values would converge by no more than 100–200 ms (Jesteadt *et al.*, 1982). In the 20 dB difference conditions (first column) this criterion is attained at 150 ms ITI for observer S8, 375 ms for AD, and 300 ms for S7. In the 40 dB difference conditions the convergence criterion is never attained. Observer S7 withdrew from the research before completing this condition. With only three exceptions, all in the 20 dB condition (S8 at both 200 and 325 ms and AD at 500 ms), all observers perform at least the same or better in the loud-reliable conditions across all ITI. After d' convergence in the 20 dB difference conditions, the d' values stay close together for S8 and AD, but less so for

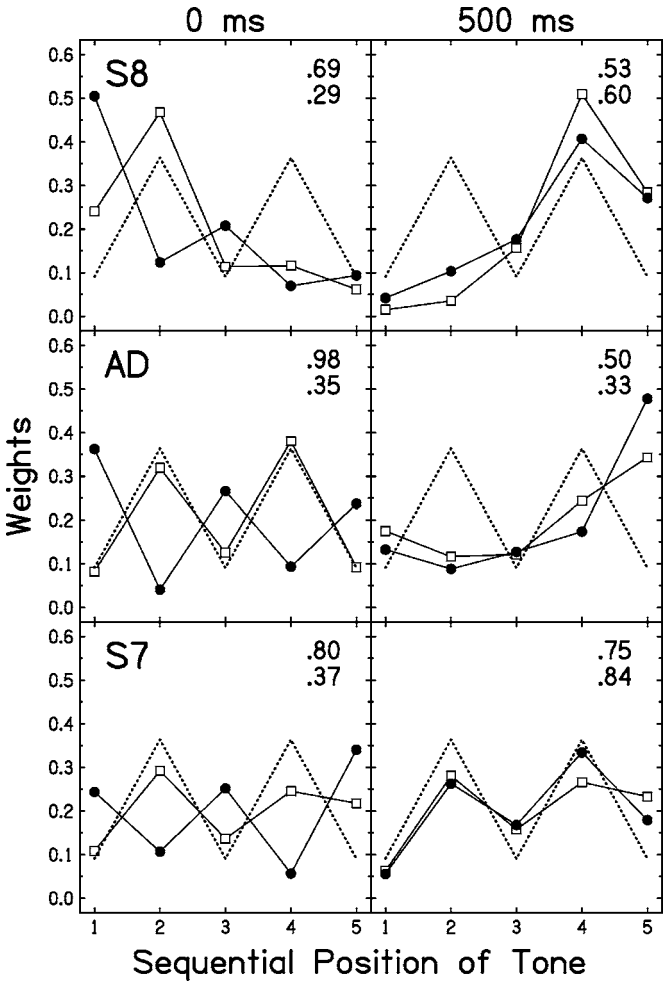


FIG. 2. Decision weights from the 20 dB difference conditions. Each row represents a single observer. Open squares represent the loud-reliable condition, filled circles represent quiet reliable. The columns represent the shortest (0 ms) ITI (inter-tone interval) and longest (500 ms) ITI, respectively. In all frames, the dotted lines represent the weights for the ideal observer. The two numbers in the upper right of each frame represent weighting efficiencies (η_{wgt}) for the loud reliable (upper number) and quiet reliable (lower number) as described in the text.

S7. What is clear throughout is that the d' values do not converge within the expected region for forward masking effects.

Estimates of the decision weights for the 20 and 40 dB difference conditions are shown in Figs. 2 and 3, respectively. Decision weights are computed as described in Berg (1990). The first column shows weights for the briefest ITI, and the second column shows weights for the longest ITI. Optimal weights are represented by the dotted lines, and the upper and lower numbers in the top right corner of each panel represent the weighting efficiency (η_{wgt}; Berg, 1990) for the loud-reliable and quiet-reliable conditions, respectively.

At 0 ms ITI, AD shows near ideal weights when the louder tones are reliable (for either level difference), whereas S8 exhibits a tendency to give greater weight to the initial tones of the sequence. Note that there is supportive evidence that the weights would not be distributed ideally even if the reliabilities were all equal, though this is not fully explored here (Berg, 1990; Pedersen and Ellermeier, 2005). For all

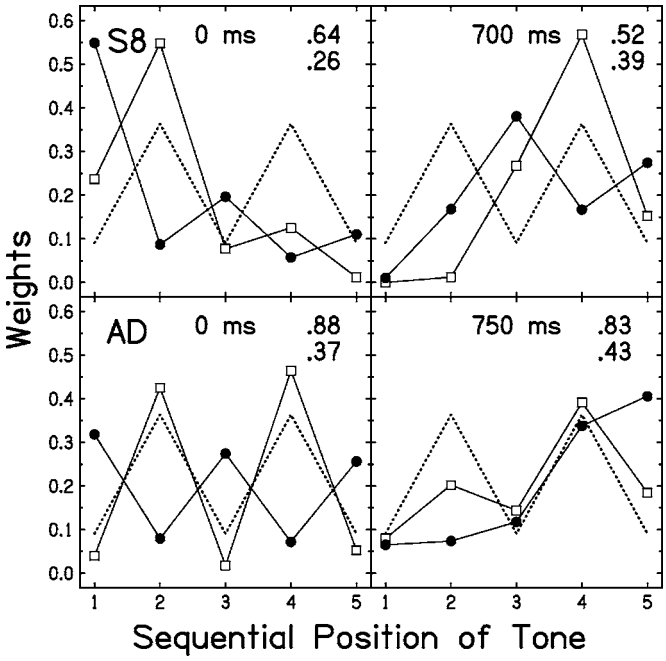


FIG. 3. Decision weights from the 40 dB difference conditions. The columns represent the shortest (0 ms) ITI and longest (700 ms for S8; 750 ms for AD) ITI, respectively. All other details as in Fig. 2.

observers the weighting efficiencies at 0 ms ITI are at least twice as great for loud reliable as they are for quiet reliable. This misuse of available information explains the differences in d'.

By the 500 ms ITI in the 20 dB difference conditions, all three observers have clearly altered their listening strategy, and appear to be doing the same thing for both loud-reliable and quiet-reliable conditions. By the longest ITI conditions this has become an emphasis on one of the final two tones for observers S8 and AD.

The results for the 40 dB difference condition are somewhat difficult to interpret because of an apparent recency effect. Note that the earliest loud tone in both conditions, the first quiet-reliable tone or the second loud-reliable tone, always receives the greatest weight at 0 ms ITI, but exhibits one of the lowest weights by the longest ITI. With an ITI of 700 or 750 ms, the stimulus interval is greater than 3 s. This seems to have encouraged a strategy of shifting maximum weight to one of the last observations of the sequence. The failure to use information from the initial observations is consistent with the observed decrements in d' for ITIs greater than 400 ms.

IV. DISCUSSION

The level dominance effect is found at ITIs much longer than expected if forward masking was the only factor. We conclude that level dominance manifests a limitation at a central stage of auditory processing, a conclusion also reached by Berg (1990) and by Lutfi and Jesteadt (in press). Differences in convergence between the 20 and 40 dB conditions suggest a decay function, as the duration of interference is longer in the 40 dB condition. The present study has established both an interesting effect as well as some techni-

cal limitations of this class of experiments. The uncommonly long stimulus duration for long ITIs may add a complicating factor, a recency effect, suggesting the use of fewer tones in the sequence in future studies.

ACKNOWLEDGMENTS

David Pricz was a significant contributor to this research but he was unavailable at the time of submission. He should be credited with the collection and the original presentation of the data. This work was supported in part by a grant from the President's Undergraduate Fellowship of the University of California, Irvine. The authors would like to thank the editor, Dr. Robert Lutfi, and two anonymous reviewers for their thoughtful comments.

Berg, B. G. (**1990**). "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158.

Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.

Lutfi, R. A. (**1989**). "Informational processing of complex sound. I. Intensity discrimination," J. Acoust. Soc. Am. **86**, 934–944.

Lutfi, R. A., and Jesteadt, W. (**2006**). "Molecular analysis of the effect of relative tone level on multitone pattern discrimination," J. Acoust. Soc. Am. **120**, 3853–3860.

Pedersen, B., and Ellermeier, W. (**2005**). "Temporal and spectral interaction in loudness perception," J. Acoust. Soc. Am. **117**, 2397.

Beamforming using spatial matched filtering with annular arrays (L)

Kang-Sik Kim, Jie Liu, and Michael F. Insana[a)]
Department of Bioengineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 3120 DCL, MC-278 1304 W. Springfield Avenue, Urbana, Illinois 61801

(Received 30 August 2006; revised 13 January 2007; accepted 17 January 2007)

A linear array beamforming method for ultrasonic *B*-mode imaging using spatial matched filtering (SMF) and a rectangular aperture geometry was recently proposed Kim *et al.*, [J. Acoust. Soc. Am. **120**, 852–861 (2006)]. This letter extends those results to include circularly symmetric apertures. SMF applied to annular arrays can improve the lateral resolution and echo signal-to-noise ratio as compared with conventional dynamic-receive delay-sum beamforming. At high frequencies, where delay and sum beamforming is problematic, SMF showed greatly improved target contrast over an extended field of view. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2642214]

PACS number(s): 43.60.Fg, 43.35.Yb [TDM] Pages: 1852–1855

I. INTRODUCTION

Circular aperture transducers, including annular arrays, are widely used particularly at high frequencies to improve focusing with a small number of transducer elements. Above 30 MHz, accurate digital beamforming is challenging for arrays with many elements.[1] Instead, mechanically scanned, circular, single-element transducers geometrical focused at a fixed depth are used to image subsurface structures in the human eye, skin, and inside small animals.[2] Although the spatial resolution can approach the wavelength in these high-frequency systems, the maximum cross-range resolution is often intentionally reduced to attain adequate depth of focus. Annular arrays with dynamic receive focusing (DRF) can extend the depth of focus at lower *f*-number, however the receive channel count must increase from 1 to about 8, and the improvement is often marginal. Alternative beamforming techniques,[2,3] including spatial matched filtering (SMF), have the potential to further increase depth of focus using low *f*-number, fixed-focus apertures while also increasing the echo signal-to-noise ratio (eSNR).

Our group recently described a SMF beamformer for one-dimensional linear array imaging with a rectangular aperture.[4] We found that the SMF approach leads to improved eSNR, lateral resolution, and contrast resolution as compared with conventional DRF provided the pulse-echo impulse response can be accurately determined. The lateral resolution was found to be equivalent to several synthetic-aperture-focusing methods for one-dimensional arrays.[5]

Our goal with this letter is to extend the SMF beamforming strategies to circular symmetric transducers, and then evaluate them through comparison with conventional delay-sum beamforming methods. We examine a large-aperture, fixed-focus technique for pulse transmission in which received signals are matched filtered by the pulse-echo point spread function. While lateral resolution and eSNR improved compared with DRF methods, axial resolution is somewhat compromised due to the two-dimensional (2D) filter. We show that spatial matched filtering of individual receive-element signals using annular arrays generates lower sidelobes than does filtering beamformed echo signals. Thus contrast resolution may be improved at the cost of increasing the computational load. Spatial filtering offers the additional advantage of not requiring delay circuits in the beamformer, which simplifies the stringent hardware requirements for imaging with arrays particularly at high frequencies. We describe the conditions under which the SMF method offers an efficient beamforming solution for circularly symmetric apertures.

II. METHOD AND ANALYSIS

A. Conventional fixed-focus beamforming (FF)

Assume we transmit and receive a broadband acoustic pulse using spherically focused circularly symmetric apertures, $a_t(x_0,y_0)$ whose radius of curvature is z_F. The source-detector plane is defined by (x_0,y_0) and (x,y,z) represents field points. The complex pulse-echo pressure field at (x,z) and at radial temporal frequency $\omega=2\pi f$ can be obtained from the Rayleigh-Sommerfeld diffraction formula. The field, $\bar{\psi}_w(x,z)$, at wavelength λ and wave number $k=2\pi/\lambda$ is

$$\begin{aligned}\bar{\psi}_\omega(x,z) &= C^2\left[\int_{-\infty}^{+\infty} dy_0 e^{jk\beta y_0^2}\int_{-\infty}^{+\infty} dx_0 e^{-jk(x/z)x_0} a_t(x_0,y_0) e^{jk\beta x_0^2}\right]^2 \\ &= C^2\left[\int_{-\infty}^{+\infty} dx_0 e^{-jk(x/z)x_0} e^{jk\beta x_0^2} b(x_0)\right]^2 \\ &= C^2[\Im\{b(x_0)e^{jk\beta x_0^2}\}_{u_x=x/\lambda z}]^2, \end{aligned} \quad (1)$$

where

[a)]Electronic mail: mfi@uiuc.edu

 0001-4966/2007/121(4)/1852/4/$23.00

$$b(x_0) = \int_{-\infty}^{+\infty} dy_0 e^{jk\beta y_0^2} a_t(x_0, y_0), \quad C = \frac{e^{jkz}}{j\lambda z},$$

$\beta = 1/2z - 1/2z_F$, and $\Im\{\cdot\}_u$ is the spatial Fourier transform of the argument evaluated at frequency u. Equation (1) is the narrowband pulse-echo field for a conventional beamformer and one standard for comparison. The goal of the SMF beamformer is to improve cross-range resolution in the near and far field as compared with conventional fixed-focused beamforming in Eq. (1).

B. Spatial matched filtering

In the SMF method, we transmit and receive fixed-focus beams as in Eq. (1). However the received echoes are matched filtered. To simplify the result, we autocorrelate Eq. (1) along the lateral (x-axis) direction to find

$$\begin{aligned}
\bar{\bar{\psi}}_{\omega,S}(x,z) = \bar{\psi}_\omega(x,z) * \bar{\psi}_\omega^*(-x,z) &= C^4 \int_{-\infty}^{+\infty} d\tau \left[\int_{-\infty}^{+\infty} dy_0 e^{jk\beta y_0^2} \int_{-\infty}^{+\infty} dx_0 e^{-jk(\tau/z)x_0} a_t(x_0,y_0) e^{jk\beta x_0^2} \right]^2 \\
&\times \left[\int_{-\infty}^{+\infty} dy_1 e^{-jk\beta y_1^2} \int_{-\infty}^{+\infty} dx_1 e^{jk[(\tau-x)/z]x_1} a_t(x_1,y_1) e^{-jk\beta x_1^2} \right]^2 \\
&= C^4 \int_{-\infty}^{+\infty} dx_0 \left[e^{-jk(x/z)x_1} e^{jk\beta x_0^2} \int_{-\infty}^{+\infty} d\tau e^{-jk[(x_0-x_1)/z]\tau} \int_{-\infty}^{+\infty} dy_0 e^{jk\beta y_0^2} a_t(x_0,y_0) \int_{-\infty}^{+\infty} dy_1 e^{-jk\beta y_1^2} \int_{-\infty}^{+\infty} dx_1 a_t(x_1,y_1) e^{-jk\beta x_1^2} \right]^2 \\
&= C^4 \left(\frac{4\pi^2}{k} \right)^2 \int_{-\infty}^{+\infty} dx_0 \left[e^{-jk(x/z)x_0} \int_{-\infty}^{+\infty} dy_0 e^{jk\beta y_0^2} a_t(x_0,y_0) \int_{-\infty}^{+\infty} dy_1 e^{-jk\beta y_1^2} a_t(x_0,y_1) \right]^2 \\
&= C^4 \left(\frac{4\pi^2}{k} \right)^2 \int_{-\infty}^{+\infty} dx_0 e^{-jk(2x/z)x_0} [b(x_0) b^*(x_0)]^2 = C^4 \left(\frac{4\pi^2}{k} \right)^2 \Im\{|b(x_0)|^4\} \Big|_{u_x = 2x/\lambda z}.
\end{aligned} \tag{2}$$

If a rectangular aperture is substituted into Eq. (2), as with linear arrays,[4] the 2D aperture function is separable into lateral (x-axis) and elevational (y-axis) components, $a_t(x_0,y_0) = a_t(x_0)a_t(y_0)$, and Eq. (2) simplifies to the spatial Fourier transform of $|a_t(x_0)|$.[4] Because circular apertures are not separable in the same way, the SMF result is proportional to $|b(x_0)|^4$ that includes an integral of the aperture times a quadratic phase factor.

The complex phase factor $\exp(jk\beta y_0^2)$ converts the normally constant aperture function a into a rapidly oscillating function. This feature contributes to sidelobe energy and beam spreading that correspond to high frequency components in the spatial Fourier domain. Despite these limitations, the SMF method improves in-plane, cross-range resolution (often called lateral resolution) over an extended depth of field. Equation (2) shows that spatial matched filtering improves lateral resolution by eliminating the lateral quadratic phase factor $\exp(jk\beta x_0^2)$ from the transmit-receive field at all depths. The inseparability of the aperture in the (x_0, y_0) plane does not provide the closed-form solution that allows us to clearly predict the effects of filtering with circular apertures. Nevertheless, the numerical and experimental results presented in the following show that matched filtering clearly improves beam quality.

It is significant that the spatial frequency variable for the SMF beamformer in Eq. (2) is scaled by a factor of 2 as compared with the FF aperture in Eq. (1). The factor of 2 means that the SMF applied to beamformed echo data has a narrower lateral main-lobe width, like synthetic aperture focusing techniques.[5]

C. Annular array imaging using spatial matched fltering

Rf data from the entire aperture are being filtered in the previous section. In this section, however, we filter receive-channel echoes from annuli before summation. A focused array aperture $a_t(x_0,y_0)$ transmits sinusoids at frequency ω. The filtered pulse-echo field from the nth element, $\psi_w(x,z;n)$, becomes

$$\psi_w(x,z;n) = \varphi_w(x,z)\varphi_w'(x,z;n), \tag{3}$$

where $\varphi_w(x,z)$ represents the transmit field using the whole aperture and $\varphi_w'(x,z;n)$ is the receive field from just the nth element. We now matched filter the receive fields along the x axis. The field is evaluated at the corresponding element before summation and without applying time delays. The pulse-echo field for the SMF beamformer, where the echoes are filtered and then summed, can be expressed by

$$\bar{\psi}_{\omega,S}(x,z) = \sum_{n=1}^{N} \psi_w(x,z;n) * [\psi_w(-x,z;n)]^*, \tag{4}$$

where N is the number of receiver elements in the annular array.

To compare the characteristics of SMF before and after summation, we rewrite the pulse echo field of SMF of Eq. (2) as

$$\begin{aligned}\bar{\bar{\psi}}_{w,S}(x,z) &= \left[\sum_{n=1}^{N}\psi_w(x,z;n)\right] * \left[\sum_{n=1}^{N}\psi_w(-x,z;n)\right]^* \\ &= \sum_{n=1}^{N}\psi_w(x,z;n) * [\psi_w(-x,z;n)]^* \\ &\quad + \sum_{l=1}^{N}\sum_{m=1,m\neq l}^{N}\psi_w(x,z;l) * [\psi_w(-x,z;m)]^* \\ &= \bar{\bar{\psi}}_{\omega,S}(x,z) + \sum_{l=1}^{N}\sum_{m=1,m\neq l}^{N}\psi_w(x,z;l) * \\ &\quad \times [\psi_w(-x,z;m)]^*. \end{aligned} \tag{5}$$

Comparing Eqs. (4) and (5), we see that if the echoes are filtered before summation, the cross terms between different elements are eliminated, which can be expected to concentrate beam energy and improve spatial resolution.

III. SIMULATION AND EXPERIMENTAL RESULTS

The above- mentioned predictions of beamformer performances were validated using rf echo simulations from Field II[6] for an annular array transducer composed of eight elements, each with equal area. The center frequency of the transmitted pulse was 10 MHz and the −6 dB bandwidth was 6 MHz. And diameter and geometric focal length of the array are 30 and 45 mm, respectively, resulting in an $f/1.5$ geometric aperture. Simulation parameters were the same as those of the associated experiment.

In the following results, "FF" beam plots describe patterns obtained when transmit and receive apertures are both focused at 45 mm depth. "DRF" plots describe results for a fixed 45 mm focal length on transmit and a dynamically focused receive aperture. "SMF(A)" and "SMF(B)" are results for SMF after the summing and SMF before element summation, respectively.

Figure 1 shows broadband, 2D, in-plane, pulse-echo point spread functions for each method. Simulation and experimental results were very similar except for noise. After beam formation, the point spread functions are envelope detected, the amplitudes are normalized to the peak values, and then log compressed with 50 dB dynamic range. The SMF results are from a 2D spatial matched filter applied in the $x-z$ plane.

As seen from Fig. 1, the broadband pulse energy is more spatially compact with the SMF method compared to the FF and DRF method in the near field and far field. The PSFs were all very similar at the transmit focal depth of 45 mm, except for the expected loss of axial resolution for the SMF methods from 2D filtering. There is generally more background noise in the experimental results, and the SMF(B) methods are most effective at suppressing that noise.

Complete comparisons of the three methods must extend beyond point reflectors to include scattering fields that generate speckle and have low-contrast targets. Figure 2 shows experimental results using a cyst phantom with 2D anechoic targets, where the center of the anechoic region was positioned anywhere from 35 to 55 mm. The SMF(B) approach provides the greatest lesion visibility.

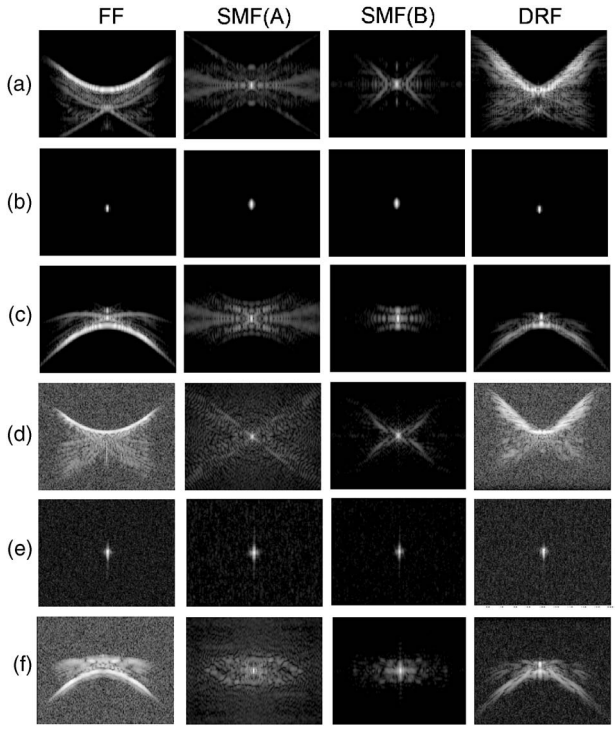


FIG. 1. Impulse response estimates are imaged using different beamformers. Simulation results (a)–(c) and experimental results (d)–(f) for a broadband, pulse-echo, 2D point spread functions are compared for the FF, DRF, and SMF beamformers at (a), (d) 25 mm depth, (b), (e) 45 mm depth, and (c), (f) 65 mm depth. The transmit focus is fixed at 45. Images are normalized to the individual peak values and displayed with 50 dB dynamic range.

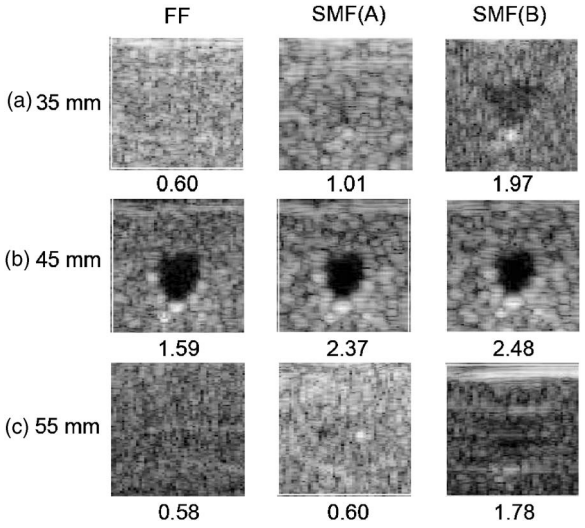


FIG. 2. Experimental images using a cyst phantom that were processed using FF and SMF methods. The transmit focus is fixed at 45 mm. The cyst diameter is 1.5 mm and the medium has a constant speed of sound. All images are displayed with 50 dB dynamic range. Depths of cyst centers are indicted in the left column. CNR values appearing below each image are computed using $\mathrm{CNR}=|\langle S_i\rangle-\langle S_o\rangle|/\sqrt{\sigma_o^2+\sigma_i^2}$, where $\langle S_{i,o}\rangle$ and $\sigma_{i,o}^2$ are the mean and variance of image pixels inside and outside the target.

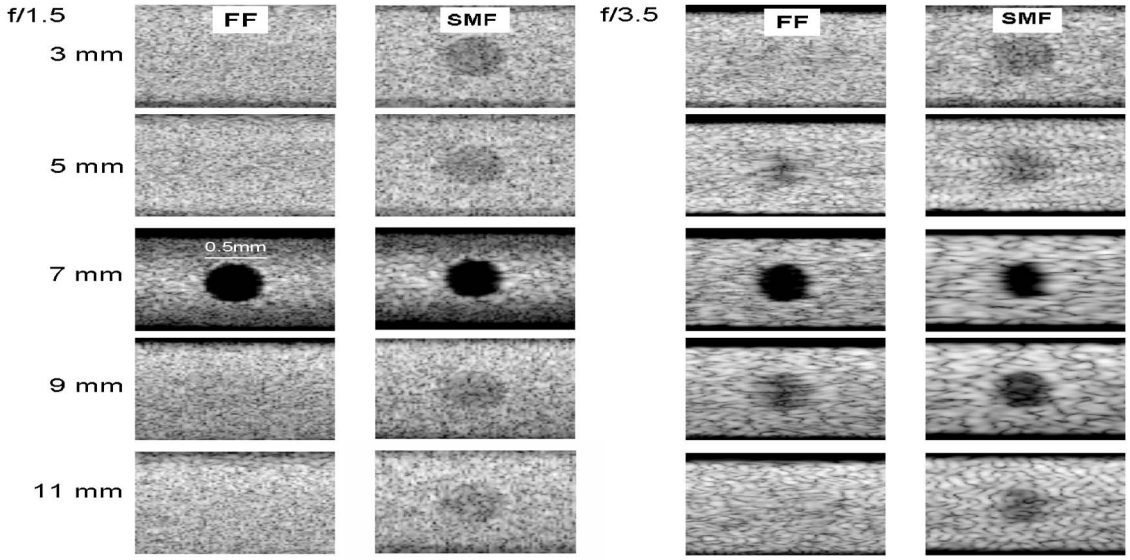


FIG. 3. The images are from echo simulations of a cyst phantom at 70 MHz center frequency that were processed using FF and SMF methods. The transmit focus is set to 7 mm. The cyst diameter is 0.5 mm and the medium has a constant speed of sound. All images are displayed with 50 dB dynamic range. Depths of cyst centers are indicated on the left.

These results differ from those of a linear array.[4] In our previous study using commercial ultrasound system, we could not completely control aperture growth. Thus the impulse response function was only approximately known. The associated errors degraded lateral resolution when SMF was applied before beamforming. Moreover, the assumption of a shift-invariant impulse response along the lateral axis of the image was violated in practice for linear arrays because the length of the subaperture varies with lateral position. Such limitations were not a factor with annular arrays because we had fine control over each parameter and the aperture was truly fixed. Thus we could obtain more accurate impulse response functions for filtering, which led to measurement results being more consistent with theoretical and simulation results.

One promising application of SMF can be high frequency imaging. At transmission frequencies over 50 MHz, it is very difficult to implement digital dynamic focusing since very high-speed digital circuits are required. SMF requires only 2D FIR filters for focusing, so that receive focusing can be quickly and efficiently implemented. To evaluate the performance of spatial filtering for high frequency application, we employed Field II simulations for a 70 MHz (45% BW), single element, circular transducer with a focal depth of 7 mm.

Figure 3 shows computer generated cyst phantom images that were processed using conventional FF and SMF method at each depth. Simulations were performed using two apertures which have different f-numbers: $f/1.5$ and $f/3.5$. The SMF method provides much improved spatial and contrast resolution in the near- and far-field regions compared with fixed focusing, suggesting greater depth of focus without compromising lateral resolution.

IV. CONCLUSIONS

SMF can effectively focus the ultrasound beam. The resulting pulse-echo point spread function yields superior lateral resolution compared to conventional FF or DRF except near the transmit focal length where they are comparable. Also, in the case of array transducers, applying the SMF before summation enhances lateral resolution more than applying it to beamformed data.

ACKNOWLEDGMENT

This material is based upon work supported by the National Cancer Institute under Award Nos. R01 CA082497 and R01 CA118294.

[1]J. A. Brown and G. R. Lockwood, "A digital beamformer for high-frequency annular arrays," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1262–1269 (2005).
[2]G. R. Lockwood, D. H. Turnbull, D. A. Christopher, and F. S. Foster, "Beyond 30 MHz—Applications of high frequency ultrasound imaging," IEEE Eng. Med. Biol. Mag. **15**, 60–71 (1996).
[3]M.-L. Li, W.-J. Guan, and P.-C. Li, "Improved synthetic aperture focusing technique with applications in high-frequency ultrasound imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 63–70 (2004).
[4]K. S. Kim, J. Liu, and M. F. Insana, "Efficient array beamformer by spatial filtering for ultrasound B-mode imaging," J. Acoust. Soc. Am. **120**, 852–861 (2006).
[5]M. Karaman, P.-C. Li, and M. O'Donnell, "Synthetic aperture imaging for small scale systems," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 429–442 (1995).
[6]J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped apodized and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).

A beam summation algorithm for wave radiation and guidance in stratified media

Tal Heilpern, Ehud Heyman,[a] and Vadim Timchenko
School of Electrical Engineering, Tel Aviv University, Tel Aviv 69978, Israel

(Received 21 June 2006; revised 15 December 2006; accepted 12 January 2007)

A Gaussian beams summation (GBS) algorithm for tracking source excited wave fields in plane stratified media is presented. In the present application the medium is described by layers with constant gradient of the wave speed, and the GB propagators are calculated recursively in a closed form. The algorithm is calibrated for numerical efficacy and accuracy by defining simple physical criteria for choosing the expansion parameters (the beam collimation and the spectral discretization and truncation) that allow for sparse representation of the source-excited angular spectrum of beams. It is validated for a source-excited example in layered media, where it provides a smooth and physically meaningful solution under multipath and caustic conditions and remains accurate for long propagation ranges where phase error tends to accumulate. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2537221]

PACS number(s): 43.20.Bi, 43.30.Dr, 43.30.Cq, 43.20.Dk [JJM] Pages: 1856–1864

I. INTRODUCTION AND PROBLEM SCOPE

Gaussian beam summation (GBS) formulations are an important tool in wave theory since they provide a framework for ray-based construction of spectrally uniform local solutions in complex configurations. In these formulations, the field is expanded into a spectrum of GB that emanate at a given set of points and directions in the source domain, and thereafter are tracked locally in the medium. Their main advantages are *spectral localization* (the summation involves only the beams that pass near the observation point) and spectral uniformity (the beam propagators are insensitive to ray catastrophes, like caustics and foci). Thus, the GBS approach, when applied judiciously, combines the uniform features of the spectral representation with the algorithmic ease of ray representation (see recent reviews in Refs. 1–3).

There are essentially two classes of GBS representations for point source configurations[4–13] and for distributed (or aperture) source configurations.[14–16] This paper utilizes the former, but may readily be extended to the latter when presented in its latest ultrawideband format.[16] The GBS is also related to other local spectral marching approaches, such as Ref. 17.

The GBS formulation for point source fields was originally derived asymptotically,[4,5] and utilized in various disciplines and applications.[6–11] In Ref. 12 it has been placed on a rigorous footing by showing that the field of an isotropic point source in uniform medium can be expressed *exactly* as a superposition of "complex source beams" (CSBs) that emerge from the source in all directions. This *expansion identity* has been extended in Ref. 13 to the time domain where the propagators are pulsed beams (see also Ref. 2).

In this work we present a GBS algorithm for calculating radiation and propagation in stratified configurations, described by layers with constant gradient (CG) of the wave speed. The CG model allows for wider layers compared with the conventional uniform layers model (thereby reducing the algorithm complexity), and eliminates the spurious reflection at the layer interfaces that characterize the latter. The GB propagators in this medium are calculated recursively in a closed form. The algorithm is calibrated for efficacy and accuracy, and it is validated numerically for a source-excited example in layered medium, where it provides a smooth and physically meaningful solution under multipath and caustic conditions and remains accurate for long propagation ranges where phase error tends to accumulate. The GBS algorithm can be extended to range dependent configurations, and can also accommodate rough surface boundaries by utilizing the GB to GB rough surface scattering matrix of the surface.[18,19]

The presentation starts in Sec. II with the GBS in free space, emphasizing the parametrization of the errors due to the spectral discretization and truncation and due to the use of the GB propagators instead of the exact CSB. The algorithm is calibrated and the expansion parameters (the beam collimation and the spectral discretization and truncation) are optimized for a sparse discretization of the source-excited angular spectrum of beams within a given error level. We then proceed with a brief summary of the basic principles of the GBS in a general inhomogeneous medium (Sec. III), and then present the algorithm for the special case of a layered medium with CG wave speed layers (Sec. IV) and demonstrate its accuracy (Sec. V).

II. GAUSSIAN BEAM SUMMATION REPRESENTATIONS FOR A POINT SOURCE IN FREE SPACE

In this section we present and calibrate the GBS representation for a point source in a free space. We start with the exact expansion identity expressed as a continuous spectrum of CBSs and then explore the errors introduced when the CSB propagators are replaced by GBs and the spectral summation is discretized and truncated. A harmonic time dependence $e^{-i\omega t}$ is assumed and suppressed throughout.

[a]Electronic mail: heyman@eng.tau.ac.il

 0001-4966/2007/121(4)/1856/9/$23.00

A. Complex source beams and Gaussian beams in free space

The CSB is modeled by extending the real source point $\mathbf{r}'$ in the conventional Green's function point source solution to the complex domain. The complex source point can be expressed in the most general form as

$$\mathbf{r}' = \mathbf{r}_0 + i\mathbf{b}. \tag{1}$$

As will be shown in the following, the real point $\mathbf{r}_0$ is the center of the beam waist, the real vector $\mathbf{b}$ defines the beam direction, and its magnitude $b \equiv |\mathbf{b}|$ is the beam collimation distance. For a given observation point $\mathbf{r}=(x,y,z) \in \mathbb{R}^3$, the "complex distance" from $\mathbf{r}'$ is defined as

$$s(\mathbf{r}) = \sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2} \tag{2a}$$

$$= \sqrt{\eta_1^2 + \eta_2^2 + (\sigma - ib)^2}, \quad \mathrm{Re}\, s > 0, \tag{2b}$$

where Eq. (2a) is the general definition, while Eq. (2b) is expressed in terms of the beam coordinates $(\sigma, \boldsymbol{\eta}) = (\sigma, \eta_1, \eta_2)$ of $\mathbf{r}$ defined such that σ extends from $\mathbf{r}_0$ along the beam axis $\mathbf{b}$ while $\boldsymbol{\eta}$ are the transversal coordinates. The single-value function $s(\mathbf{r})$ is obtained by setting $\mathrm{Re}\, s > 0$ for all $\mathbf{r} \in \mathbb{R}^3$. This introduces a branch cut where from Eq. (2b) the set of branch points is the circle $\sqrt{\eta_1^2 + \eta_2^2} = b$ in the $\sigma = 0$ plane and the cut is the flat disk $\sqrt{\eta_1^2 + \eta_2^2} < b$ in that plane, which is referred to as the "source disk."[1,2]

The field due to the source in Eq. (1) in a uniform medium with wave speed c is obtained as an analytic extension of the real source solution, i.e.,

$$B_{\mathrm{CSB}}(\mathbf{r};\mathbf{r}') = (-ibe^{-kb})\frac{e^{iks(\mathbf{r})}}{s(\mathbf{r})}, \quad k = \omega/c, \tag{3}$$

where for subsequent use we added the normalization factor $-ibe^{-kb}$, so that this solution reduces to the *normalized* GB in Eq. (5). In view of the properties of $s(\mathbf{r})$, the field in Eq. (3) is a *globally exact* solution of the wave equation that propagates outward from the source disk. It behaves like a beam field that propagate along the positive σ axis and decays exponentially away to minimum along the $-\sigma$ axis. Further properties of this solution are discussed, e.g., in Refs. 1 and 2. Here, we are mainly interested in the paraxial zone $|\boldsymbol{\eta}| \ll \sqrt{\sigma^2 + b^2}$ near the positive σ axis, where from Eq. (2b)

$$s \simeq \sigma - ib + |\boldsymbol{\eta}|^2/2(\sigma - ib) - |\boldsymbol{\eta}|^4/8(\sigma - ib)^3 + \cdots,$$

$$\sigma > 0. \tag{4}$$

Substituting into Eq. (3) and keeping terms to second order only, we obtain there

$$B_{\mathrm{CSB}} \simeq \frac{-ib}{\sigma - ib}\exp\left\{ik\left[\sigma + \frac{1}{2}\frac{|\boldsymbol{\eta}|^2}{\sigma - ib}\right]\right\}$$

$$\equiv B_{\mathrm{GB}}(\mathbf{r};\mathbf{r}_0,\hat{\boldsymbol{\sigma}},b). \tag{5}$$

Expression (5) has the standard form of a stigmatic (circular symmetric) GB in free space, emerging from $\mathbf{r}_0$ in the direction $\hat{\boldsymbol{\sigma}}$ with collimation parameter b, where the caret denotes a unit vector. To parametrize this expression we define $1/(\sigma - ib) = 1/R(\sigma) + i/kW^2(\sigma)$ with

$$R(\sigma) = \sigma + b^2/\sigma, \quad W(\sigma) = W_0\sqrt{1 + (\sigma/b)^2} \text{ with}$$

$$W_0 = \sqrt{b/k}. \tag{6}$$

Substituting in the exponent of Eq. (5) one identifies $R(\sigma)$ and $W(\sigma)$ as the wave front radius of curvature and the beamwidth (the $e^{-1/2}$ half-diameter) along the σ axis with b being the collimation or (Rayleigh) length: For $\sigma \ll b$, $W(\sigma) \approx W_0$ and the GB propagates essentially without decay or spreading. For $\sigma \gg b$, it diverges along a cone of angle

$$\theta_D \simeq W(\sigma)/\sigma \simeq \sqrt{1/kb}, \tag{7}$$

hence collimation requires $kb \gg 1$. The paraxial GB approximation in Eq. (5) is valid only if the phase error due to the omission of the fourth order term in Eq. (4) is $\ll 1$ within the main beam region. For $\sigma \gg b$ we use $|\boldsymbol{\eta}| \sim \sigma\theta_D$, obtaining the validity condition

$$\sigma \ll kb^2. \tag{8}$$

B. Beam expansion identities in uniform medium

We consider the field $g_s = e^{ikR}/R$, $R = |\mathbf{r} - \mathbf{r}_s|$ due to a real point source at $\mathbf{r}_s = (x_s, y_s, z_s)$ in free space. The *exact* beam expansion identity is

$$g(\mathbf{r},\mathbf{r}_s) = \frac{ki}{2\pi}\frac{1}{1 - e^{-2k\beta}}\int_{4\pi} d\Omega B_{\mathrm{CSB}}(\mathbf{r};\mathbf{r}' = \mathbf{r}_s + i\hat{\boldsymbol{\sigma}}(\Omega)\beta),$$

$$R > |\beta|, \tag{9}$$

where $\hat{\boldsymbol{\sigma}}(\Omega)$ is a unit vector in the spherical angle direction $\Omega = (\theta, \phi)$ about $\mathbf{r}_s$ and $\beta = b + i\beta_i$ with $b > 0$ is a complex parameter. Equation (9) expresses the field as a continuum of CSBs that emerge from $\mathbf{r}_s$ in all 4π spherical directions. Referring to Eq. (1) and the discussion following Eq. (3), each beam propagates away from $\mathbf{r}_s$ in the $\hat{\boldsymbol{\sigma}}(\Omega)$ direction with collimation distance b and waist at the *real* point $\mathrm{Re}\,\mathbf{r}' = \mathbf{r}_s - \hat{\boldsymbol{\sigma}}(\Omega)\beta_i$, i.e., at the point $\sigma = -\beta_i$ along the axis of that beam. Choosing $\beta_i = 0$ implies that the waists of these beams are located at $\mathbf{r}_s$. In certain cases [see the discussion following Eq. (12)] one may prefer to use $\beta_i < 0$ so that the propagators converge as they emerge from $\mathbf{r}_s$ to a waist at $\sigma = -\beta_i$ away from $\mathbf{r}_s$ and then diverge. Finally the constraint $R > |\beta|$ in Eq. (9) is due to the "source disk" singularity discussed after Eq. (2), hence this restriction can be removed when the CSBs are replaced by their GB approximation in Eq. (5) as done in Eq. (10b). A proof of Eq. (9) is given in Ref. 12. A time domain expansion identity using pulsed beam (PB) waves is developed in Refs. 13 and 2. These references also present an alternative proof that demonstrates the convergence rate of the expansion.

The identity in Eq. (9) applies for any kb, but henceforth we only consider collimated propagators with $kb \gg 1$. In this case, only those beams that propagate near the observation direction from $\mathbf{r}_s$ to $\mathbf{r}$ contribute there, hence the beam spectrum may be truncated to narrow cone of angle ϑ about this direction where ϑ will be determined in the following. Noting, in addition that $e^{-2k\beta} \to 0$ here, we obtain from Eq. (9)

$$g_s \simeq \frac{ki}{2\pi}\int_{\vartheta} d\Omega B_{\mathrm{CSB}}(\mathbf{r};\mathbf{r}_s + i\hat{\boldsymbol{\sigma}}(\Omega)\beta) \tag{10a}$$

$$\simeq \frac{ki}{2\pi}\sum_j \delta\Omega_j B_{\mathrm{GB}}(\mathbf{r};\mathbf{r}_s,\hat{\boldsymbol{\sigma}}_j,\beta), \tag{10b}$$

where in Eq. (10b), the integral is discretized in terms of a lattice of beam directions $\hat{\boldsymbol{\sigma}}_j$ associated with spherical angle differentials $\delta\Omega_j$, and the exact CSBs are replaced by GBs that emerge from $\mathbf{r}_s$ in the $\hat{\boldsymbol{\sigma}}_j$ direction. The latter are given by Eq. (5) with b replaced by $\beta=b+i\beta_i$, i.e., they have the same properties as the GBs in Eq. (5) except that the waists are now located at $\sigma=-\beta_i$. The replacement in Eq. (10b) of the CSBs by the GBs removes the source disc singularity and hence the $R>|\beta|$ constraint in Eq. (9), yet, as will be demonstrated in the following, the convergence deteriorates for $R \ll b$.

C. Calibration of the beam expansion scheme

In the following we calibrate the beam expansion (10b) and parametrize the tradeoff between accuracy and numerical efficacy (i.e., the number of beams needed to model the field). The truncation error can be parametrized by evaluating the integral representation in Eq. (10a) in closed form. Introducing the polar and azimuthal angles (θ', ϕ') about the observation direction we obtain

$$\begin{aligned}
&\frac{ki}{2\pi}\int_{\vartheta} d\Omega B_{\mathrm{CSB}}(\mathbf{r};\mathbf{r}_s + i\hat{\boldsymbol{\sigma}}(\Omega)\beta) \\
&= \frac{k\beta e^{-k\beta}}{2\pi}\int_0^{2\pi} d\phi' \int_0^{\vartheta} d\theta' \sin\theta' \frac{e^{iks(\theta')}}{s(\theta')} \\
&= \frac{-ki}{R}e^{-k\beta}\int_{s(0)}^{s(\vartheta)} ds\, e^{iks} = \frac{e^{ikR}}{R} - \frac{e^{ik(i\beta+s(\vartheta))}}{R},
\end{aligned} \tag{11}$$

where in the first equality we inserted B_{CSB} from Eq. (3) with b replaced by β and we also expressed $s(\mathbf{r})$ as $s(\theta') = \sqrt{R^2+(i\beta)^2-2i\beta R\cos\theta'}$. In the second equality we noted that the ϕ' integral yields 2π and replaced the integration variable θ' by $s(\theta')$. The final result follows from a closed form evaluation of that integral, using also $s(\theta'=0) = R-i\beta$.

The first term in the final result in Eq. (11) is the exact field, hence the second term is the error due to the ϑ truncation. Since we use collimated beams, ϑ is small and the observation point $\mathbf{r}$ is in the paraxial zone of all the beams in the expansion, including the most off-axis one in the ϑ direction. We may therefore use $s(\vartheta) = \sqrt{(R-i\beta)^2+2i\beta R(1-\cos\vartheta)} \simeq R-i\beta+i\beta R\vartheta^2/2(R-i\beta)$ and hence the relative error in Eq. (11) is

$$\begin{aligned}
\varepsilon &= \left|e^{ik(i\beta+s(\vartheta))}\right| \simeq \left|e^{-(1/2)kR\beta\vartheta^2/(R-i\beta)}\right| \\
&= e^{-(1/2)kbR^2\vartheta^2/[(R+\beta_i)^2+b^2]} = e^{-(1/2)\eta_\vartheta^2/W^2(R)} \to e^{-(1/2)\vartheta^2/\theta_D^2},
\end{aligned} \tag{12}$$

where in the last equality $\eta_\vartheta = R\vartheta$ is the η coordinate of $\mathbf{r}$ corresponding to the last (i.e., the ϑ) beam in the summation and $W(R)=\sqrt{b/k}\sqrt{1+(R+\beta_i)^2/b^2}$ is the beamwidth at R [see Eq. (6) with σ replaced by the distance $R+\beta_i$ from the beam-waist]. Finally the last expression is the limiting case in the far zone $R \gg b$. Thus, for a given error level, the spectral truncation angle ϑ is determined via

$$\eta_\vartheta \approx \chi_a W(R), \ \to \vartheta \approx \chi_a \theta_D, \tag{13}$$

where χ_a is an error parameter to be chosen: for a 1% error it is sufficient to use $\chi_a \approx 3$. The second condition in Eq. (13) is a simplification of the first one for $R \gg b$. As will be demonstrated in Fig. 1(a), for $R \gg b$ the condition in (13) requires a rather narrow spectral spread of beams ϑ which is independent of R, but for $R<b$, ϑ and thereby the number of beams in the summation increase as R decreases.

The parameters b and β_i are chosen now to minimize the error at a given range R, and also to minimize the beamwidth along the propagation trajectory. The optimal choice is $b=-\beta_i=R/2$ (i.e., the beam waist is placed half way between $\mathbf{r}_s$ and $\mathbf{r}$, and both $\mathbf{r}_s$ and $\mathbf{r}$ are at the Rayleigh distance of the beams). In the present work, however, the beams waists are taken to be at $\mathbf{r}_s$ (i.e., $\beta_i=0$) hence we choose $b \sim R$ where R is a representative value of the range. This choice optimizes between two requirements: It is desired that the beams will be collimated and narrow at the observation point, leading to $b \gtrsim R$, but it is also desired to have a narrow spectral spread of beams, leading from Eq. (13) to $b \lesssim R$.

D. Numerical implementation: A hierarchial scheme for a wide observation range

The lattice of beam directions in Eq. (10b) is formed by dividing the unit sphere into a mesh of triangles with common nodes. The corners define the beam directions $\hat{\boldsymbol{\sigma}}_j$ and the spherical angle associated with the jth beam is

$$\delta\Omega_j = \frac{1}{3}\sum_{j'} \Delta_{j'}, \tag{14}$$

where $\Delta_{j'}$ are the areas of the triangles associated with the jth node. We shall not dwell here on meshing techniques to obtain approximately uniform mesh. Typically there are six triangles per node, and hence the average beam discretization angle $\delta\Omega$ is roughly twice the average triangle size Δ. $\delta\Omega$ can be parametrized by the spherical angle spot-size of the beam, defined by [see Eq. (7)]

$$\Theta_D = \pi\theta_D^2 = \pi/kb. \tag{15}$$

In constructing the mesh we therefore require

$$\max \delta\Omega_j \leqslant \chi_b \Theta_D, \tag{16}$$

where χ_b is a proportionality parameter. From the below-presented numerical experiments it follows that using $\chi_b=1$ yields a 1% error and that a finer discretization (smaller χ_b) does not improve the result significantly. Combining Eq. (16) with the second condition in Eq. (13) one finds that the number of beams that are required to synthesize the ray field in the range $R>b$ is given by $\pi\vartheta^2/\chi_b\Theta_D = \chi_a^2/\chi_b$ [see also Fig. 1(b)].

The above-mentioned considerations are utilized now to construct a *hierarchial* GBS scheme that accommodates a

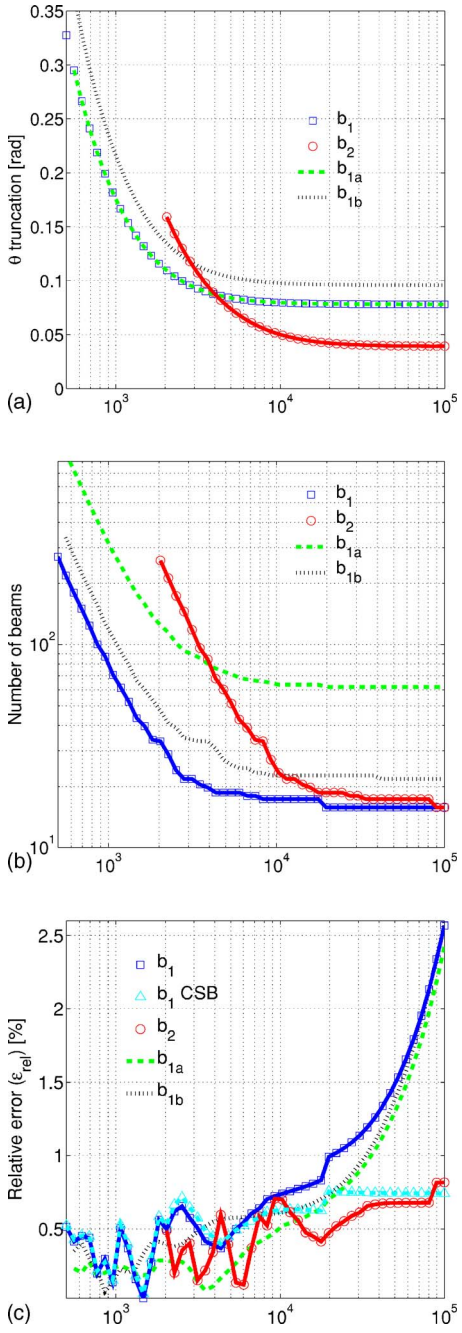


FIG. 1. (Color online) Performance of the GBS as a function of R, for the four realizations denoted in the text as cases (b_1), (b_2), (b_{1a}), and (b_{1b}) where the expansion parameters are: Case (b_1): $b_1=2\times10^3$, $(\chi_a,\chi_b)=(3.5,0.94)$ giving $\delta\Omega_1=1.48\times10^{-3}$. Case (b_2): $b_2=8\times10^3$, $(\chi_a,\chi_b)=(3.5,0.94)$ giving $\delta\Omega_2=3.70\times10^{-4}$. Case (b_{1a}): As in case (b_1) but with $\chi_b=0.23$ giving $\delta\Omega=\delta\Omega_2=3.70\times10^{-4}$. Case (b_{1b}): As in case (b_1) but with $\chi_a=4.3$. (a) The number of beams. (b) The spectral spread ϑ. (c) The relative error. We also show case $(b_{1\mathrm{CSB}})$ which corresponds to case (b_1) but with the GBs replaced by the CSBs.

wide range of R. As noted earlier, larger R requires larger b and hence, from Eq. (16), a finer mesh. It is therefore required to use different meshes (different beam lattices) for different observation ranges. In order to reduce the calculation time needed for tracking the beam propagation in the medium, we construct a self-consistent hierarchy of meshes so that the coarse lattices for shorter R are decimated versions of the finest lattice for the largest R. This implies that only one set of beams for the finer lattice needs to be tracked in the medium, whereas for the coarser lattices, one may use properly decimated subsets.

Following this approach we divide the propagation range $(R_{\min},R_{\max})$ into two-octave range bins $R\in[R_{\min};4R_{\min}]$, $R\in[4R_{\min};16R_{\min}]$,…, and assign them the indexes $i=1,2,\ldots$, respectively. For the shortest $i=1$ range bin, we choose $b_1=R_{\min}$ and then calculate the coarser mesh according to Eq. (16) with some chosen value of χ_b. For the $i=2$ range bin we then choose $b_2=4R_{\min}$ and create a finer mesh by subdividing each edge of the triangle into two, thus dividing each triangle into four. The new mesh complies, again with Eq. (16), for the same value of χ_b. This process goes on and in the ith range bin we use $b_i=2^{2(i-1)}R_{\min}$ and a fine mesh whose triangle areas are $2^{2(i-1)}$ times smaller than those for the smallest range bin.

The performances of the GBS formulation are explored in Fig. 1. Following the above-outlined procedure we consider two observation ranges: $R\in[2\times10^3,8\times10^3]$ and $R\in[8\times10^3,32\times10^3]$ wherein we use, respectively, $b_1=2\times10^3$ and $b_2=8\times10^3$. Assuming $k=1$ we find via Eq. (15) with b_1 that $\Theta_{D_1}=1.57\times10^{-3}$ and hence using Eq. (16) we construct a mesh with max $\delta\Omega_1=1.48\times10^{-3}$ (i.e., $\chi_{b_1}=0.94$). For b_2 we have $\Theta_{D_2}=3.93\times10^{-4}$, hence we divide each of the triangles in the former mesh into four as described earlier, obtaining max $\delta\Omega_2=3.70\times10^{-4}$, which again complies with Eq. (16) for the same value of χ_b. The spectral truncation in both cases has been determined via Eq. (13) with $\chi_a=3.5$.

Figure 1 explores the properties of the GBS formulation by using the beam sets b_1 and b_2 to calculate the field not only within the range bins they have been designed for, but in entire range $R\in[10^2,10^5]$. In order to determine the role of the various parameters, we also show in Fig. 1 the results for the beams b_1 but with the four times denser mesh (actually we use the mesh $\delta\Omega_2$ discussed earlier), as well as the results of the same beams b_1 with the original mesh $\delta\Omega_1$ but with a wider spectral truncation using $\chi_a=4.3$. These results are denoted by curves (b_{1a}) and (b_{2b}), respectively.

Figures 1(a) and 1(b) show the spectral truncation angle ϑ and the number of beams used in the summation as a function of R and for all four cases described earlier. In all cases, ϑ and the number of beams increase dramatically for $R<b$, as follows from the first condition in Eq. (13). For $R>b$, the number of beams needed to describe the field is small, and complies with the estimate χ_a^2/χ_b discussed after Eq. (16). For the "optimal" beam-lattices in cases (b_1) and (b_2) where $(\chi_a,\chi_b)=(3.5,0.94)$ one observes that this number is indeed ~13, whereas in cases (b_{1a}) and (b_{1b}) where $(\chi_a,\chi_b)=(3.5,0.23)$ and (4.3, 0.94), respectively, the number is larger because of the denser lattice in case (b_{1a}) and the wider spectral range in case (b_{1b}).

Figure 1(c) depicts the expansion error as a function of R. In cases (b_1) and (b_2) the error is less than 1% not only in the relevant ranges they were designed for, but also for $R<b$ where the expansion efficiency decays (see the discussion in Figs. 1(a) and 1(b)). There is, however, an error buildup at large distances $R\gtrsim20b$ which are beyond the relevant range bin of a given b. This error is due to the the

phase error in the GB propagators discussed in Eq. (8), and hence it cannot be reduced by using a denser discretization lattice or a wider spectral range as in cases (b_{1a}) and (b_{1b}). It fully disappears if the GB propagators are replaced by CSBs as shown by curve $b_{1\mathrm{CSB}}$. In general, however, we prefer to use GBs since they can be tracked in inhomogeneous medium, as done in the following sections. The far zone error buildup for a fixed b has already been noted in Ref. 8, but it is shown here to be due to the GB approximation of the exact CSB propagators.

III. THE GBS IN GENERAL INHOMOGENEOUS MEDIUM

The GBS formulation of Sec. II can be extended to the case of a source in smoothly inhomogeneous medium with wave speed $c(\mathbf{r})$. The source at $\mathbf{r}_s$ is identified by the radiation pattern $f(\hat{\boldsymbol{\sigma}})$, $\hat{\boldsymbol{\sigma}}$ being a unit vector pointing away from $\mathbf{r}_s$. f is normalized such that the radiated field in free space is given by $f(\hat{\boldsymbol{\sigma}})e^{ikR}/R$. The GBS extension of Eq. (10b) is[4,5,7,8]

$$u(\mathbf{r}) = \frac{ki}{2\pi}\sum_j \delta\Omega_j f(\hat{\boldsymbol{\sigma}}_j) B_{\mathrm{GB}}(\mathbf{r};\mathbf{r}_s,\hat{\boldsymbol{\sigma}}_j,\beta). \tag{17}$$

As in Eq. (10b) $B_{\mathrm{GB}}(\mathbf{r};\mathbf{r}_s,\hat{\boldsymbol{\sigma}}_j,\beta)$ are GB propagators emerging from $\mathbf{r}_s$ in the direction $\hat{\boldsymbol{\sigma}}_j$, normalized such that near $\mathbf{r}_s$ they behave like B_{GB} in Eq. (5) (but with b replaced for generality by β). Following Eq. (13), the j summation involves all the beams passing within the three-beamwidth neighborhood of $\mathbf{r}$. These beams cover the spectral zone associated with the geometrical (i.e., Fermat) ray from $\mathbf{r}_s$ to $\mathbf{r}$. Under multipath conditions where there are several geometrical rays, the j summation involves the beams adjacent to each of these rays.

The beam propagators B_{GB} are affected by the local medium properties along the ray trajectories S_j emerging from $\mathbf{r}_s$ in the direction $\hat{\boldsymbol{\sigma}}_j$. Following Ref. 20 (see also Refs. 1, 2, 4, 5, 7, and 8) the procedure starts by defining an orthogonal ray-fixed coordinate system $(\sigma,\boldsymbol{\eta})=(\sigma,\eta_1,\eta_2)$, where σ is an arclength along S_j measured from $\mathbf{r}_s$. The transversal coordinates $\boldsymbol{\eta}$ are related to the conventional normal and binormal to S_j ($\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}_b=\hat{\boldsymbol{\sigma}}\times\hat{\mathbf{n}}$, respectively) via

$$\begin{pmatrix}\hat{\boldsymbol{\eta}}_1\\ \hat{\boldsymbol{\eta}}_2\end{pmatrix} = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}\begin{pmatrix}\hat{\mathbf{n}}\\ \hat{\mathbf{n}}_b\end{pmatrix}, \tag{18}$$

where the rotation angle is related to the torsion of S_j via $\partial_\sigma\varphi(\sigma)=\kappa(\sigma)$. The GB solution along S_j is given now by

$$B_{\mathrm{GB}}(\mathbf{r},\mathbf{r}_s;\hat{\boldsymbol{\sigma}}_1,\beta) = \left[\frac{c_0(\sigma)}{c_0(0)}\frac{\det\mathbf{Q}(0)}{\det\mathbf{Q}(\sigma)}\right]^{1/2} \times\exp\left[i\omega\int_0^\sigma\frac{d\sigma'}{c_0(\sigma')} + \frac{i\omega}{2c_0(\sigma)}\boldsymbol{\eta}^T\boldsymbol{\Gamma}(\sigma)\boldsymbol{\eta}\right], \tag{19}$$

where $c_0(\sigma)$ is the wave speed along S_j so that the integral in the exponent is simply the "axial delay." The second term in the exponent represents the paraxial phase correction. Here, the superscript T denotes the transpose of a vector and $\boldsymbol{\Gamma}(\sigma)$ is a 2×2 complex symmetrical matrix with $\mathrm{Im}\,\boldsymbol{\Gamma}(\sigma)$ positive definite. Consequently, the imaginary part of the quadratic form $\boldsymbol{\eta}^T\boldsymbol{\Gamma}(\sigma)\boldsymbol{\eta}=\eta_1^2\Gamma_{11}+2\eta_1\eta_2\Gamma_{12}+\eta_2^2\Gamma_{22}$ increases as $\boldsymbol{\eta}$ increases away from the beam axis, leading to the Gaussian decay of B_{GB}. $\mathrm{Im}\,\boldsymbol{\Gamma}$ is recognized as the *beam-envelope matrix* while $\mathrm{Re}\,\boldsymbol{\Gamma}$ is the *wave-front-curvature matrix*. Both $\mathrm{Im}\,\boldsymbol{\Gamma}$ and $\mathrm{Re}\,\boldsymbol{\Gamma}$ are, in general, astigmatic (noncircular symmetric), but since they are real and symmetric, they may be diagonalized by rotation about the σ axis, thus defining the principal axes of the beam envelope and of the wave front curvature, respectively. $\boldsymbol{\Gamma}(\sigma)$ is calculated by setting $\boldsymbol{\Gamma}=c_0(\sigma)\mathbf{P}\mathbf{Q}^{-1}$ and solving the set of two linear differential equations along S_j,

$$\mathbf{Q}' = c_0\mathbf{P}, \quad \mathbf{P}' = -c_0^{-2}\mathbf{C}_2\mathbf{Q} \tag{20}$$

subject to the initial conditions $\mathbf{Q}(0)=\mathbf{I}$ and $\mathbf{P}(0)=c_0^{-1}(0)\boldsymbol{\Gamma}(0)$, where $\mathbf{C}_2(\sigma)=\partial_{ij}c(\mathbf{r})|_{S_j}$ is the second transversal derivative matrix of the wave speed along the beam axis. It may be shown that the solution of Eq. (20) satisfies $\mathrm{Im}\,\boldsymbol{\Gamma}(\sigma)$ positive definite for all σ provided that $\mathrm{Im}\,\boldsymbol{\Gamma}(0)$ is positive definite. Finally in view of Eq. (5), the initial condition for the beams in Eq. (17) is $\boldsymbol{\Gamma}(0)=(i/\beta)\mathbf{I}$.

If the GB hits an interface of wave speed discontinuity, it launches reflected and transmitted GBs according to Snell's refraction law. These GB have the functional form of Eq. (19) with their initial axial amplitudes described by Fresnel (plane wave) reflection and transmission coefficients. The initial values of the $\boldsymbol{\Gamma}$ matrixes on the interface are determined by $\boldsymbol{\Gamma}$ of the incident GB via phase matching on the interface[20] (see also Refs. 1 and 2). For the algorithm of Sec. IV C we need to account not only for the wave speed discontinuity but also for the local wave speed gradient near the interface. Such a expressions are derived in the Appendix.

IV. THE GBS ALGORITHM IN STRATIFIED MEDIA

A. Constant-gradient layered medium

As mentioned in Sec. I, the present GBS algorithm for stratified media utilizes the closed form solution of a GB in a medium with a constant gradient of the wave speed. We therefore approximate the given medium by layers with constant gradient of the wave speed. Considering the plane stratified medium where the wave speed $c(z)$ depends only on the vertical coordinate z, we divide it into layers with constant gradient of $c(z)$, such that in the jth layer

$$c(z) \simeq c_j^+ + \alpha_j(z-z_j), \quad z_j < z < z_{j+1}, \tag{21}$$

where $c_j^+=c(z_j^+)$ and $\alpha_j=(c_{j+1}^- - c_j^+)/(z_{j+1}-z_j)$ is the wave speed gradient in this layer. Note that unless z_j is a true physical interface, the wave speed is continuous at z_j (i.e., $c_j^+=c_j^-$) since these points are chosen such that they provide a convenient model of the medium. The beam propagation within the jth layer is treated in Sec. IV B, while Sec. IV C deals with the transition through an interface between the layers.

B. GB tracking in a CG layer

The solution is similar to the dynamic ray solution in a CG medium, which may be found for example in Ref. 21, except that here the "wave front curvature matrix" $\mathbf{\Gamma}(\sigma)$ is complex. Without loss of generality we place the source on the z axis at $\mathbf{r}_s=(0,0,z_s)$. Due to the symmetry, the rays' trajectories remain in planes of constant ϕ and are described by the (ρ,z) coordinates, where (ρ,ϕ) are the radial and azimuthal coordinates about the z axis. The beam-coordinates (η_1,η_2) are chosen, conveniently, such that $\hat{\boldsymbol{\eta}}_1=\hat{\boldsymbol{\phi}}$ is normal to the ray-propagation plane ϕ=const, and $\hat{\boldsymbol{\eta}}_2=\hat{\boldsymbol{\sigma}}\times\hat{\boldsymbol{\eta}}_1$.

We consider a particular GB that enters the jth layer from "below" at $\mathbf{r}_j=(\rho_j,z_j)$ and an angle θ_j from the z axis. Its initial values at $\mathbf{r}_j$ are: σ_j=the accumulated ray path from the source, ψ_j=the phase, A_j=the axial amplitude and $\mathbf{\Gamma}_j$. The expressions in Eq. (19) can now be calculated in a closed form. A similar solution is obtained if the beam enters the layer from "above" at z_{j+1}.

Since the ray trajectories in plane stratified media satisfy Snell's law $c^{-1}(z)\sin\theta(z)$=const where θ is the ray angle from the z axis, we obtain in the jth layer,

$$\sin\theta(z)=(1+(z-z_j)\alpha_j/c_j)\sin\theta_j. \tag{22}$$

This ray may have a turning point at $z_t=z_j+(c_j/\alpha_j)[\csc\theta_j-1]$ in this layer if $z_t<z_{j+1}$. Substituting Eq. (22) into $\rho(z)=\rho_j+\int_{z_j}^{z}dz'\tan\theta(z')$ we find that the trajectory is a circular arc with radius $c_j/|\alpha_j|\sin\theta_j$,

$$(\rho-\rho_j-c_j/\alpha_j\tan\theta_j)^2+(z-z_j+c_j/\alpha_j)^2=(c_j/\alpha_j\sin\theta_j)^2, \tag{23}$$

whose center is at $\rho_j+c_j/\alpha_j\tan\theta_j$ and z_j-c_j/α_j (note that this center is outside the jth layer, located either below or above the layer for $\alpha_j\gtrless 0$, respectively). The arclength σ along this ray is therefore related to the angle θ along the ray axis via

$$\sigma=\sigma_j+(\theta-\theta_j)c_j/\alpha_j\sin\theta_j \tag{24}$$

(note that θ increases or decreases along the trajectory for $\alpha_j\gtrless 0$). From Eqs. (22) and (24), the axial phase along the beam axis is given by

$$\psi(\sigma)=\psi_j+\int_{\sigma_j}^{\sigma}\frac{d\sigma'}{c_0(\sigma')}=\psi_j+\frac{1}{\alpha_j}\ln\left[\frac{(1-\cos\theta)}{(1-\cos\theta_j)}\frac{\sin\theta_j}{\sin\theta}\right]. \tag{25}$$

Since the wave speed is linear with z we have $\mathbf{C}_2=\mathbf{0}$ in Eq. (20) and hence

$$\mathbf{P}=\mathbf{P}(0)=\mathbf{\Gamma}_j/c_j, \tag{26a}$$

$$\mathbf{Q}=\mathbf{Q}(0)+\mathbf{P}\int_{\sigma_j}^{\sigma}c_0(\sigma)d\sigma'=\mathbf{I}+S(\sigma)\mathbf{\Gamma}_j, \tag{26b}$$

where

$$S(\sigma)=\frac{1}{c_j}\int_{\sigma_j}^{\sigma}c_0(\sigma')d\sigma'=c_j\frac{\cos\theta_j-\cos\theta}{\alpha_j\sin^2\theta_j}. \tag{27}$$

Note that $\partial S/\partial\sigma>0$. The field in Eq. (19) may now be calculated using

$$\mathbf{\Gamma}(\sigma)=c_0(\sigma)\mathbf{P}\mathbf{Q}^{-1}=\frac{c_0(\sigma)}{c_j}(\mathbf{\Gamma}_j^{-1}+S(\sigma)\mathbf{I})^{-1}, \tag{28}$$

$$\det\mathbf{Q}(\sigma)=1+S(\sigma)\ \mathrm{trace}\ \mathbf{\Gamma}_j+S^2(\sigma)\det\mathbf{\Gamma}_j. \tag{29}$$

Expression (28) has the form of the free-space expression $\mathbf{\Gamma}(\sigma)=(\mathbf{\Gamma}^{-1}(0)+\sigma\mathbf{I})^{-1}$ with the ray-integral $S(\sigma)\mathbf{I}$ replacing the free space propagation $\sigma\mathbf{I}$. Note that the GBs that emerge from the point source have $\mathbf{\Gamma}(0)=(i/\beta)\mathbf{I}$ [see discussion after Eq. (20)], hence due to the problem symmetry, $\mathbf{\Gamma}$ remains diagonal throughout.

C. GB reflection and transmission at a planar interface between CG layers

As discussed after Eq. (20), the reflected and transmitted GBs due to a GB that hits an interface between two media, satisfy Snell's refraction law and their amplitudes are determined by the Fresnel reflection and transmission coefficients. It is required to calculate only the initial values of the respective matrices $\mathbf{\Gamma}_{r,t}$ from the matrix $\mathbf{\Gamma}_i$ of the incident GB. For generality we assume in the following that not only α but also c are discontinuous at z_j' (i.e., $c_j^-\neq c_j^+$).

We assume that the incident GB with matrix $\mathbf{\Gamma}_i$ hits the interface from $z<z_j$ at an angle θ_i. Denoting the limiting values of the wave speed and of the wave speed gradient on both sides of the interface z_j as $c_j^{\pm}$ and $\alpha_j^{\pm}$, respectively, the transmitted beam at $z>z_j$ emerges at an angle θ_t that satisfies Snell's law

$$(c_j^-)^{-1}\sin\theta_i=(c_j^+)^{-1}\sin\theta_t \tag{30}$$

and its initial matrix $\mathbf{\Gamma}_t$ is given by (see derivation in the Appendix)

$$\frac{1}{c_j^+}\mathbf{\Theta}_t\mathbf{\Gamma}_t\mathbf{\Theta}_t=\frac{1}{c_j^-}\mathbf{\Theta}_i\mathbf{\Gamma}_i\mathbf{\Theta}_i+\begin{pmatrix}0&0\\0&1\end{pmatrix}\left(\frac{\alpha_j^-}{(c_j^-)^2}\sin^2\theta_i\cos\theta_i-\frac{\alpha_j^+}{(c_j^+)^2}\sin^2\theta_t\cos\theta_t\right), \tag{31}$$

where the matrixes $\mathbf{\Theta}_{i,t}$ are defined by

$$\mathbf{\Theta}_{i,t}=\begin{pmatrix}1&0\\1&\cos\theta_{i,t}\end{pmatrix}. \tag{32}$$

In most cases c is continuous at the interface, and only α is discontinuous. In this case there is no reflection, while the transmission coefficient is 1. The ray angle is unchanged across the interface (i.e., $\theta_t=\theta_i$) so that Eq. (31) reduces to

$$\mathbf{\Gamma}_t=\mathbf{\Gamma}_i+\begin{pmatrix}0&0\\0&1\end{pmatrix}(\alpha_j^--\alpha_j^+)\frac{\sin^2\theta_i}{c_j\cos\theta_i}. \tag{33}$$

Note that the above-noted expressions apply for a general incident matrix $\mathbf{\Gamma}_i$ that is not necessarily diagonal, as in our case.

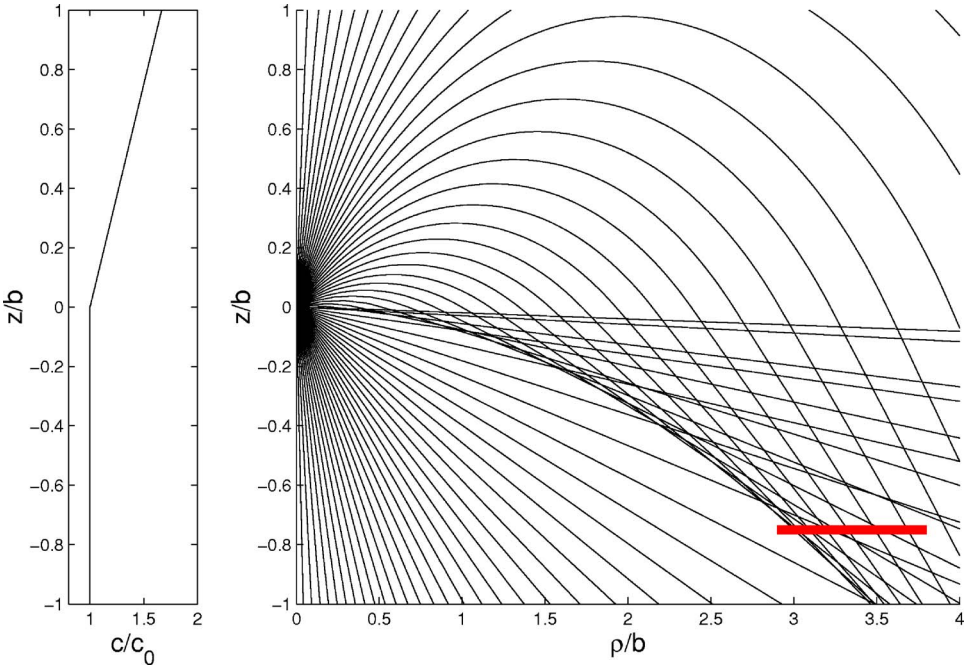


FIG. 2. (Color online) Ray trajectories for the numerical example: The left-hand panel shows the wave speed profile, while the right-hand panel shows the ray trajectories. Note the horizontal line $z/b=-0.75$, $2.9<x/b<3.8$ along which the field is plotted in Fig. 3.

V. NUMERICAL EXAMPLE

The GBS algorithm has been tested in various plane stratified configurations. We present here the relatively simple, yet interesting, example in Fig. 2, consisting of a CG medium above a homogeneous half space in $z<0$. The rays refracting from the CG medium form a caustic in $z<0$ hence the field at points on the illuminated side of the caustic is a sum of the direct ray from the source and two refracted rays. This specific example has been chosen since the three-ray interference provides a sensitive testbed to explore the phase error accumulation of the GBS representation (see Fig. 3). We compared the closed form ray solution, obtained via the technique in Ref. 21, to the GBS result where several GBs are required in order to synthesize the field of each one of the rays. One observes that the GBS algorithm not only provides a uniform solution across the caustic where the ray solution fails, but also recovers remarkably well the fine three-ray interference far from the caustic.

The isotropic point source is located in the origin $\mathbf{r}'=(0,0,0)$ of a three-dimensional space, on the interface between the layers where the wave speed is

$$c(z)=\begin{cases} c_0+\alpha_0 z & \text{for } z>0 \\ c_0 & \text{for } z<0, \end{cases} \tag{34}$$

where $c_0=1500$ [m/s] and $\alpha_0=0.1$ [s^{-1}]. The frequency f has been taken such that $\lambda_0\equiv c_0/f=2\pi$ [m]. Figure 3 compares the GBS results and the ray solution for observation points along the horizontal line $z=-7500$ shown in Fig. 2. Figure 3(b) zooms in on the caustic region at $\rho\simeq 30\,000$.

In the GBS algorithm we chose $b=10\,000$, which is of the same order as the range (see the discussion in Sec. II D). Recalling that $k_0=1$ here, we have $\theta_D=10^{-2}$ rad. The beam discretization mesh was was chosen according Eq. (16) with $\chi_b=\frac{1}{3}$, and we used the truncation condition in Eq. (13) with $\chi_a=3.5$.

As discussed earlier, the GBS reconstructs the smooth transition in Fig. 3(a) from the shadow side of the caustic, where the field is described by the direct ray from the source, to the illuminated side of the caustic where it is described, in addition, by the two refracted rays. The perfect match in Fig. 3 of the fine three-ray interference structure beyond the caustic verifies the phase accuracy of the GBS algorithm to within a fraction of a wavelength. Following the discussion in connection with Fig. 1(c), a phase error accumulation is expected to appear around $\rho\sim 10b=100\,000$.

ACKNOWLEDGMENT

This work was supported in part by the Israel Science Foundation under Grant No. 216/02.

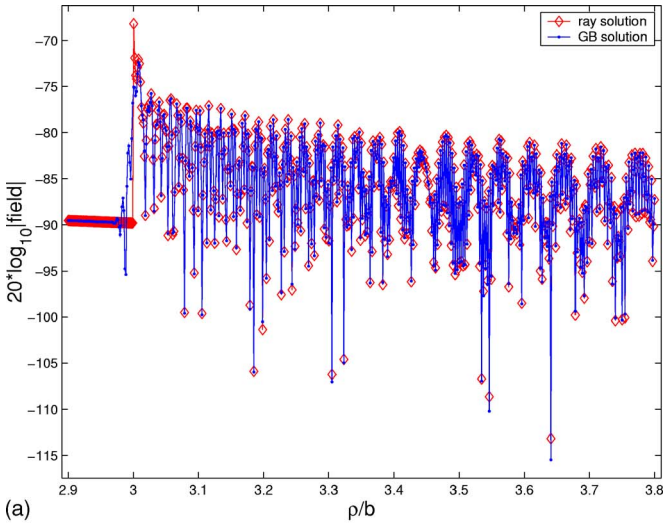
APPENDIX: DERIVATION OF EQ. (31)

Referring to Fig. 4, we consider the interface at $z=0$ between two media described by

$$c(c)=\begin{cases} c_1+\alpha_1 z, & z<0 \\ c_2+\alpha_2 z, & z>0. \end{cases} \tag{A1}$$

An incident GB hits the interface from $z<0$ at an angle θ_i and matrix $\mathbf{\Gamma}_i$. It is described in the beam coordinates (σ,η_1,η_2) where, referring to the beam coordinates discussed in Sec. IV B, $\hat{\boldsymbol{\eta}}_1$ is chosen to be normal to the plane of incidence, and $\hat{\boldsymbol{\eta}}_2=\hat{\boldsymbol{\sigma}}\times\hat{\boldsymbol{\eta}}_1$. Referring to Eq. (23), the incident and transmitted beam trajectories are circular arcs with radii $R_1=c_1/|\alpha_1|\sin\theta_i$ and $R_2=c_2/|\alpha_2|\sin\theta_t$ (Fig. 4 depicts the case $\alpha_{1,2}>0$).

Next we match the paraxial incident and transmitted beam fields at points on the interface near the beam axis. To this end, we express the beam fields in terms of the local interface coordinates $\mathbf{x}=(x_1,x_2)$ that are centered at the point of incidence: we choose x_1 to be normal to the plane of



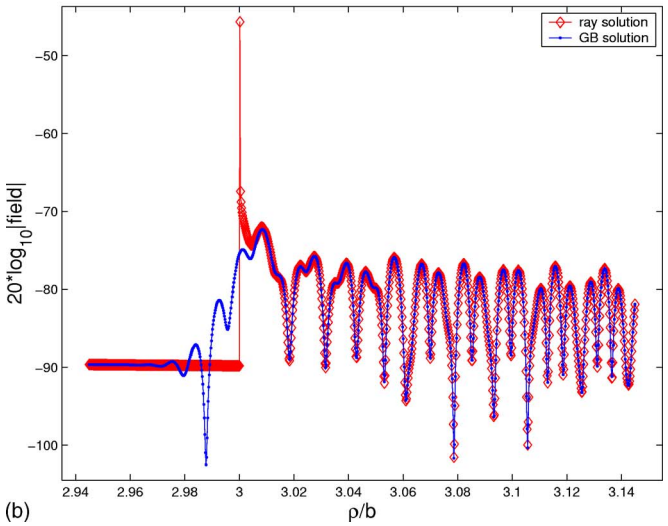


FIG. 3. (Color online) A comparison of the GBS and the ray solutions along the line in Fig. 2 at $z/b=-0.75$. (a) The field along the entire line. (b) Zoom on the caustic region. Note that the GBS has a smooth transition at the caustic, and it models the fine structure of the field afterwards, due to the three-ray interference.

incidence, and $\hat{\mathbf{x}}_2=\hat{\mathbf{z}}\times\hat{\mathbf{x}}_1$. The $\boldsymbol{\eta}_i$ coordinates of the incident beam corresponding to a given off-axis point $\mathbf{x}$ on the interface are found by rotation

$$\boldsymbol{\eta}_i=\boldsymbol{\Theta}_i\mathbf{x}, \tag{A2}$$

where $\boldsymbol{\Theta}_i$ is defined in Eq. (32). In order to determine the σ_i coordinate corresponding to $\mathbf{x}$ we refer to Fig. 4 and consider

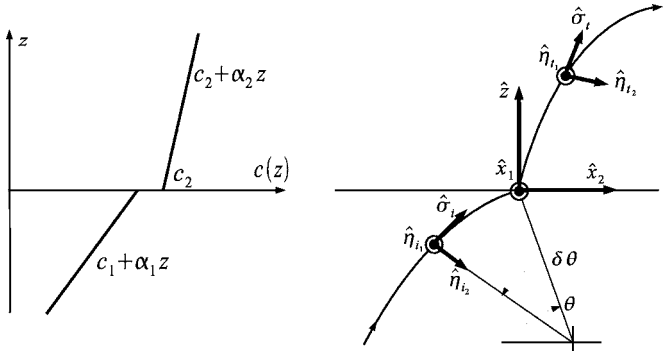


FIG. 4. A GB hitting a planar interface between CG layers.

a point on the incident beam trajectory identified by $\delta\sigma=-R_1\delta\theta<0$. The x_2 coordinate corresponding to this point is

$$x_2=R_1\sin\theta_i[\cot(\theta_i-\delta\theta)-\cot(\theta_i)]\simeq R_1[(\sin\theta_i)^{-1}\delta\theta+\cos\theta_i(\sin\theta_i)^{-2}(\delta\theta)^2]$$

$$=-(\sin\theta_i)^{-1}\delta\sigma+\frac{\alpha_1}{c_1}\cos\theta_i(\sin\theta_i)^{-1}(\delta\sigma)^2.$$

Note that in the derivation we referred to the situation in Fig. 4 where $\alpha_1>0$, but the last expression applies for both $\alpha_1\gtrless 0$. From this expression we find to second order that the beam coordinate σ corresponding to an off-axis point x_2 on the interface is

$$\delta\sigma=-x_2\sin\theta_i+x_2^2\alpha_1c_1^{-1}\cos\theta_i\sin^2\theta_i. \tag{A3}$$

The axial phase $\psi_0(\delta\sigma)$ corresponding to that point is given by [see Eq. (25)]

$$\psi_0(\delta\sigma)=\int_0^{\delta\sigma}\frac{d\sigma}{c_1+\alpha_1\sigma\cos\theta_i}\simeq\frac{\delta\sigma}{c_1}-\frac{\alpha_1(\delta\sigma)^2}{2c_1^2}\cos\theta_i$$

$$=-\frac{x_2}{c_1}\sin\theta_i+\frac{1}{2}x_2^2\frac{\alpha_1}{c_1^2}\cos\theta_i\sin^2\theta_i. \tag{A4}$$

Substituting Eqs. (A2) and (A4) in the exponent of Eq. (19) $\psi_0(\delta\sigma)+(1/2c_1)\boldsymbol{\eta}_i^T\boldsymbol{\Gamma}_i\boldsymbol{\eta}_i$ we obtain the total phase of the incident GB at point $\mathbf{x}$ on the interface

$$\psi_i(\mathbf{x})=-\frac{x_2}{c_1}\sin\theta_i+\frac{1}{2c_1}\mathbf{x}^T\boldsymbol{\Theta}_i^T\left(\boldsymbol{\Gamma}_i+\begin{pmatrix}0&0\\0&1\end{pmatrix}\frac{\alpha_1}{c_1}\frac{\sin^2\theta_i}{\cos\theta_i}\right)\boldsymbol{\Theta}_i\mathbf{x}. \tag{A5}$$

Next we repeat this procedure for the transmitted field and match the complex phases of the incident and transmitted beams for points $\mathbf{x}$ on the interface. The first-order terms in $\mathbf{x}$ yield Snell's refraction law (30). The second-order terms in $\mathbf{x}$ yields

$$\frac{1}{c_i}\boldsymbol{\Theta}_i^T\left(\boldsymbol{\Gamma}_i+\begin{pmatrix}0&0\\0&1\end{pmatrix}\frac{\alpha_1}{c_1}\frac{\sin^2\theta_i}{\cos\theta_i}\right)\boldsymbol{\Theta}_i=\frac{1}{c_2}\boldsymbol{\Theta}_t^T\left(\boldsymbol{\Gamma}_t+\begin{pmatrix}0&0\\0&1\end{pmatrix}\frac{\alpha_2}{c_2}\frac{\sin^2\theta_t}{\cos\theta_t}\right)\boldsymbol{\Theta}_t \tag{A6}$$

leading to the final result in Eq. (31) (one may remove the transpose sign since $\boldsymbol{\Theta}$ are symmetrical in our specific choice of the coordinate system).

[1]E. Heyman and L. B. Felsen, "Gaussian beam and pulsed beam dynamics: Complex source and spectrum formulations within and beyond paraxial asymptotics," J. Opt. Soc. Am. A **18**, 1588–1610 (2001).
[2]E. Heyman, "Pulsed beam solutions for propagation and scattering problems," in *Scattering: Scattering and Inverse Scattering in Pure and Applied Science*, edited by R. Pike and P. Sabatier (Academic, New York, 2002), Vol. **1**, chap. 1.5.4, pp. 295–315.
[3]R. L. Nowack, "Calculation of synthetic seismograms with Gaussian beams," PAGEOPH **160**, 487–507 (2003).
[4]M. M. Popov, "A new method of computation of wave fields using Gaussian beams," Wave Motion **4**, 85–97 (1982).
[5]V. Červený, M. M. Popov, and I. Pšenčhik, "Computation of wave films in inhomogeneous media-Gaussian beam approach," Geophys. J. R. Astron. Soc. **70**, 109–128 (1982).
[6]G. Muller, "Efficient calculation of Gaussian-beam seismograms for two

dimensional inhomogeneous media," Geophys. J. R. Astron. Soc. **79**, 153–166 (1984).
[7]V. Červený, "Gaussian beam synthetic seismogram," J. Geophys. **58**, 44–72 (1985).
[8]V. M. Babič and M. M. Popov, "Gaussian summation method (review)," Radiophys. Quantum Electron. **39**, 1063–1081 (1989).
[9]M. B. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**, 1349–1359 (1987).
[10]L. Klimes, "Gaussian packets in the computation of seismic wavefields," Geophys. J. Int. **99**, 421–433 (1989).
[11]K. Zacek, "Gaussian packet pre-stack depth migration," Society of Exploration Geophysicists 74th Meeting Denver, CO, October 2004, Expanded Abstracts Vol. **23**, pp. 957–960, doi:10.1190/1.1845325.
[12]A. N. Norris, "Complex point-source representation of real sources and the Gaussian beam summation method," J. Opt. Soc. Am. A **3**, 2005–2010 (1986).
[13]E. Heyman, "Complex source pulsed beam expansion of transient radiation," Wave Motion **11**, 337–349 (1989).
[14]M. J. Bastiaans, "The expansion of an optical signal into a discrete set of Gaussian beams," Optik (Stuttgart) **57**, 95–102 (1980).
[15]B. Z. Steinberg, E. Heyman, and L. B. Felsen, "Phase-space beam summation for time-harmonic radiation from large apertures," J. Opt. Soc. Am. A **8**, 41–59 (1991).
[16]A. Shlivinski, E. Heyman, A. Boag, and C. Letrou, "A phase-space beam summation formulation for wideband radiation," IEEE Trans. Antennas Propag. **52**, 2042–2056 (2004).
[17]B. Z. Steinberg and J. McCoy, "Marching acoustic fields in a phase space," J. Acoust. Soc. Am. **93**, 188–204 (1993).
[18]G. Gordon, E. Heyman, and R. Mazar, "A phase-space Gaussian beam summation representation of rough surface scattering," J. Acoust. Soc. Am. **117**, 1911–1921 (2005).
[19]G. Gordon, E. Heyman, and R. Mazar, "Phase space beam summation analysis of rough surface waveguide," J. Acoust. Soc. Am. **117**, 1922–1932 (2005).
[20]V. M. Babič and V. S. Buldyrev, *Short-Wavelength Diffraction Theory: Asymptotic Methods*, translated by E. F. Kuester (Springer, Berlin, 1990), Secs. 9.5-6. [Original Russian edition: *Asymptotic Methods in Short-Wavelength Diffraction Problems: The Model Problem Method* (Nauka, Moscow, 1972)].
[21]V. Červený, *Seismic Ray Theory* (Cambridge University Press, Cambridge, UK, 2001), Section 4.8.3, p. 343.

Periodic orbit theory in acoustics: spectral fluctuations in circular and annular waveguides

M. C. M. Wright[a] and C. J. Ham[b]
Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, United Kingdom

(Received 29 June 2006; revised 23 January 2007; accepted 24 January 2007)

Formulas based on the theory of Weyl are widely used to obtain the average number of modes at or below a given frequency in acoustic and vibrational waveguides. These formulas are valid at asymptotically high frequencies; at finite frequencies they are subject to some error, due to fluctuations in the mode count, which depend on the shape of the waveguide. The periodic orbit theory of semiclassical physics is used to give estimates of the variance of these fluctuations and these results are compared with numerical estimates based on eigenvalues obtained by root-finding. The comparison is good but shows errors that can be related to the nature of the periodic orbit theory. Engineering formulas are provided that give an accurate approximation without significant computational cost. The results are valid for membranes, ducts, and thin plates with clamped and/or simply supported boundary conditions. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2709407]

PACS number(s): 43.20.Ks, 43.20.Bi, 43.20.Dk, 43.20.Mv [ADP] Pages: 1865–1872

I. INTRODUCTION

Circular and annular waveguides are frequently encountered in engineering applications of acoustics, for example in the form of ducts carrying sound from fans that inevitably have axial symmetry (at least at the fan face). In some applications frequencies are high enough that a statistical approach to sound transmission, such as statistical energy analysis (SEA), is appropriated.[1] In this case the modal density is an important parameter, which is usually estimated from Weyl-type formulas for the mode count. In aeroacoustics, too, a high frequency model of cascade noise has been proposed.[2] At such frequencies modal populations are high and approximate methods based on modal averaging can be applied. It has also been shown that the far-field radiation pattern arising from multiple incoherent modes propagating in an open duct can be estimated once modal sound power weighting is known as a function of the cut-on ratios of the modes.[3] Once again the calculation involves a continuous approximation to the mode count distribution function.

The average number of modes present at or below a given frequency or wave number can be estimated from Weyl-type formulas,[4] which are asymptotically accurate at high frequency. At finite frequencies, however, it inevitably has some error with respect to the true mode count; so far this error has not been quantified. The purpose of this paper is to do so for circular and annular waveguides.

II. CIRCULAR WAVEGUIDES

A. Basic properties

Consider a circular domain of radius R containing a field which obeys the Helmholtz equation

$$(\nabla^2 + k^2)\psi = 0, \tag{1}$$

and which may obey the Dirichlet ($\psi=0$) or Neumann condition ($\partial\psi/\partial n=0$) on the boundary.

The spectral density function of this waveguide is defined by

$$\rho(k) = \sum_{i=1}^{\infty} \delta(k - k_i), \tag{2}$$

where the k_i are the ordered solutions of

$$\mathrm{J}_m(kR) = 0 \tag{3}$$

for the Dirichlet condition or

$$\mathrm{J}'_m(kR) = 0 \tag{4}$$

for the Neumann condition. The mode count function is

$$N(k) = \int_0^k \rho(k')\,\mathrm{d}k' = \sum_{i=1}^{\infty} \Theta(k - k_i), \tag{5}$$

where Θ is the Heaviside or "step" function. The mode count can be written as the sum of smooth and oscillating parts

$$N(k) = \bar{N}(k) + N_{\mathrm{osc}}(k), \tag{6}$$

and the average number of modes $\bar{N}$ in a two-dimensional waveguide of area S and perimeter length ∂S at wave number k is given by[4]

$$\bar{N}(k) = \frac{S}{4\pi}k^2 \mp \frac{\partial S}{4\pi}k + \frac{1}{4}, \qquad = \frac{R^2}{4}k^2 \mp \frac{R}{2}k + \frac{1}{4}, \tag{7}$$

for Dirichlet (Neumann) conditions.

Equivalent Weyl formulas are used in many branches of physics and engineering. The nature and magnitude of the

[a]Electronic mail: mcmw@isvr.soton.ac.uk
[b]Electronic mail: ch@isvr.soton.ac.uk

fluctuations about this mean has also been investigated in many fields, one of the first examples being room acoustics.[5–7] In order to quantify the error incurred in using $\bar{N}$ in calculations instead of N it is desirable to be able to estimate the variance (in some normalized sense) of N_{osc}. It can be shown[8] that N_{osc} grows as $\sqrt{k}$ so the quantity to be obtained is $\sigma^2(N_{\mathrm{osc}}(k)/\sqrt{k})$. This quantity is closely related to a quantity known as the Dyson-Mehta statistic Δ,[9] which measures the deviation of the mode count from the best fitting straight line when plotted against k^2 and determines the so-called "spectral rigidity."[8] In a previous article[10] a corresponding result was obtained for rectangular ducts using the periodic orbit theory of semiclassical physics.[11] This theory was developed in the context of quantum physics but is equally applicable to the classical wave problems that concern acousticians.[12–15] This method, and in particular Berry's semiclassical theory of spectral rigidity,[8] will be used to find equivalent formulas for circular and concentric annular waveguides.

B. Periodic orbit theory

The following outline derivation of the semiclassical trace formula is based on that of Balian and Bloch.[13] "Semiclassical" here means "high wave number;" the term arises in quantum physics when considering the motion of a particle at high energies. In a bounded domain with no potential (a so-called "quantum billiard") the governing Schrödinger equation is formally identical to the Helmholtz equation for a waveguide of the same shape. The "classical" limit with infinite wave number is singular, therefore the description must remain "semiclassical" with $k \to \infty$.

A waveguide has a Green's function G satisfying

$$\nabla^2 G + k^2 G = \delta(\mathbf{r} - \mathbf{r}_0) \tag{8}$$

in its domain D and

$$G(\mathbf{r}, \mathbf{r}_0; k) = 0 \tag{9}$$

on its boundary ∂D (Dirichlet conditions are assumed for now). The Green's function describes wave propagation from $\mathbf{r}$ to $\mathbf{r}_0$ and can be related to G_0, the Green's function for free space, by writing the following double layer potential:[16]

$$G(\mathbf{r}, \mathbf{r}_0; k) = G_0(\mathbf{r}, \mathbf{r}_0; k) + \int_{\partial D} \frac{\partial G_0(\mathbf{r}, \alpha)}{\partial n_\alpha} f(\alpha, \mathbf{r}_0)\, \mathrm{d}\alpha, \tag{10}$$

where f is to be determined. Balian and Bloch[13] wrote the solution by successive approximation as

$$\begin{aligned} G(\mathbf{r}, \mathbf{r}_0; k) = {} & G_0(\mathbf{r}, \mathbf{r}_0; k) - 2 \int_{\partial D} \frac{\partial G_0(\mathbf{r}, \alpha)}{\partial n_\alpha} G_0(\alpha, \mathbf{r}_0)\, \mathrm{d}\alpha \\ & + 2^2 \int \int_{\partial D \times \partial D} \frac{\partial G_0(\mathbf{r}, \alpha)}{\partial n_\alpha} \frac{\partial G_0(\alpha, \beta)}{\partial n_\beta} \\ & \times G_0(\beta, \mathbf{r}_0)\, \mathrm{d}\alpha\, \mathrm{d}\beta + \cdots, \end{aligned} \tag{11}$$

where each successive integral corresponds to another reflection of waves from the boundary. For large k the free space Green's function will tend to its large argument asymptote, in two dimensions

$$\begin{aligned} G_0(\mathbf{r}, \mathbf{r}_0; k) &= \frac{1}{4\mathrm{i}} \mathrm{H}_0^{(1)}(k|\mathbf{r} - \mathbf{r}_0|), \\ &\sim \frac{(1+\mathrm{i})}{4\sqrt{\pi k|\mathbf{r} - \mathbf{r}_0|}} \mathrm{e}^{\mathrm{i}k|\mathbf{r}-\mathbf{r}_0|}, \quad \text{as } k \to \infty, \end{aligned} \tag{12}$$

where $H_0^{(1)}$ is a Hankel function of the first kind and of order 0. The integrals in Eq. (11) will all take the form

$$\int \cdots \int g(\mathbf{r}) \exp(\mathrm{i}k|\mathbf{r} - \mathbf{r}_0|)\, \mathrm{d}\alpha \cdots \mathrm{d}\omega,$$

so the method of stationary phase[17] can be applied. Under this approximation each integral will be dominated by the contribution from specularly reflecting paths and the semiclassical approximation to the Green's function will be

$$G(\mathbf{r}, \mathbf{r}_0) \approx \sum_j a_j(\mathbf{r}, \mathbf{r}_0) \exp(\mathrm{i}kL_j + \mathrm{i}\phi_j), \tag{13}$$

where the sum is over all possible ray paths; a_j is a geometrical prefactor, which can be obtained from the geometry of the ray path; L_j is the length of the ray path; and ϕ_j, known as the Maslov phase, is related to phase changes undergone by a ray in traversing the path of length L_j.

The spectral density of the system can be related to the Green's function of the system by

$$\rho(k) = -\frac{1}{\pi} \lim_{\epsilon \to 0} \Im \int_D G(\mathbf{r}, \mathbf{r}; k + \mathrm{i}\epsilon)\, \mathrm{d}\mathbf{r}, \tag{14}$$

where $G(\mathbf{r}, \mathbf{r}; k)$ is known as the trace Green's function, which is singular for $k = k_i$, necessitating the limiting process. The spectral density can be written as the sum of a smooth part and an oscillatory part

$$\rho(k) = \bar{\rho}(k) + \rho_{osc}(k), \tag{15}$$

and substituting the semiclassical Green's function of Eq. (13) into Eq. (14) gives the semi-classical trace formula

$$\rho(k) = \bar{\rho}(k) + \sum_j C_j(k) \cos(kL_j + \phi_j) + \cdots, \tag{16}$$

where the L_j are the lengths of periodic orbits (ray paths that repeat forever), C_j gives the amplitude of their contribution, and ϕ_j is the phase change accumulated by the orbit over one period. By accounting for the different effects of Dirichlet and Neumann boundary conditions on the orbit's phase the effect of each type of boundary condition on the spectrum can be included. In what follows, however, this phase turns out to be irrelevant to the variance of fluctuations that is the main goal here.

The first few periodic orbits for a circular waveguide are as shown in Fig. 1, indexed by v the number of vertices and w the winding number around the center. All possible orbits can be described this way. The oscillating part of the spectrum has been found to be[11,18]

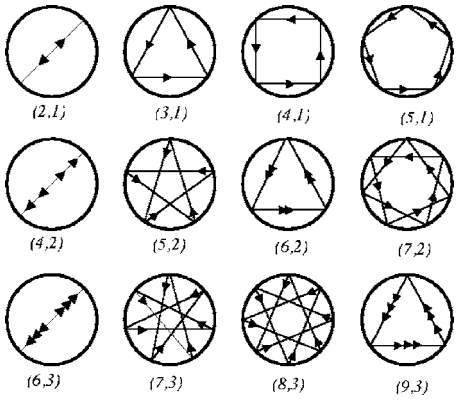


FIG. 1. Periodic orbits in a circular domain and their indices (v,w). Multiple arrow heads indicate multiple traversals (after Balian and Bloch[13]).

$$\rho_{\mathrm{osc}} \approx 2\sqrt{\frac{kR^3}{\pi}}\sum_{w=1}^{\infty}\sum_{v=2w}^{\infty} f_{vw}\sqrt{\frac{\sin^3\varphi_{vw}}{v}} \times \cos\left(kL_{vw} \pm v\frac{\pi}{2} + \frac{\pi}{4}\right) \tag{17}$$

for Dirichlet (Neumann) conditions, where $f_{vw}=1$ when $w=2v$ and 2 when $w>2v$, so as to count orbits without time-reversal invariance twice. The length of the (v,w) orbit is

$$L_{vw} = 2vR\sin\varphi_{vw}, \tag{18}$$

and the angle between the orbit and a tangent to the boundary is

$$\varphi_{vw} = \frac{\pi w}{v}. \tag{19}$$

An equivalent expression for N_{osc} can be found, by integrating ρ_{osc},

$$N_{\mathrm{osc}}(k) \approx \sqrt{\frac{kR}{\pi}}\sum_{w=1}^{\infty}\sum_{v=2w}^{\infty} f_{vw}\sqrt{\frac{\sin\varphi_{vw}}{v^3}} \times \sin\left(kL_{vw} \pm v\frac{\pi}{2} + \frac{\pi}{4}\right). \tag{20}$$

The terms involving Fresnel functions that arise during this integration have been neglected, i.e.,

$$\int_0^k \sqrt{k'}\cos(k'L + C)\,\mathrm{d}k' = \frac{\sqrt{k}}{L}\sin(kL + C) + O(k^0), \tag{21}$$

and terms of $O(k^0)$ have been neglected. Figure 2 shows Eq. (20) plus $\bar{N}$, plotted against the true staircase found by finding zeros of Bessel functions, along with $\bar{N}$, the average mode count. Although the trace formula is in principle only asymptotically valid as $k\to\infty$, nonetheless in this case it provides a good approximation at very low wave numbers. Reasons for this are discussed by Fulling.[19] The agreement is not so good at low wave numbers in the Neumann case, but asymptotic results derived from the formulas are equally valid.

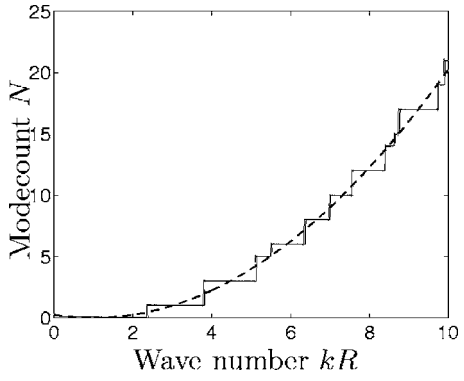


FIG. 2. Mode count staircase for a circular waveguide with Dirichlet conditions calculated from exact eigenvalues and by Eq. (20) (solid lines), and the average mode count (dashed line).

C. Variance formula

The variance of $N_{\mathrm{osc}}/\sqrt{k}$ can now be found by applying the definition of variance directly to the formula obtained above:

$$\sigma^2\left(\frac{N_{\mathrm{osc}}}{\sqrt{k}}\right) = \lim_{\Omega\to\infty}\frac{1}{\Omega}\int_0^{\Omega}\left(\frac{N_{\mathrm{osc}}}{\sqrt{k}}\right)^2\mathrm{d}k. \tag{22}$$

Substituting Eq. (20) into this formula gives an explicit formula for the variance in a circular waveguide. No two distinct (v,w) pairs have the same L_{vw} so that only products of terms with the same values of v and w survive in the limit. The result is then

$$\sigma^2\left(\frac{N_{\mathrm{osc}}}{\sqrt{k}}\right) = \frac{R}{2\pi}\sum_{w=1}^{\infty}\sum_{v=2w}^{\infty} f_{vw}^2\frac{\sin\varphi_{vw}}{v^3} \approx 0.092\,46R. \tag{23}$$

Note that, as predicted, the Maslov phase plays no part in the variance estimate, which is therefore the same for Dirichlet and Neumann conditions. This result agrees closely with estimates from computed eigenvalues.[20] A reasonable engineering approximation would be $\sigma^2 = R/11$.

III. CONCENTRIC ANNULAR WAVEGUIDES

A. Basic properties

The condition to be satisfied for k_i to be an eigenvalue of the Helmholtz operator on a concentric annular domain with Dirichlet boundary conditions is

$$\mathrm{J}_m(\beta k_i R)\mathrm{Y}_m(k_i R) - \mathrm{Y}_m(\beta k_i R)\mathrm{J}_m(k_i R) = 0, \tag{24}$$

or, with Neumann boundary conditions,

$$\mathrm{J}'_m(\beta k_i R)\mathrm{Y}'_m(k_i R) - \mathrm{Y}'_m(\beta k_i R)\mathrm{J}'_m(k_i R) = 0, \tag{25}$$

where J_m are Bessel functions of the first kind and Y_m are Neumann functions of order m. The outer radius of the annulus is R and the inner radius is βR. Chapman[21] has shown that the WKB method can be used to find the eigenvalues of a concentric annulus, however this method still requires numerical root finding.

The average mode count for the annulus is

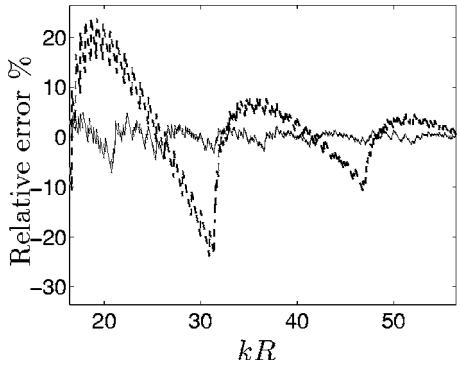


FIG. 3. Relative percentage error between the exact and average mode counts $(N(k)-\bar{N}(k))/N(k)\times 100\%$ for $\beta=0.4$ (solid line) and $\beta=0.8$ (dashed line).

$$\bar{N}(k)=\frac{kR(1+\beta)}{2}\left(\frac{kR(1-\beta)}{2}\mp 1\right)+\frac{1}{4}. \tag{26}$$

The error when $N(k)$ is approximated by $\bar{N}(k)$ is $N_{\mathrm{osc}}(k)$. The relative difference between the exact mode count and the average mode count is therefore $(N(k)-\bar{N}(k))/N(k)$. This quantity is plotted as a percentage in Fig. 3. As expected the relative error falls as k increases, but it does so slowly enough that it may be significant in some applications—kR has to reach 80 before it falls to 2% when $\beta=0.8$.

As before $N_{\mathrm{osc}}/\sqrt{k}$ is stationary and is plotted for two different values of β in Fig. 4. The large amplitude oscillations seen when β is large indicate strong clustering of eigenvalues, corresponding to eigenvalues of modes with the same azimuthal order, but different radial orders being more closely spaced than *vice versa*.

B. Annular trace formula

Richter *et al.*[22] derived the trace formula for the annulus using the Berry-Tabor[14,15] method for integrable systems. An alternative, more geometric, derivation, based on the method of Creagh and Littlejohn,[23,24] is given in the Appendix .

There are three types of periodic orbits in the annulus: type I orbits, which do not touch the central inclusion; type II orbits, which do; and type III orbits, which follow a "bouncing ball" trajectory between the inner and outer boundaries (Fig. 5). The contributions to the formula are

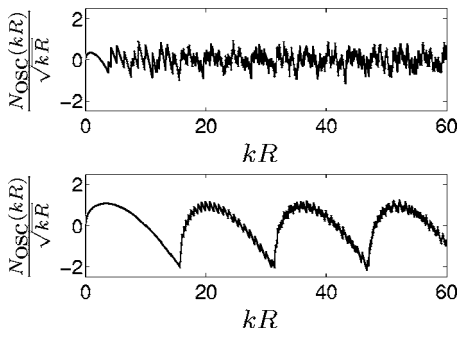


FIG. 4. Oscillating part of the mode count normalized by the square root of the wave number, calculated using the exact eigenvalues for shape ratio $\beta=0.2$ (top) and $\beta=0.8$ (bottom).

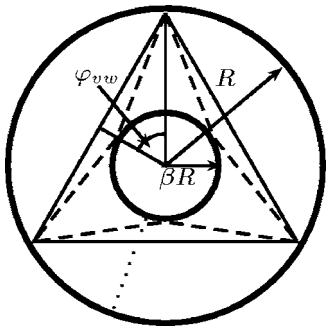


FIG. 5. The three types of periodic orbit within an annular waveguide: type I (solid line), type II (dashed line), and type III (dotted line). The type I and type II orbits have $v=3$ and $w=1$.

$$\rho^{\mathrm{I}}_{\mathrm{osc}}(k)=\sqrt{\frac{2k}{\pi}}\sum_{w=1}^{\infty}\sum_{v=v_0}^{\infty}\frac{(L^{\mathrm{I}}_{vw})^{3/2}}{v^2}\cos\left(kL^{\mathrm{I}}_{vw}\pm v\frac{\pi}{2}+\frac{\pi}{4}\right) \tag{27}$$

for type I orbits and

$$\rho^{\mathrm{II}}_{\mathrm{osc}}(k)=\sqrt{\frac{8k}{\pi}}\sum_{w=1}^{\infty}\sum_{v=v_0}^{\infty}2\frac{R^2A_{vw}}{\sqrt{L^{\mathrm{II}}_{vw}}}\cos\left(kL^{\mathrm{II}}_{vw}-\frac{\pi}{4}\right) \tag{28}$$

for type II orbits, where the plus is for Dirichlet boundary conditions and the minus for Neumann. The type III orbits contribute

$$\rho^{\mathrm{III}}_{\mathrm{osc}}(k)=\sqrt{\frac{8k}{\pi}}\sum_{v=1}^{\infty}\frac{R^2(1-\beta)\sqrt{\beta}}{\sqrt{L^{\mathrm{III}}_{v}}}\cos\left(kL^{\mathrm{III}}_{vw}-\frac{\pi}{4}\right). \tag{29}$$

The type III orbits can also be considered to be type II orbits with $w=0$, but the classification given above is more convenient for the present purpose. The lengths of the orbits are

$$L^{\mathrm{I}}_{vw}=2vR\sin\varphi_{vw}, \tag{30}$$

$$L^{\mathrm{II}}_{vw}=2vR\sqrt{1-2\beta\cos\varphi_{vw}+\beta^2}, \tag{31}$$

$$L^{\mathrm{III}}_{v}=2vR(1-\beta), \tag{32}$$

where the angle $\varphi_{vw}=\pi w/v$ as before. The index v starts from

$$v_0=\lceil \pi w/\cos^{-1}(\beta)\rceil, \tag{33}$$

where $\lceil x\rceil$ is the ceiling of x, that is, the smallest integer greater than or equal to x (this is given incorrectly as $v_0=\lfloor w/\pi/\cos^{-1}\beta\rfloor$ in Richter *et al.*[22]) Finally,

$$A_{vw}=\sqrt{(1-\beta\cos\varphi_{vw})\beta(\cos\varphi_{vw}-\beta)}. \tag{34}$$

The level density is then

$$\rho(k)\approx\bar{\rho}(k)+\rho^{\mathrm{I}}_{\mathrm{osc}}(k)+\rho^{\mathrm{II}}_{\mathrm{osc}}(k)+\rho^{\mathrm{III}}_{\mathrm{osc}}(k). \tag{35}$$

The mode count contribution for the type I orbits is therefore [neglecting $O(k^0)$ terms as before]

$$N_{\mathrm{osc}}^{\mathrm{I}}(k) = 4\sqrt{\frac{kR^3}{\pi}}\sum_{w=1}^{\infty}\sum_{v=v_0}^{\infty}\sqrt{\frac{\sin^3\varphi_{vw}}{v}} \times \frac{\sin\left(kL_{vw}^{\mathrm{I}} \pm v\pi/2 + \pi/4\right)}{L_{vw}^{\mathrm{I}}}, \tag{36}$$

the mode count for the type II orbits is

$$N_{\mathrm{osc}}^{\mathrm{II}}(k) = 4\sqrt{\frac{kR^3}{\pi}}\sum_{w=1}^{\infty}\sum_{v=v_0}^{\infty} \times \frac{A_{vw}}{\sqrt{v}(1-2\beta\cos\varphi_{vw}+\beta^2)^{1/4}} \times \frac{\sin\left(kL_{vw}^{\mathrm{II}} - \pi/4\right)}{L_{vw}^{\mathrm{II}}}, \tag{37}$$

and the mode count for the type III orbits is

$$N_{\mathrm{osc}}^{\mathrm{III}}(k) = \sqrt{\frac{kR}{\pi}}\sqrt{\frac{\beta}{1-\beta}}\sum_{v=1}^{\infty}\frac{\sin\left(kL_v^{\mathrm{III}} - \pi/4\right)}{v^{3/2}}. \tag{38}$$

The total mode count is therefore

$$N(k) \approx \bar{N}(k) + N_{\mathrm{osc}}^{\mathrm{I}}(k) + N_{\mathrm{osc}}^{\mathrm{II}}(k) + N_{\mathrm{osc}}^{\mathrm{III}}(k). \tag{39}$$

C. Variance formula for the annulus

Equation (22) can be used to calculate an estimate for the variance of the difference between true and average mode counts for the annulus. The total variance is then

$$\sigma^2\left(\frac{N_{\mathrm{osc}}^{\mathrm{I}}(k)}{\sqrt{k}}\right) = \sigma_{\mathrm{I}}^2 + \sigma_{\mathrm{II}}^2 + \sigma_{\mathrm{III}}^2 + \sigma_{\mathrm{degen.}}^2, \tag{40}$$

where

$$\sigma_{\mathrm{I}}^2 = \frac{2R}{\pi}\sum_{w=1}^{\infty}\sum_{v=v_0}^{\infty}\frac{\sin\varphi_{vw}}{v^3}, \tag{41}$$

$$\sigma_{\mathrm{II}}^2 = \frac{2R}{\pi}\sum_{w=1}^{\infty}\sum_{v=v_0}^{\infty}\frac{A_{vw}^2}{v^3(1-2\beta\cos\varphi_{vw}+\beta^2)^{3/2}}, \tag{42}$$

$$\sigma_{\mathrm{III}}^2 = \frac{2R}{\pi}\sum_{v=1}^{\infty}\frac{R^2(1-\beta)\beta}{v(L_v^{\mathrm{III}})^2} = \frac{\beta R}{2(1-\beta)}\frac{\zeta(3)}{\pi}, \tag{43}$$

where $\zeta(3)=\Sigma_{n=1}^{\infty}n^{-3}=1.202\cdots$ is Apéry's constant. The contribution from degeneracies between orbits, $\sigma_{\mathrm{degen.}}^2$, is shown below to be small.

This estimate of variance (excluding degeneracies) is plotted with the variance calculated using the exact eigenvalues (Fig. 6) in the range $150 \le k \le 250$. The same results are found using Neumann eigenvalues, as predicted by the semiclassical theory in which the phase contributions vanish when the variance is formed. The dashed line on Fig. 6 shows that the contributions of the type III orbits to the total variance dominates for significant β. In this regime the expression for $N_{\mathrm{osc}}^{\mathrm{III}}$ can be used to estimate N_{osc} to reasonable accuracy, and therefore to predict frequencies at which mode clustering will occur.

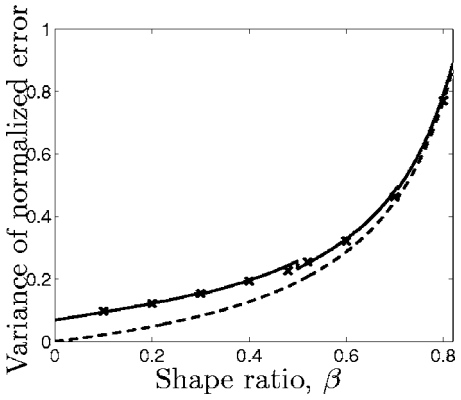


FIG. 6. Variance calculated using the contributions from all three types of orbit but excluding contributions from degeneracies (solid line) and the variance estimated using exact eigenvalues (×). The contribution from type III orbits alone is also shown (dashed line).

D. Degeneracies

Degeneracies occur when two different orbits have the same length. No degeneracies are possible between two orbits of the same type, therefore there are three classes of degeneracy: class A between orbits of type I and type II, class B between type I and type III, and class C between type II and type III. These occur at specific values of β given by

$$\beta_A = \cos\varphi_2 - \sqrt{r^2\sin^2\varphi_1 + \sin^2\varphi_2}, \tag{44}$$

$$\beta_B = 1 - r\sin\varphi_1, \tag{45}$$

$$\beta_C = \frac{r^2\cos\varphi_1 - 1 + \sqrt{(\cos\varphi_1 - 1)r^2(r^2(\cos\varphi_1+1)-2)}}{r^2-1}, \tag{46}$$

where $r=v_1/v_2$ and $\varphi_1=\varphi_{v_1w_1}$ etc. The subscripts 1, 2 refer to orbits of the two different types for each class, in the order stated above. The degeneracy-induced variance is then

$$\sigma_{\mathrm{degen.}}^2 = \sigma_A^2 + \sigma_B^2 + \sigma_C^2, \tag{47}$$

with

$$\sigma_A^2 = \frac{2R}{\pi}\sum_{w_2=1}^{\infty}\sum_{v_2=v_0}^{\infty}\sum_{w_1=1}^{\infty}\sum_{v_1=v_0}^{\infty} \times \frac{A_{v_1w_1}\sqrt{\sin\varphi_1}}{(1-2\beta_A\cos\varphi_2+\beta_A^2)^{3/4}}\frac{\delta(\beta-\beta_A)}{v_1^{3/2}v_2^{3/2}}, \tag{48}$$

$$\sigma_B^2 = \frac{R}{\pi}\sqrt{\frac{\beta_B}{1-\beta_B}}\sum_{v_2=1}^{\infty}\sum_{w_1=1}^{\infty}\sum_{v_1=v_0}^{\infty}\sqrt{\sin\varphi_{v_1w_1}}\frac{\delta(\beta-\beta_B)}{v_1^{3/2}v_2^{3/2}}, \tag{49}$$

$$\sigma_C^2 = \frac{R}{\pi}\sqrt{\frac{\beta_C}{1-\beta_C}}\sum_{v_2=1}^{\infty}\sum_{w_1=1}^{\infty}\sum_{v_1=v_0}^{v_2-1} \times \frac{A_{v_1w_1}}{(1-2\beta_C\cos\varphi_1+\beta_C^2)^{3/4}}\frac{\delta(\beta-\beta_C)}{v_1^{3/2}v_2^{3/2}}, \tag{50}$$

Evaluation of $\sigma_{\mathrm{degen.}}^2$ is complicated by two issues: Firstly v_0,

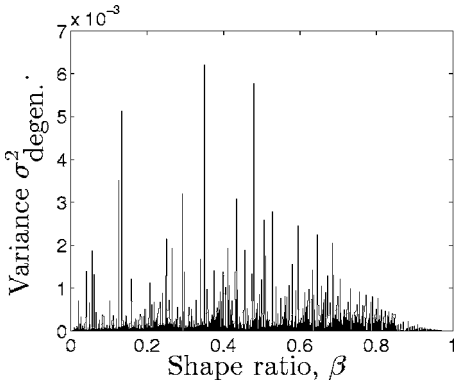


FIG. 7. Contributions to the normalized variance due to degenerate orbit pairs; delta functions are represented by vertical lines.

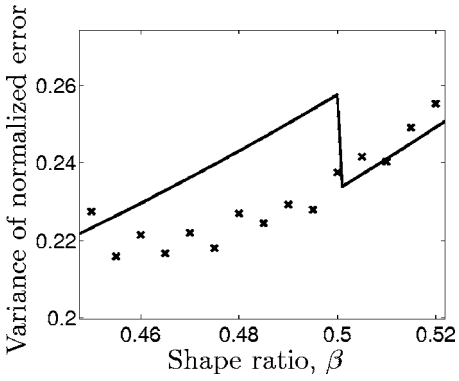


FIG. 8. Expanded section of Fig. 6, showing that the semiclassical formula (solid line) is more accurate after the jump than before, when compared with numerical estimates from eigenvalues (×).

and A_{vw} where it occurs, depend on the value of β_A, β_B, or β_C for the particular degeneracy in question; in practice it is easier to sum from $v_i=2w_i$ and then discard the terms for which $v_i<v_0$. Secondly, even the slightest numerical error could cause contributions that occur at slightly different values of β to be erroneously added to one another. Therefore the degeneracies were calculated using symbolic algebra (Mathematica) and only converted to numerical values after summing. The results are shown in Fig. 7 with delta functions represented by vertical lines whose height corresponds to the magnitude of the delta function. Each v or w index was limited to a maximum value of 20. The apparent relative rarity of degeneracies for values of β close to 1 is due to the fact that for such shapes very high v values are required for an orbit to exist. This plot provides a good indication that the degeneracies can safely be neglected.

IV. DISCUSSION

A. Accuracy

The variances predicted by the semiclassical formulas differ from those obtained from numerical estimates of ensembles of eigenvalues in a number of respects.

Consider the behavior as $\beta\to 0$. It is tempting to think that when the center is reduced to a point the behavior will be that of a circle with all the modes that do not have a node at the center supressed. In fact, as pointed out by Rayleigh[25] and further discussed by Gottlieb,[26] these modes survive and, as can be seen by comparing Eqs. (7) and (26), the average mode count for the annulus with $\beta\to 0$ is the same as that for the circle. The semiclassical prediction however is, from Eq. (41),

$$\frac{2R}{\pi}\sum_{w=1}^{\infty}\sum_{v=2w+1}^{\infty}\frac{\sin\varphi_{vw}}{v^3}=0.068\,55R \tag{51}$$

for infinitesimal β, clearly an underestimate. The precise behavior for very small β is difficult to observe because accurate numerical solution of Eq. (24) or Eq. (25) for high eigenvalues becomes unreliable in this regime.

Steps are observed at $\beta=0.174,0.223,0.309,\frac{1}{2},\sqrt{2}/2,0.809,\ldots$, i.e., $\beta=\cos\varphi_{vw}$ for small v and w, the steps being most pronounced where L^{I}_{vw} is shortest. At these values a family of type I orbits is annihilated, for example when $\beta>\frac{1}{2}$ the family of triangular orbits with $v=3w$ is destroyed. Figure 8 shows that these steps are a semiclassical artefact and are not observed in the actual eigenvalues.

It might be thought the error in this region is due to neglecting diffracted periodic orbits. Robinett[27] has shown that the presence of such diffracted orbits can be detected in the Fourier transform of the spectral density (the "length spectrum"). But neglecting these diffractive orbits would be expected to have the greatest effect just after a family of orbits had been destroyed, where the diffracted orbits have to creep the shortest distance around the center in order to be periodic. As Fig. 8 shows, the error is actually greater *before this point* than after it. Just before a step the type I and type II orbits with the same (v,w) are very similar and it is possible that interference between the two orbits causes some error in the steepest descent approximation upon which the trace formula is founded. However, the class A degeneracy formula Eq. (48) shows that this contribution vanishes as the two orbits approach each other. A possible explanation is that as the grazing incidence of the type II orbits against the inner boundary becomes shallower, the steepest-descent approximation, upon which the semiclassical trace formula depends, becomes less accurate.

A simple, but accurate, estimate of the variance, which actually gives better agreement with the numerical estimates than the full semiclassical formula, is found by interpolating Eqs. (23) and (43) to give

$$\sigma^2\left(\frac{N_{\mathrm{osc}}}{\sqrt{k}}\right)\approx\left[0.092\,46(1-\beta)+0.1913\frac{\beta}{1-\beta}\right]R, \tag{52}$$

$$\approx\left(\frac{\beta^2+\beta/5+1}{1-\beta}\right)\frac{R}{11}. \tag{53}$$

B. Extension to thin plates

Bogomolny and Hugues[28] have shown that the same periodic orbit theory applies to thin plates that obey the biharmonic equation

$$(\nabla^4 - k^4)\psi = 0, \tag{54}$$

with simply supported or clamped boundary conditions. They give expressions for the phase change for a ray reflecting from a straight boundary with either of these boundary conditions. As has been shown above this phase plays no part in the variance and so it can be concluded that the expressions given above will apply equally to such plates, including annular plates with different conditions on their inner and outer boundaries. Bogomolny and Hugues also give a corresponding expression for a free edge, but in this case additional edge waves may be generated so the same conclusion cannot be drawn.

V. CONCLUSIONS

Periodic orbit theory has been used to estimate the variance of fluctuations in the mode count of circular and annular waveguides. The results broadly agree with numerical estimates but also show errors that can be related to the semiclassical nature of the periodic orbit theory and its approximation. Engineering formulas have been given that estimate the variance to reasonable accuracy but do not require summation of a series.

This theory is only valid for the concentric annulus; when the annulus is eccentric ray paths are chaotic.[20,29] This leads to qualitatively different behavior, which will be explored in future work.

ACKNOWLEDGMENTS

The authors thank C. L. Morfey for suggesting the problem and S. C. Creagh for helpful discussions. C. J. Ham was supported by a Rayleigh Scholarship from the ISVR and M. C. M. Wright was supported by an EPSRC Advanced Research Fellowship.

APPENDIX: DERIVATION OF ANNULUS TRACE FORMULA BY THE CREAGH-LITTLEJOHN METHOD

Creagh and Littlejohn[23] give a general formula for the amplitude and phase terms in Eq. (16) for a wide class of Hamiltonian systems. In the case of a two-dimensional waveguide with axial symmetry it can be reduced to

$$\rho_{\mathrm{osc}}(k) \approx \sqrt{\frac{2kR}{\pi}} \sum_{\gamma} \frac{L_\gamma \sqrt{\cos\psi}}{\alpha_\gamma \left|(\partial\Theta/\partial\psi)_\gamma\right|^{1/2}} \times \cos\left(kL_\gamma - \mu_\gamma \pi/2 - \pi/4\right), \tag{A1}$$

where L_γ is the length of the periodic orbit and α_γ is the discrete rotational symmetry of the periodic orbit. The angle Θ measures the amount by which the periodic orbit fails to close when ψ, the initial angle with respect to a radial line, has been perturbed.

For the type I orbits $\cos\psi = \sin\varphi_{vw}$. Figure 9(a) shows the effect on the type I orbit of a small perturbation. Basic geometry gives

$$\left|\frac{\partial\Theta}{\partial\psi}\right|_{\mathrm{I}} = 2v. \tag{A2}$$

The angle ψ for the type II orbits is

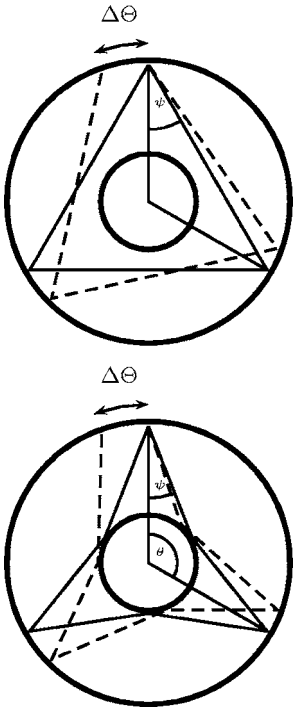


FIG. 9. Unperturbed periodic orbit (solid line) and perturbed orbit (dashed line) for type I and type II orbits with $v=3$ and $w=1$.

$$\cos\psi = \frac{(1 - \beta\cos\varphi_{vw})}{\sqrt{1 - 2\beta\cos\varphi_{vw} + \beta^2}}. \tag{A3}$$

Figure 9(b) shows the effect on the type II orbit of a small perturbation. Here

$$\left|\frac{\partial\Theta}{\partial\psi}\right|_{\mathrm{II}} = \frac{2v(1 - 2\beta\cos\varphi_{vw} + \beta^2)}{\beta(\cos\varphi_{vw} - \beta)}. \tag{A4}$$

The corresponding formula for type III orbits can be obtained by setting $w=0$ to give

$$\left|\frac{\partial\Theta}{\partial\psi}\right|_{\mathrm{III}} = \frac{2v(1-\beta)}{\beta}. \tag{A5}$$

The Maslov index μ_γ counts the number of times the orbit γ grazes a caustic; $\mu_{\mathrm{c}} = +1$ for each caustic touched. It also counts the number of reflections from a Dirichlet boundary condition; $\mu_{\mathrm{b}} = +2$ for each. There is no contribution from a reflection from a Neumann boundary condition. If the sign of $\partial\Theta/\partial\psi$ is negative, $\delta=0$; if positive, $\delta=1$.[30,31] The total Maslov index is

$$\mu_\gamma = \mu_{\mathrm{c}} + \mu_{\mathrm{b}} - \delta. \tag{A6}$$

See Table I for particular values. Maslov indices for annuli with different boundary conditions on their inner and outer boundaries can be constructed in a similar manner.

TABLE I. Maslov index for different orbits and boundary conditions.

Orbit	Boundary condition	μ_b	μ_c	δ	μ_γ
Type I	Dirichlet	$2v$	v	1	$3v-1$
Type I	Neumann	0	v	1	$v-1$
Type II	Dirichlet	$4v$	0	0	$4v$
Type II	Neumann	0	0	0	0

[1]R. H. Lyon, *Statistical Energy Analysis of Dynamical Systems* (MIT, Cambridge, 1975).
[2]E. Envia, "A high frequency model of cascade noise," *4th AIAA/CEAS Aeroacoustics Conference*, 98-2318 (1998).
[3]E. J. Rice, "Multimodal far-field acoustic radiation pattern using mode cutoff ratio," AIAA J. **16**, 906–911 (1978).
[4]H. P. Baltes and E. R. Hilf, *Spectra of Finite Systems* (Bibliographisches Insititut Wissenschaftsverlag, Mannheim, 1976).
[5]R. H. Bolt, "Normal frequency spacing statistics," J. Acoust. Soc. Am. **19**, 79–90 (1947).
[6]G. M. Roe, "Frequency distribution of normal modes," J. Acoust. Soc. Am. **13**, 1–7 (1941).
[7]R. H. Bolt and R. W. Roop, "Frequency response fluctuations in rooms," J. Acoust. Soc. Am. **22**, 280–289 (1950).
[8]M. V. Berry, "Semiclassical theory of spectral rigidity," Proc. R. Soc. London, Ser. A **400**, 229–251 (1985).
[9]F. J. Dyson and M. L. Mehta, "Statistical theory of energy levels of complex systems IV," J. Math. Phys. **4**, 701–712 (1963).
[10]M. C. M. Wright, "Variance of deviations from the average mode count for rectangular wave guides," ARLO **2**, 19–24 (2001).
[11]M. Brack and R. K. Bhaduri, *Semiclassical Physics* (Addison–Wesley, Reading, MA, 1997).
[12]M. C. Gutzwiller, "Periodic orbits and classical quantization conditions," J. Math. Phys. **12**, 343–358 (1971).
[13]R. Balian and C. Bloch, "Distribution of eigenfrequencies for the wave equation in a finite domain: III. Eigenfrequency density oscillations," Ann. Phys. **69**, 76–160 (1972).
[14]M. V. Berry and M. Tabor, "Closed orbits and the regular bound spectrum," Proc. R. Soc. London, Ser. A **349**, 101–123 (1976).
[15]M. V. Berry and M. Tabor, "Calculating the bound spectrum by path summation in action–angle variables," J. Phys. A **10**, 371–379 (1977).
[16]P. Filippi, D. Habault, J.-P. Lefebvre, and A. Bergassoli, *Acoustics: Basic Physics, Theory and Methods* (Academic, San Diego, 1989).
[17]R. H. Self, "Asymptotic expansion of integrals," in *Lecture Notes on the Mathematics of Acoustics*, edited by M. C. M. Wright (Imperial College, London, 2005), Chap. 4, pp. 91–105.
[18]S. M. Reimann, M. Brack, A. G. Magner, and M. V. N. Murthy, "Applications of classical periodic orbit theory to circular billiards with small scattering centers," Surf. Rev. Lett. **3**, 19–23 (1996).
[19]S. A. Fulling, "Spectral oscillations, periodic orbits, and scaling," J. Phys. A **35**, 4049–4066 (2002).
[20]M. C. M. Wright, C. L. Morfey, and S. H. Yoon, "On the modal distribution for circular and annular ducts," in *9th AIAA/CEAS Aeroacoustics Conference*, 2003-3141 (2003).
[21]S. J. Chapman, "On the approximation of the eigenvalues of an annulus using complex rays," Eur. J. Appl. Math. **10**, 225–236 (1999).
[22]K. Richter, D. Ullmo, and R. A. Jalabert, "Orbital magnetism in the ballistic regime: geometrical effects," Phys. Rep. **276**, 1–83 (1996).
[23]S. C. Creagh and R. G. Littlejohn, "Semiclassical trace formulas in the presence of continuous symmetries," Phys. Rev. A **44**, 836–850 (1991).
[24]S. C. Creagh and R. G. Littlejohn, "Semiclassical trace formulas for systems with non-Abelian symmetry," J. Phys. A **25**, 1643–1669 (1992).
[25]Lord Rayleigh, *Theory of Sound*, 2nd ed. (Dover, New York, 1945), Vol. **1**.
[26]H. P. W. Gottlieb, "On pinned and collared membranes," J. Sound Vib. **225**, 1000–1004 (1999).
[27]R. W. Robinett, "Periodic orbit theory analysis of the circular disk or annular billiard: Nonclassical effect and the distribution of energy eigenvalues," Am. J. Phys. **67**, 67–77 (1999).
[28]E. Bogomolny and E. Hugues, "Semiclassical theory of flexural vibrations of plates," Phys. Rev. E **57**, 5404–5424 (1998).
[29]G. Gouesbet, S. Meunier-Guttin-Cluzel, and G. Grehan, "Periodic orbits in Hamiltonian chaos of the annular billiard," Phys. Rev. E **65**, 016212 (2001).
[30]E. B. Bogomolny, "Smoothed wave-functions of chaotic quantum-systems," Physica D **31**, 169–189 (1988).
[31]S. C. Creagh, "Trace formula for broken symmetry," Ann. Phys. **248**, 60–94 (1996).

Nonlinear surface waves in soft, weakly compressible elastic media

Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton[a)]
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713–8029

(Received 25 May 2006; revised 16 January 2007; accepted 22 January 2007)

Nonlinear surface waves in soft, weakly compressible elastic media are investigated theoretically, with a focus on propagation in tissue-like media. The model is obtained as a limiting case of the theory developed by Zabolotskaya [J. Acoust. Soc. Am. **91**, 2569–2575 (1992)] for nonlinear surface waves in arbitrary isotropic elastic media, and it is consistent with the results obtained by Fu and Devenish [Q. J. Mech. Appl. Math. **49**, 65–80 (1996)] for incompressible isotropic elastic media. In particular, the quadratic nonlinearity is found to be independent of the third-order elastic constants of the medium, and it is inversely proportional to the shear modulus. The Gol'dberg number characterizing the degree of waveform distortion due to quadratic nonlinearity is proportional to the square root of the shear modulus and inversely proportional to the shear viscosity. Simulations are presented for propagation in tissue-like media. [DOI: 10.1121/1.2697098]

PACS number(s): 43.25.Dc [RR] Pages: 1873–1878

I. INTRODUCTION

Experiments have been reported on the linear[1,2] propagation of surface waves in gels, and on both the linear[3] and nonlinear[4] propagation of surface waves in soft tissue and tissue phantoms. Experiments conducted by Klochkov[4] on nonlinear surface waves in human forearm tissue reveal the generation of harmonics, combination frequencies, and subharmonics. The aim of the present contribution is to model and interpret nonlinear surface wave propagation in the class of soft, weakly compressible media represented by gels and soft tissue.

Gels and soft tissue are characterized by a shear modulus μ that is at least five orders of magnitude smaller than the bulk modulus K.[5] Velocities of the shear (transverse) and compressional (longitudinal) waves in such media thus differ by about three orders of magnitude. When modeling the propagation of shear waves in weakly compressible media, defined here by $\mu/K \ll 1$, it is reasonable to assume that the medium is incompressible because of the negligible coupling between the shear and compressional waves. The assumption of incompressibility is appropriate for nonlinear as well as linear shear waves.[6]

Rayleigh's surface wave solution applies to arbitrary homogeneous isotropic media, but he also evaluated and discussed his solution in the limiting case of an incompressible elastic medium.[7] In the linear approximation, Rayleigh waves consist of contributions that satisfy either the compressional wave equation or the shear wave equation. Since the nonlinearity in compressional waves is quadratic and in shear waves it is cubic, and since compressional waves are prohibited in incompressible media, it is not obvious *a priori* whether the nonlinearity in Rayleigh waves is quadratic or cubic at leading order in weakly compressible media.

Fu and Devenish[8] have derived an evolution equation for nonlinear Rayleigh waves in prestressed incompressible elastic media. Their model indicates that the Rayleigh wave nonlinearity is quadratic. It also reveals that the quadratic nonlinearity does not depend on any third-order elastic constants of the medium (e.g., A, B, and C in the notation of Landau and Lifshitz[9]). The quadratic nonlinearity is therefore purely geometric and due exclusively to the nonlinearity in the strain tensor.

We take a different approach. Instead of beginning with the assumption of an incompressible elastic medium and proceeding from there to derive an evolution equation, we begin with an evolution equation for Rayleigh waves in arbitrary isotropic media[10] and evaluate its coefficients for $\mu/K \ll 1$. While any one of several models of nonlinear Rayleigh waves might be used (e.g., see those reviewed by Mayer[11]), we are most familiar with the one developed in Ref. 10, and which has been shown[12] to be equivalent to the model developed previously by Parker.[13] The coefficients are evaluated using relations developed previously[14] between second- and third-order elastic constants for $\mu/K \ll 1$. Following this procedure, we recover the result that the nonlinearity is quadratic and independent of third-order elastic constants, in agreement with Fu and Devenish.[8]

In contrast with the model developed by Fu and Devenish,[8] ours accounts for energy loss. Gels and soft tissue can produce strong viscous attenuation of surface waves, and it is essential to take this into account when making estimates of waveform distortion and harmonic generation. It is the ratio of nonlinear effects to viscous effects, not nonlinearity alone, that determines the extent to which nonlinearity affects the propagation of a wave. We define a Gol'dberg number that accounts for this ratio and can be used to predict whether shock formation is possible. Numerical simulations with viscosity taken into account are compared with reported measurements of both nonlinear surface waves[4] and nonlinear shear waves[15] in soft tissue-like media.

[a)]Electronic mail: hamilton@mail.utexas.edu

While the motivation for this work is recent experiments in tissue-like media, it should be noted that real tissues are complex media that can exhibit anisoptropy, prestress, and other properties,[16] and they may therefore differ considerably from the homogeneous isotropic media considered here. The present contribution is thus intended just as a first step toward modeling nonlinear surface waves in tissue and tissue-like media.

II. LINEAR SURFACE WAVES

A brief discussion of linear theory is necessary to establish notation and provide context for the nonlinear theory in Sec. III. We begin by noting two possible approaches to obtaining the equations of motion for surface wave propagation in an incompressible elastic medium. One is to first derive the equations for a surface wave in a compressible elastic medium, and then take their limits for an incompressible elastic medium. This is the approach followed by Rayleigh.[7] The linear stress-strain relation for a compressible isotropic elastic medium is[9]

$$\sigma_{ik} = K u_{ll} \delta_{ik} + 2\mu\left(u_{ik} - \tfrac{1}{3} u_{ll} \delta_{ik}\right), \tag{1}$$

where σ_{ik} is the stress tensor,

$$u_{ik} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right) \tag{2}$$

is the linear approximation of the strain tensor, repeated subscripts indicate summation, δ_{ik} is the Kronecker delta, K the bulk modulus, and μ the shear modulus. This constitutive relation leads to the classical results for longitudinal waves that propagate at speed $c_l = [(3K+4\mu)/3\rho]^{1/2}$, transverse waves that propagate at speed $c_t = (\mu/\rho)^{1/2}$, and Rayleigh waves that propagate at a speed c_R that is a function of the ratio c_t/c_l. The incompressible limit for each of these three modes can be obtained by letting $K=\infty$.

Alternatively, one may start with the stress-strain relation for an incompressible elastic medium.[17–19] In this limit, while the bulk modulus becomes infinite, the dilitation u_{ll} vanishes, but the product of the two quantities is finite and equals the negative of the hydrostatic pressure, $K u_{ll} \to -p$, such that Eq. (1) becomes

$$\sigma_{ik} = -p\,\delta_{ik} + 2\mu u_{ik}. \tag{3}$$

In this approach pressure appears as a new field variable to be eliminated.

Our approach is to take the limit of the solutions for compressible elastic media. Letting $K=\infty$, one obtains for the longitudinal ($\mathbf{u}_l$) and transverse ($\mathbf{u}_t$) components of the particle displacement vectors in an incompressible medium

$$\nabla^2 \mathbf{u}_l = \mathbf{0}, \quad \nabla^2 \mathbf{u}_t = \frac{1}{c_t^2}\frac{\partial^2 \mathbf{u}_t}{\partial t^2}. \tag{4}$$

The longitudinal component satisfies Laplace's equation, which by itself does not admit solutions for wave propagation. The longitudinal component participates in surface wave propagation through coupling to the transverse component at the free surface. The total displacement vector is $\mathbf{u} = \mathbf{u}_l + \mathbf{u}_t$, which must satisfy a stress-free boundary condition.

The medium is assumed to occupy the half-space $z \leqslant 0$, and propagation is in the x direction. An harmonic solution for the x (horizontal) and z (vertical) components of a Rayleigh wave in a compressible elastic medium may be expressed in the form[9]

$$u_x = i(\xi_t e^{\xi_t kz} + \eta e^{\xi_l kz}) u_0 e^{i(kx-\omega t)}, \tag{5}$$

$$u_z = (e^{\xi_t kz} + \eta \xi_l e^{\xi_l kz}) u_0 e^{i(kx-\omega t)}, \tag{6}$$

where u_0 is an arbitrary displacement amplitude, ω the angular frequency, and $k=\omega/c_R$ the wave number. The quantities $e^{\xi_t kz}$ and $e^{\xi_l kz}$ come from the solutions for $\mathbf{u}_t$ and $\mathbf{u}_l$, respectively, in which the coefficients ξ_t and ξ_l ensure that the respective dispersion relations are satisfied independently. The coefficient η appears when the solutions for $\mathbf{u}_t$ and $\mathbf{u}_l$ are combined and required to satisfy the stress-free boundary condition.

The Rayleigh wave speed is expressed as $c_R = \xi c_t$. For $K=\infty$, ξ is a root of the following equation:[7]

$$\xi^6 - 8\xi^4 + 24\xi^2 - 16 = 0. \tag{7}$$

The desired root is

$$\xi = 0.9553\ldots, \tag{8}$$

such that the result $c_R/c_t = 0.9553$ is the maximum possible value of this quantity for a compressible elastic medium. Accordingly, the values of the remaining coefficients in Eqs. (5) and (6) are

$$\xi_l = 1, \tag{9}$$

$$\xi_t = \sqrt{1-\xi^2} = 0.2956, \tag{10}$$

$$\eta = -\frac{2\sqrt{1-\xi^2}}{2-\xi^2} = -0.5437. \tag{11}$$

While the coefficient $\xi_l = 1$ in Eqs. (5) and (6) may appear to be superfluous in the present context, it is a function of c_t/c_l for compressible elastic media, and it is retained here to distinguish between contributions due to $\mathbf{u}_t$ and $\mathbf{u}_l$. For example, the penetration depth of the transverse component is seen to be $\xi_l/\xi_t \simeq 3.4$ times that of the longitudinal component.

The effect of viscosity is now considered. In the linear approximation, the stress tensor for an incompressible viscoelastic medium is

$$\sigma_{ik} = -p\,\delta_{ik} + 2\mu u_{ik} + 2\eta_v \frac{\partial u_{ik}}{\partial t}, \tag{12}$$

where η_v is the coefficient of shear viscosity. Equation (12) follows from the dissipative stress tensor for a compressible elastic medium.[9] For a time-harmonic disturbance, Eq. (12) can be expressed as

$$\sigma_{ik} = -p\,\delta_{ik} + 2\tilde{\mu} u_{ik}, \tag{13}$$

where $\tilde{\mu} = \mu - i\omega\eta_v$ is a complex shear modulus accounting for viscous loss. Provided $\omega\eta_v/\mu \ll 1$, the attenuation coefficient may be obtained as the small imaginary part of $\tilde{k} = \omega/\tilde{c}_R$, where $\tilde{c}_R = \xi(\tilde{\mu}/\rho)^{1/2}$, such that

$$\tilde{k} = \frac{\omega}{c_R}\left(1 - \frac{i\omega\eta_v}{\mu}\right)^{-1/2} \simeq \frac{\omega}{c_R} + i\frac{\eta_v\omega^2}{2\mu c_R}. \tag{14}$$

With $\tilde{k}$ replacing k in the exponentials $e^{i(kx-\omega t)}$ in Eqs. (5) and (6), the viscous attenuation coefficient for the Rayleigh wave is thus

$$\alpha_v = \frac{\eta_v\omega^2}{2\mu c_R} = \frac{\eta_v\rho^{1/2}\omega^2}{2\xi\mu^{3/2}}. \tag{15}$$

For $\omega\eta_v/\mu \ll 1$, the effect of replacing k by $\tilde{k}$ in the exponentials $e^{\xi_t kz}$ and $e^{\xi_l kz}$ in Eqs. (5) and (6) is negligible and therefore ignored.

The assumption that the only effect of shear viscosity is to introduce simple exponential attenuation in the solution for a Rayleigh wave in a lossless medium relies on the ratio $\omega\eta_v/\mu$ being sufficiently small. A detailed analysis of Rayleigh waves in viscoelastic media by Currie *et al.*[20] reveals that for incompressible media the assumption is valid only for $\omega\eta_v/\mu < 0.159$. For $\omega\eta_v/\mu > 0.159$ two Rayleigh waves are possible, one with the usual retrograde particle motion at the surface, and the other with prograde motion. For all numerical simulations in Sec. IV we have $\omega\eta_v/\mu < 0.1$, and therefore the assumption is valid for our purposes.

Equations (8)–(11) and (14) are taken as approximate values of the coefficients in Eqs. (5) and (6) for any weakly compressible medium characterized by $\mu/K \ll 1$, and in particular for soft gels and tissue-like media in which K exceeds μ by five orders of magnitude.

III. NONLINEAR SURFACE WAVES

We now consider the form taken for $\mu/K \ll 1$ by an evolution equation derived previously for nonlinear Rayleigh waves in arbitrary isotropic media.[10] The particle velocity components $v_{x,z} = \partial u_{x,z}/\partial t$ are expressed as modal expansions of harmonics whose depth dependence is described by linear theory, Eqs. (5) and (6):

$$v_x(x,z,\tau) = \frac{i}{2}\sum_{n=1}^{\infty} v_n(x)(\xi_t e^{n\xi_t kz} + \eta e^{nkz})e^{-in\omega\tau} + \text{c.c.}, \tag{16}$$

$$v_z(x,z,\tau) = \frac{1}{2}\sum_{n=1}^{\infty} v_n(x)(e^{n\xi_t kz} + \eta e^{nkz})e^{-in\omega\tau} + \text{c.c.}, \tag{17}$$

where $\tau = t - x/c_R$ is a retarded time, ω here is the fundamental angular frequency of the periodic waveform, often the source frequency, and c.c. designates the complex conjugate of all preceding terms.

Nonlinear evolution of the spectral amplitudes $v_n(x)$ is determined by the following set of coupled equations:[21]

$$\frac{dv_n}{dx} + n^2\alpha_v v_n = -\frac{n^2\omega\rho}{2\mu\xi^4\zeta}\sum_{l+m=n}\frac{lm}{|lm|}R_{lm}v_l v_m, \tag{18}$$

where α_v is the attenuation coefficient at the fundamental frequency ω, and

$$\zeta = \xi_t + \xi_t^{-1} + 4\eta + 2\eta^2 = 2.095. \tag{19}$$

The nonlinearity matrix is given by

$$\begin{aligned} R_{lm} = {} & \frac{\alpha'}{|l| + |m|\xi_t + |l+m|\xi_t} + \frac{\alpha'}{|l|\xi_t + |m| + |l+m|\xi_t} \\ & + \frac{\alpha'}{|l|\xi_t + |m|\xi_t + |l+m|} + \frac{\beta'}{|l|\xi_t + |m| + |l+m|} \\ & + \frac{\beta'}{|l| + |m|\xi_t + |l+m|} + \frac{\beta'}{|l| + |m| + |l+m|\xi_t} \\ & + \frac{3\gamma'}{|l| + |m| + |l+m|}, \end{aligned} \tag{20}$$

where the coefficients in the numerators are combinations of the second- and third-order elastic constants for the medium. Relations (9)–(11) were taken into account in Eqs. (19) and (20). Definitions of α', β', and γ' for an arbitrary isotropic medium, in both Landau-Lifshitz and Murnaghan notation, are provided elsewhere.[12] We use the notation of Landau and Lifshitz,[9] in which the second-order constants are μ and K, and the third-order constants are A, B, and C.

The next step is to determine the limiting forms of the coefficients α', β', and γ' for $\mu/K \ll 1$. For soft tissue-like media, the values of μ and A are between five and six orders of magnitude smaller than those of K, B, and C.[5] From the relations between the third-order elastic constants and the properties of liquids one concludes that $B = -K + O(\mu)$.[14,22] All that can be concluded for the other third-order elastic constants is that $A = O(\mu)$ and $C = O(K)$, which is sufficient for taking the required limit.

The limiting procedure is illustrated by carrying out the calculation for just one of the three coefficients in Eq. (20). The definition of β' for an arbitrary isotropic medium is[12]

$$\beta' = -\frac{\eta^2\xi_t}{\mu}\left(\frac{7}{3}\mu + K + A + 2B\right)(1 - \xi_l^4), \tag{21}$$

where $\xi_l^2 = 1 - \xi^2 c_t^2/c_l^2$. Substitution of the relations for c_t and c_l yields $\xi_l^2 = 1 - \xi^2\mu/K + O(\mu^2/K^2)$. The asymptotic form of Eq. (21) is thus

$$\beta' = -\eta^2\xi_t\frac{K}{\mu}\left[1 + O\left(\frac{\mu}{K}\right)\right]\left[2\xi^2\frac{\mu}{K} + O\left(\frac{\mu^2}{K^2}\right)\right] \tag{22}$$

$$= -2\eta^2\xi^2\xi_t + O(\mu/K). \tag{23}$$

Proceeding in this manner with the remaining coefficients, one obtains the following asymptotic values after ignoring terms of $O(\mu/K)$:

$$\alpha' = -2\eta\xi^2\xi_t^2 = 0.08671, \tag{24}$$

$$\beta' = -2\eta^2\xi^2\xi_t = -0.1595, \tag{25}$$

$$\gamma' = -\tfrac{4}{3}\eta\xi^2(\eta^2 - \xi_t) \simeq 0. \tag{26}$$

For an incompressible elastic medium, Eq. (26) is found to be identically zero by making use of Eq. (7). The term containing γ' in Eq. (20) is associated with three-wave interactions (i.e., sum or difference frequency generation) involving only longitudinal components of the particle motion. The other terms are associated with the interactions of both longitudinal and transverse components. Whereas there are

three-wave interactions involving only transverse components for compressible media, they do not occur in incompressible media.

Note also that the values of α', β', and γ' do not depend on any elastic constants, and therefore the nonlinearity matrix R_{lm} is universal for media with $\mu/K \ll 1$. Equation (18) thus indicates that the rate of nonlinear distortion is inversely proportional to μ, and moreover that this distortion rate depends on no other elastic properties of the medium. In particular, since no third-order elastic constants are involved for $\mu/K \ll 1$, the quadratic nonlinearity of Rayleigh waves is a purely geometric effect due to the nonlinearity of the strain tensor, as discovered by Fu and Devenish[8] for incompressible elastic media.

IV. NUMERICAL CALCULATIONS

Numerical integration of Eq. (18) is facilitated by introducing the following dimensionless quantities:

$$V_n = v_n/v_0, \quad X = x/x_0, \quad A_n = n^2 \alpha_v x_0, \tag{27}$$

where v_0 is a characteristic velocity amplitude,

$$x_0 = \frac{\mu \xi^4 \zeta}{4|R_{11}| v_0 \omega \rho} \tag{28}$$

is a characteristic nonlinear distortion length, and

$$R_{11} = -0.03296. \tag{29}$$

Substitution in Eq. (18) yields

$$\frac{dV_n}{dX} + A_n V_n = -\frac{n^2}{8|R_{11}|}\left(\sum_{m=1}^{n-1} R_{m,n-m} V_m V_{n-m} - 2 \sum_{m=n+1}^{\infty} R_{m,n-m} V_m V^*_{m-n} \right), \tag{30}$$

where the asterisk designates complex conjugate. For a finite number of harmonics retained in Eqs. (16) and (17), that number N replaces ∞ in the second summation in Eq. (30). The first summation corresponds to sum frequency generation, the second to difference frequency generation.

A monofrequency source condition is expressed as

$$v_n(0) = \begin{cases} v_0, & n = 1 \quad (31) \\ 0, & n > 1, \quad (32) \end{cases}$$

substitution of which in Eqs. (16) and (17) yields

$$v_x(0,0,t) = (\xi_t + \eta) v_0 \sin \omega t = -0.2481 v_0 \sin \omega t, \tag{33}$$

$$v_z(0,0,t) = (1 + \eta) v_0 \cos \omega t = 0.4563 v_0 \cos \omega t. \tag{34}$$

For this source condition, Eq. (28) approximates the shock formation distance in the absence of viscosity.[23] The instantaneous source velocity $(v_x^2 + v_z^2)^{1/2}$ varies between $|\xi_t + \eta| v_0$ and $|1 + \eta| v_0$, the nominal value of which is its rms value. This rms value is what we shall identify as the source amplitude v_s, such that

$$v_0 = \frac{\sqrt{2} v_s}{\sqrt{(\xi_t + \eta)^2 + (1 + \eta)^2}} = 2.723 v_s, \tag{35}$$

which permits conversion of source velocities reported for experiments to the reference velocity v_0 employed in the theory. This conversion is convenient for the comparisons below with measurements of nonlinear shear waves. In experiments with nonlinear surface waves, ordinarily the vertical velocity component v_{z0} is reported, in which case one should use the relation $v_0 = v_{z0}/(1+\eta) = 2.191 v_{z0}$.

Inspection of Eq. (30) reveals that the degree to which an initially sinusoidal waveform distorts during propagation depends entirely on the numerical value of $A_1 = \alpha_v x_0$. Indeed, the definition of the Gol'dberg number used to characterize the nonlinearity of acoustic waves in fluids is[24] $\Gamma = 1/\alpha_v x_0$. Substantial acoustic waveform distortion and shock formation occur only for $\Gamma \gg 1$. For $\Gamma \lesssim 1$, second-harmonic distortion is less than 10%, and the waveform remains nearly sinusoidal. Evaluation of the Gol'dberg number for Rayleigh wave propagation in weakly compressible media, using Eqs. (15) and (28), yields

$$\Gamma = \frac{8|R_{11}|}{\xi^3 \zeta} \frac{v_0 \sqrt{\mu \rho}}{\eta_v \omega} = 0.06256 \frac{v_s \sqrt{\mu \rho}}{\eta_v f}, \tag{36}$$

where $f = \omega/2\pi$ is the source frequency.

Equation (36) also displays the entire set of parameters required to perform numerical simulations for real materials. Nominal values of material parameters for soft gel are provided in, or can be inferred from, two recent papers reporting experiments on tissue phantoms. The first, by Chen *et al.*,[25] reports the average values $\mu = 4.1$ kPa and $\eta_v = 0.17$ Pa s, and the corresponding values in the second, by Catheline *et al.*,[15] are $\mu = 2.8$ kPa and $\eta_v = 0.4$ Pa s. Simulations are presented below using both sets of material parameters, but using the same density $\rho = 1100$ kg/m^3 as reported by Catheline *et al.*

For the source conditions we use the values in the experiments performed on nonlinear shear waves by Catheline *et al.*,[15] where the velocity amplitude was $v_s = 0.6$ m/s and the frequency was $f = 100$ Hz. Their measurements of sawtooth-like waveforms containing shocks demonstrate strong nonlinear distortion of a shear wave under these conditions. However, Eq. (36) yields $\Gamma = 1.6$ for their experimental conditions. Thus in this particular case, weak nonlinear distortion is predicted for a Rayleigh wave under the same conditions for which strong nonlinear distortion occurs for a shear wave. This may be counterintuitive, because whereas Rayleigh wave nonlinearity is quadratic, shear wave nonlinearity is cubic. For the values of μ and η_v reported by Chen *et al.*[25] but for all other parameters (ρ, f, v_s) the same, the Gol'dberg number is $\Gamma = 4.7$.

Figure 1 shows the horizontal (v_x) and vertical (v_z) components of the velocity waveforms calculated numerically at $x = x_0$ for the conditions described above. The solid lines are the source waveforms, the dashed lines correspond to $\mu = 4.1$ kPa and $\eta_v = 0.17$ Pa s ($x_0 = 4.8$ cm, $\Gamma = 4.7$), and the dot-dash lines correspond to $\mu = 2.8$ kPa and $\eta_v = 0.4$ Pa s ($x_0 = 3.3$ cm, $\Gamma = 1.6$).

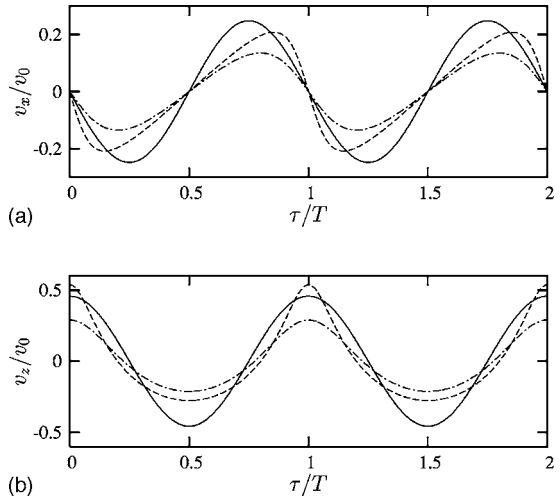


FIG. 1. Numerical simulations of waveforms for the horizontal (a) and vertical (b) components of the particle velocities at $(x,z)=(x_0,0)$ for $v_s=0.6$ m/s, $f=100$ Hz, $\rho=1100$ kg/m^3, and two different pairs of shear moduli and viscosity coefficients: $\mu=4.1$ kPa and $\eta_v=0.17$ Pa s (dashed lines, $\Gamma=4.7$); $\mu=2.8$ kPa and $\eta_v=0.4$ Pa·s (dot-dash lines, $\Gamma=1.6$). The solid lines are the source waveforms at $x=0$, and $T=2\pi/\omega$ is the period.

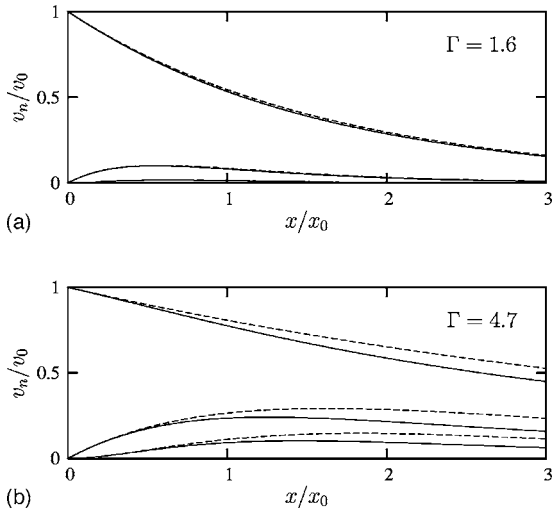


FIG. 2. Harmonic propagation curves for (a) $\Gamma=1.6$ and (b) $\Gamma=4.7$. From top to bottom in each panel, $n=1$, 2, and 3. Solid lines are numerical solutions of Eq. (30), dashed lines are Eqs. (39)–(41).

As predicted on the basis of the Gol'dberg number, the dot-dash waveforms in Fig. 1 for the parameters reported by Catheline *et al.*[15] exhibit considerably weaker nonlinearity than the shear waveforms containing shocks that they measured in their experiment. The difference in waveform distortion may be interpreted by considering the Gol'dberg number for a shear wave. The shock formation distance for a shear wave is $x_t=c_t^3/\beta\omega v_s^2$, where β is the coefficient of nonlinearity defined in Ref. 6, and the viscous attenuation coefficient is $\alpha_t=\eta_v\omega^2/2\mu c_t$. The corresponding Gol'dberg number $\Gamma_t=1/\alpha_t x_t$ is thus

$$\Gamma_t=\left(1+\frac{\frac{1}{2}A+D}{\mu}\right)\frac{3v_s^2\rho}{\eta_v\omega}, \tag{37}$$

where A and D are third- and fourth-order elastic constants, respectively, defining the coefficient β. Forming the ratio with Eq. (36) yields

$$\frac{\Gamma_t}{\Gamma}=7.6\left(1+\frac{\frac{1}{2}A+D}{\mu}\right)\frac{v_s}{c_t}. \tag{38}$$

For soft tissue-like media, μ, A, and D are of the same order.[6,14] By comparing measurements with numerical simulations, Catheline *et al.*[15] estimate that $\mu+\frac{1}{2}A+D=5.1$ kPa (this is the value of their generic elastic constant γ), and therefore $(\frac{1}{2}A+D)/\mu=0.82$, such that for their elastic medium $\Gamma_t/\Gamma=14v_s/c_t$. For their experiment $v_s/c_t=0.6/1.6=0.375$, and thus $\Gamma_t=5.3\Gamma=8.4$, which is in nominal agreement with the value 8.6 estimated by Catheline *et al.* The Gol'dberg numbers become equal only if the source amplitude is reduced to $v_s/c_t=1/14$, for which $\Gamma=\Gamma_t=1.6$, and nonlinear effects in both the surface and shear waves are relatively weak (recall the dot-dash waveforms in Fig. 1, for which $\Gamma=1.6$).

Figure 2 shows the corresponding harmonic propagation curves for the two simulations, with $\Gamma=1.6$ in Fig. 2(a), and $\Gamma=4.7$ in Fig. 2(b). The solid lines are the simulations obtained by direct numerical integration of Eq. (30). For sufficiently weak nonlinearity, one may use the following perturbation solutions of Eq. (30) for the first several harmonics:

$$v_1/v_0=e^{-\alpha_v x}, \tag{39}$$

$$v_2/v_0=0.25\Gamma(e^{-2\alpha_v x}-e^{-4\alpha_v x}), \tag{40}$$

$$v_3/v_0=0.030\Gamma^2(2e^{-3\alpha_v x}-3e^{-5\alpha_v x}+e^{-9\alpha_v x}). \tag{41}$$

These solutions appear as the dashed lines in Fig. 2. Whereas the agreement is very close for $\Gamma=1.6$, the solutions become invalid with increasing distance for $\Gamma=4.7$. The overestimation by Eq. (39), which corresponds to linear theory, indicates failure to account for nonlinear losses at the source frequency due to harmonic generation. Likewise, Eqs. (40) and (41) only account for energy gained from nonlinear interaction, and not for energy lost.

Klochkov[4] reports vibroacoustic experiments on nonlinear surface waves in human forearm tissue. While insufficient information is provided to determine the attenuation coefficient α_v required for numerical simulation, qualitative comparison can be made with the relative levels of the harmonics that were measured. The source amplitude reported for Klochkov's Fig. 3 corresponds to $v_s=0.6$ m/s, the same value that was used in the simulations for our Figs. 1 and 2, and the reported source frequency is $f=124$ Hz. The relative levels of the harmonics in the frequency spectrum presented in Klochkov's Fig. 3 are approximately -14 dB for v_2/v_1 and -28 dB for v_3/v_1 (note that Klochkov's frequency spectrum pertains to acceleration, not velocity). These relative levels are indicative of quadratic nonlinearity, and a Gol'dberg number midway between the two values considered in Fig. 2.

While the present theory indicates that Rayleigh wave nonlinearity in soft tissue-like media is quadratic at leading order, it is possible that cubic nonlinearity could become important and even dominate the quadratic nonlinearity at sufficiently high amplitude. It is straightforward to include cubic nonlinearity in the present theoretical model by ex-

tending the same Hamiltonian formalism used to account for quadratic nonlinearity.[10] But even if straightforward, the calculations are tedious, and they may be unnecessary in the absence of experimental evidence suggesting that cubic nonlinearity can be important in soft elastic media.

V. SUMMARY

Surface waves in elastic media for which μ/K is of order 1 differ fundamentally from surface waves in weakly compressible elastic media defined by $\mu/K \ll 1$. The latter case can be described by the theory for incompressible elastic media. While both the longitudinal and transverse components of surface waves in compressible media obey wave equations, only the transverse component obeys a wave equation in incompressible media; the longitudinal component obeys Laplace's equation. The surface wave is thus purely solenoidal, just like a shear wave. But unlike a shear wave, which possesses cubic nonlinearity, the nonlinearity of the surface wave is quadratic at leading order.

The quadratic nonlinearity of surface waves in weakly compressible elastic media is independent of third-order elastic constants. It is therefore independent of material nonlinearity, and it is due exclusively to geometric nonlinearity. This nonlinearity is inversely proportional to the shear modulus. The extent of harmonic generation and waveform distortion increases with the Gol'dberg number, which is a measure of the ratio of nonlinearity to dissipation. The Gol'dberg number is proportional to the square root of the shear modulus and inversely proportional to the shear viscosity.

ACKNOWLEDGMENTS

We are grateful to one of the reviewers for bringing several key references to our attention. This work was supported by National Institutes of Health Grant No. EB004336, and the Internal Research and Development Program at Applied Research Laboratories.

[1]H. Kikuchi, K. Sakai, and K. Takagi, "Complex propagation of surface waves on soft gels," Phys. Rev. B **49**, 3061–3064 (1994).
[2]P. K. Choi, E. Jyounoutil, K. Yuuki, and Y. Onodera, "Experimental observation of pseudocapillary and Rayleigh modes on soft gels," J. Acoust. Soc. Am. **106**, 1591–1593 (1999).
[3]T. J. Royston, H. A. Mansy, and R. H. Sandler, "Excitation and propagation of surface waves on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **106**, 3678–3686 (1999).
[4]B. N. Klochkov, "Nonlinear vibroacoustic processes at the surface of a biological tissue," Acoust. Phys. **46**, 621–623 (2000).
[5]S. Catheline, J.-L. Gennisson, and M. Fink, "Measurement of elastic nonlinearity of soft solid with transient elastography," J. Acoust. Soc. Am. **114**, 3087–3091 (2003).
[6]E. A. Zabolotskaya, M. F. Hamilton, Yu. A. Ilinskii, and G. D. Meegan, "Modeling of nonlinear shear waves in soft elastic media," J. Acoust. Soc. Am. **116**, 2807–2813 (2004).
[7]Lord Rayleigh, "On waves propagated along the plane surface of an elastic solid," Proc. London Math. Soc. **17**, 4–11 (1885).
[8]Y. Fu and B. Devenish, "Effects of pre-stresses on the propagation of nonlinear surface waves in an incompressible elastic half-space," Q. J. Mech. Appl. Math. **49**, 65–80 (1996).
[9]L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 3rd ed. (Pergamon, New York, 1986).
[10]E. A. Zabolotskaya, "Nonlinear propagation of plane and circular Rayleigh waves in isotropic solids," J. Acoust. Soc. Am. **91**, 2569–2575 (1992).
[11]A. P. Mayer, "Surface acoustic waves in nonlinear elastic media," Phys. Rep. **256**, 237–366 (1995).
[12]E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "General theory for the spectral evolution of nonlinear Rayleigh waves," J. Acoust. Soc. Am. **102**, 1402–1417 (1997), Appendix A.
[13]D. F. Parker, "Waveform evolution for nonlinear surface acoustic waves," Int. J. Eng. Sci. **26**, 59–75 (1988).
[14]M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Separation of compressibility and shear deformation in the elastic energy density (L)," J. Acoust. Soc. Am. **116**, 41–44 (2004).
[15]S. Catheline, J.-L. Gennisson, M. Tanter, and M. Fink, "Observation of shock transverse waves in elastic media," Phys. Rev. Lett. **91**, 164301-1–164301-4 (2003).
[16]Y. C. Fung, *Biomechanics: Mechanical Properties of Living Tissues*, 2nd ed. (Springer, New York, 1993).
[17]S. Nair and D. A. Sotiropoulos, "Elastic waves in orthotropic incompressible materials and reflection from an interface," J. Acoust. Soc. Am. **102**, 102–109 (1997).
[18]M. Destrade, "Surface waves in orthotropic incompressible materials," J. Acoust. Soc. Am. **110**, 837–840 (2001).
[19]R. W. Ogden and P. Chi Vinh, "On Rayleigh waves in incompressible orthotropic elastic solids," J. Acoust. Soc. Am. **115**, 530–533 (2004).
[20]P. K. Currie, M. A. Hayes, and P. M. O'Leary, "Viscoelastic Rayleigh waves," Q. Appl. Math. **35**, 35–53 (1977).
[21]M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Evolution equations for nonlinear Rayleigh waves," J. Acoust. Soc. Am. **97**, 891–897 (1995).
[22]S. Kostek, B. K. Sinha, and A. N. Norris, "Third-order elastic constants for an inviscid fluid," J. Acoust. Soc. Am. **94**, 3014–3017 (1993).
[23]E. Yu. Knight, M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "On Rayleigh wave nonlinearity, and analytical approximation of the shock formation distance," J. Acoust. Soc. Am. **102**, 2529–2535 (1997), Eq. (28).
[24]D. T. Blackstock, M. F. Hamilton, and A. D. Pierce, "Progressive waves in lossless and lossy fluids," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 4, Sec. 5.4.
[25]S. Chen, M. Fatemi, and J. F. Greenleaf, "Remote measurement of material properties from radiation force induced vibration of an embedded sphere," J. Acoust. Soc. Am. **112**, 884–889 (2002).

Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data

Yong-Min Jiang[a] and N. Ross Chapman[b]
School of Earth and Ocean Sciences, University of Victoria, PO Box 3055 Victoria, British Columbia V8W 3P6, Canada

Mohsen Badiey
College of Marine Studies, University of Delaware, Newark, Delaware 19716

(Received 20 October 2006; revised 16 January 2007; accepted 17 January 2007)

This paper describes geoacoustic inversion of low frequency air gun data acquired during an experiment on the New Jersey shelf. Hybrid optimization and Bayesian inversion techniques based on matched field processing were applied to multiple shots from three air gun data sets recorded by a vertical line array in a long-range shallow water geometry. For the Bayesian inversions, full data error covariance matrix was estimated from a set of consecutive shots that had high temporal coherence and small spatial variation in source position. The effect of different data error information on the geoacoustic parameter uncertainty estimates was investigated by using the full data error covariance matrix, a diagonalized version of the full error covariance, and a diagonal matrix with identical variances. The comparison demonstrated that inversion using the full data error information provided the most reliable parameter uncertainty estimates. The inversions were highly sensitive to the near sea floor geoacoustic parameters, including sediment attenuation, of a simple single-layer geoacoustic model. The estimated parameter values of the model were consistent with depth averaged values (over wavelength scales) of a high resolution geoacoustic model developed from extensive ground truth information. The interpretation of the frequency dependence of the estimated attenuation is also discussed.
[DOI: 10.1121/1.2642137]

PACS number(s): 43.30.Pc, 43.60.Pt [AIT] Pages: 1879–1894

I. INTRODUCTION

Sound propagation in shallow water and littoral environments is strongly influenced by the sea floor and subbottom properties. Information about the parameters of geoacoustic models of the ocean bottom and their spatial variability is essential for a wide range of applications that involve making predictions of the intensity of the sound field. Consequently, there has been considerable research effort to develop inversion methods for estimating geoacoustic models and the uncertainties of the model parameter estimates from measurements of the acoustic field in the water.

Matched field inversion (MFI) is an effective geoacoustic inversion technique that has been benchmarked with simulated data,[1–11] and applied to extract simple but physically meaningful geoacoustic models for several different experimental environments.[12–18] Point estimation by optimization algorithms and appraisal analysis based on Bayesian theory are the two major approaches. MFI is a highly nonlinear inverse problem that has no analytical solution. Bayesian inversion is considered to provide a complete solution to the geoacoustic inverse problem, since it generates comprehensive information about the model parameters and their uncertainties in the form of *maximum a posteriori* (MAP) estimates, one-dimensional and two-dimensional joint marginal probability distributions, and interparameter correlations.[19–23]

In the Bayesian MFI approach, the likelihood function most commonly chosen to describe the misfit between the data and the model is based on the assumption that the data errors (the difference between the measured and modeled acoustic fields) for the hydrophones across an array are Gaussian distributed and spatially uncorrelated, and therefore the data error covariance matrix is considered to be diagonal with identical variances on the diagonal elements. Since this assumption on the data errors is not generally true, researchers have started to investigate the influence of data error statistics on the model uncertainty estimates and develop methods to incorporate the relevant data error information into matched-field inversion.[22,24–26] Huang *et al.*[26] present their uncertainty analysis by incorporating full data error covariance matrices in a simulation case. Dosso *et al.*[24] demonstrate an approach for the case in which the amount of measured data is limited. They assume that the residuals (i.e., the data error, the difference between the measured and modeled data from the optimized model parameters) for the hydrophones on the array at different frequencies are ergodic random processes in the spatial domain. Consequently the ensemble average was replaced by a finite spatial average, which leads to a Toeplitz form of the data covariance matrix.

In this paper, we present results of geoacoustic inversion of broadband experimental data that demonstrate the impact of highly spatially correlated data errors on the inversion.

[a]Electronic mail: minj@uvic.ca
[b]Electronic mail: chapman@uvic.ca

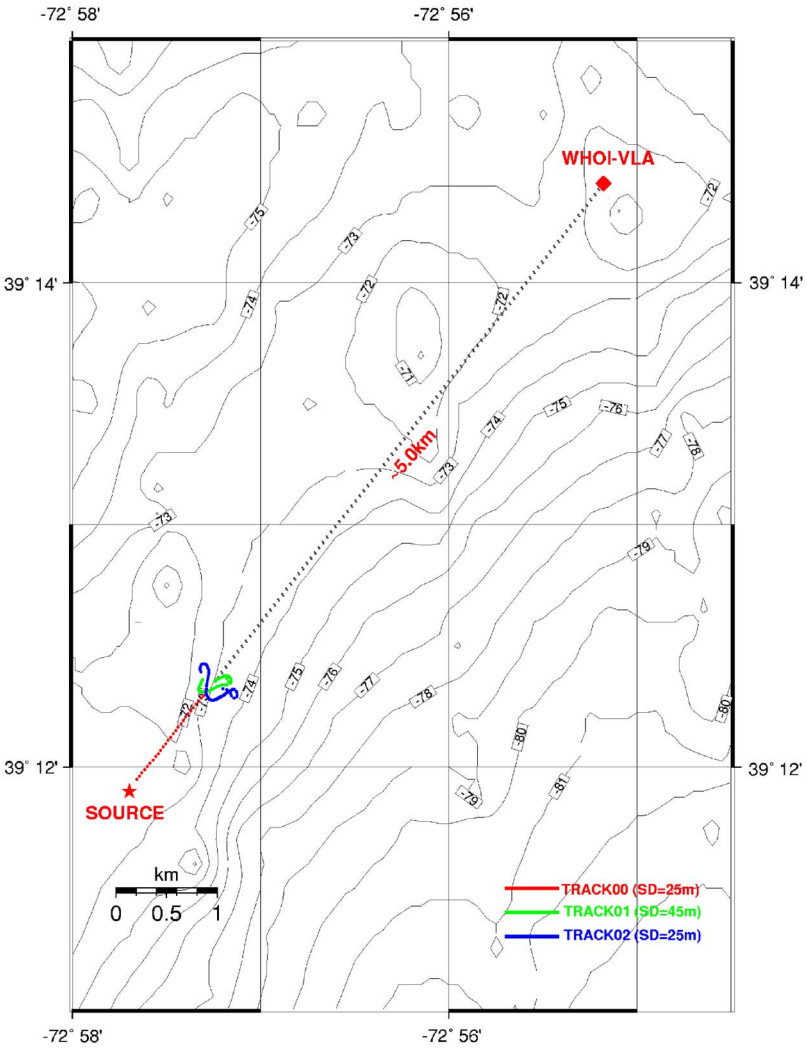


FIG. 1. (Color online) Bathymetry of experimental site. Star indicates the starting point of the source ship; the straight dotted line near the source is the track for closing course to the VLA; the other two dotted lines are the source ship maneuvering routes of Track01 and Track02 that are displayed in Fig. 2.

The data were obtained in long-range experiments (~70 water depths) using small air guns as sound sources in the SWARM'95 (Shallow Water Acoustics in Random Media) experiment carried out on the continental shelf off the New Jersey coast.[27] We introduce a method for estimating the full data error covariance matrix by averaging over the residuals from an ensemble of multiple air gun shot transmissions. The method takes account of the temporal and spatial coherence of the spectral components of the signal to construct a multifrequency Bartlett misfit function. The Bayesian inversion results obtained using the full covariance matrix, a diagonalized full covariance matrix (by neglecting the off-diagonal elements of the matrix), and a diagonal matrix with identical variances that are derived from effective-variance estimates[24] are compared in order to investigate the effect of different data error information on geoacoustic parameter uncertainty estimates. The statistical tests used to validate the assumptions of stationarity, randomness, and Gaussianality of data error residuals in our approach are shown as well.

The motivation for the inversions presented here was to estimate a realistic geoacoustic model for the shallow water site. Of particular interest in the analysis was the estimation of sediment attenuation at low frequencies. The sensitivity to attenuation in MFI is strongly dependent on the design of the experiment. Since the effect of attenuation on the acoustic field accumulates with range, longer range experimental geometries generally provide greater sensitivity. The air gun source used in the study reported here provided high quality, long-range data in two low frequency bubble pulse bands centered at 65 and 150 Hz. We show that the inversions of the data from this long-range experiment were sensitive to sediment attenuation, and present a geoacoustic model for the upper portion of the sediment column that is effective for depths within a few wavelengths of the sea floor. The estimated geoacoustic model parameters are compared with high resolution ground truth information from various other surveys that have been done in the vicinity. These include *in situ* sediment probes, grab samples and shallow cores, and high resolution chirp sonar surveys.

The sediment attenuation estimated from the inversion of the data from this experiment is interpreted as an effective attenuation that includes the contributions of different mechanisms of energy loss in long-range acoustic propagation, such as intrinsic absorption in the sea floor material, sound scattering due to spatial inhomogeneities of the seabed and the roughness of the sea floor, multiple interbed reflections, etc.[28] Although our results show a frequency-dependent attenuation that is nonlinear, we do not advocate that they support the predictions of any model of sound

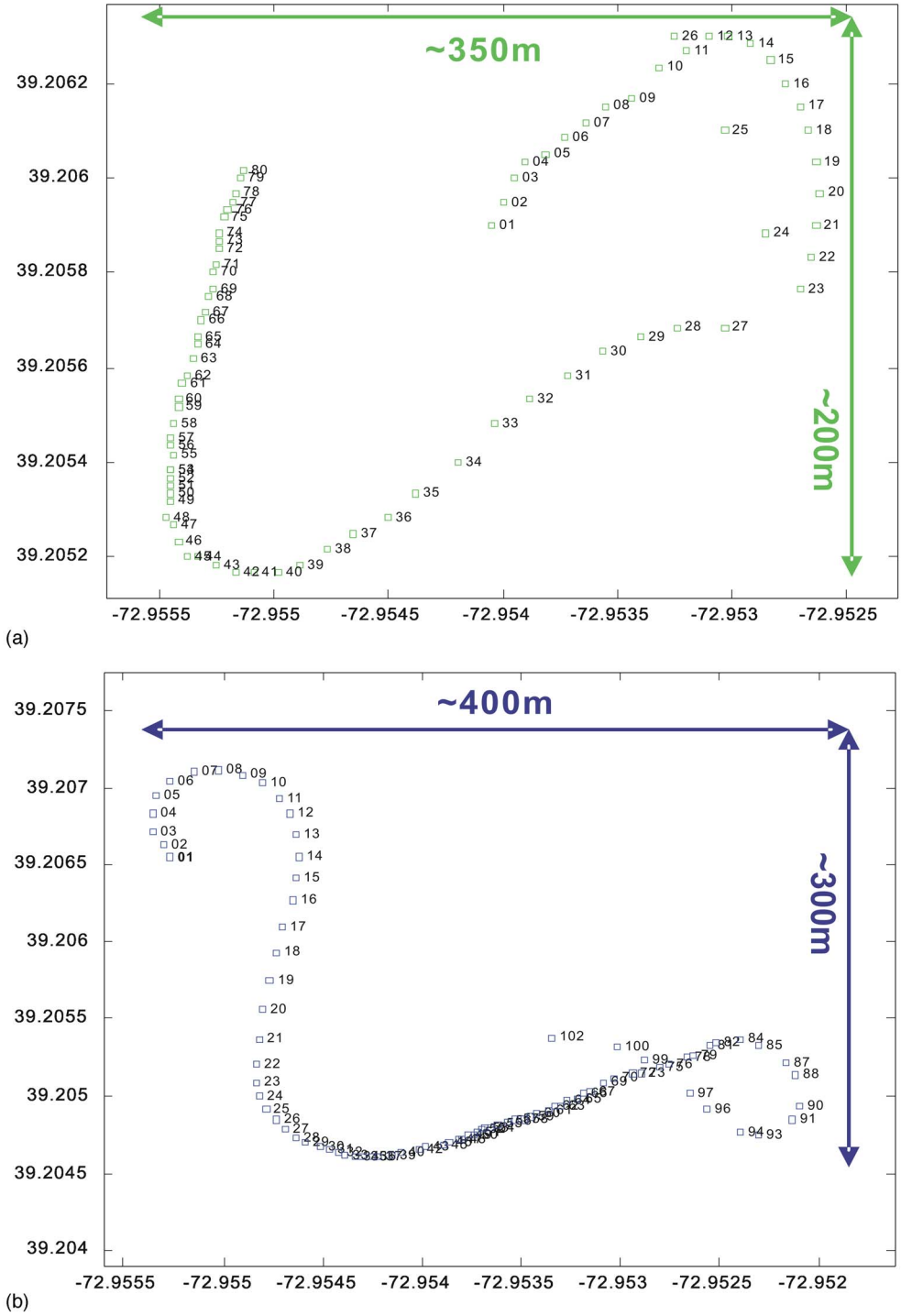


FIG. 2. (Color online) Source positions and maneuvering routes for (a) Track01 and (b) Track02.

propagation in marine sediments. Instead we interpret the results in the context of partly due to the depth-penetration in the sediment and the interaction with a layered system of sediment material with depth-dependent geoacoustic properties.

The remainder of this paper is organized as follows. Section II briefly describes New Jersey Shelf environment, the air gun data, and the relevant information of the SWARM experiment. Section III describes the geoacoustic model, reviews the inversion procedure, outlines the method for estimating full covariance matrices, and describes the statistical tests that were applied to the standardized data error residuals. Section IV presents the inversion results from optimization and Bayesian inversion approaches, discusses the consistency of the geoacoustic model estimates at different source depths and different frequency bands, and compares

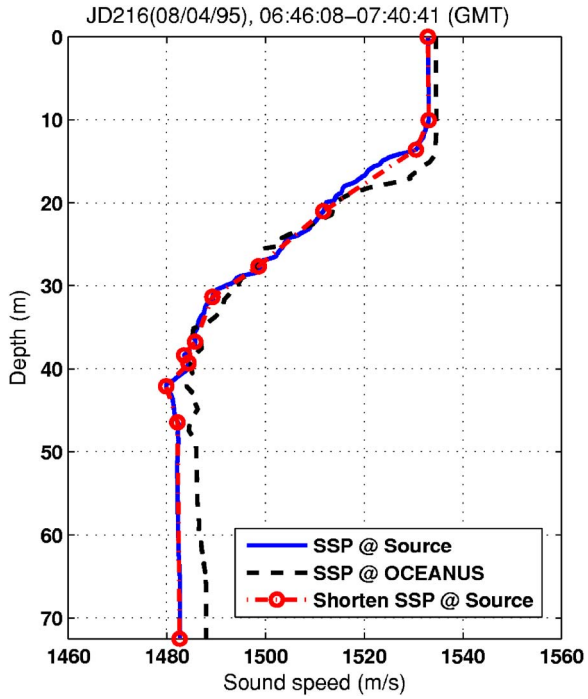


FIG. 3. (Color online) Sound speed profile for the data analyzed.

the uncertainty estimates from Bayesian inversions using three different data covariance matrix estimates. Section V summarizes the work.

II. DESCRIPTION OF THE EXPERIMENT

A. Experimental geometry and the source tracks

The SWARM experiment was carried out in the Mid-Atlantic Bight continental shelf region off the coast of New Jersey in the summer of 1995.[27] The data used in this paper were recorded in a subexperiment on 4 August 1995 (corresponding to JD 216) at the site shown in Fig. 1. The receiver was a vertical line array (VLA) from Woods Hole Oceanography Institute (WHOI), which was deployed at the location of (72°55′10.92″W, 39°14′24.72″N). The WHOI-VLA consisted of 16 hydrophones evenly spaced at a separation of 3.5 m from 14.9 to 67.4 m below the sea surface. The data we analyze in this paper are three data sets (we refer to these three data sets as Track00, Track01, and Track02 hereafter) that were collected in the time segment between JD216 06:46:08 and JD216 07:40:41 coordinated universal time (UTC) along the source ship maneuvering tracks shown at the lower-left corner of Fig. 1. The bathymetric contours[29] shown in Fig. 1 indicate that the variation of the water depth between the source and the receiver was small, from 72 to 73 m.

A small 20 $in.^3$ air gun source was deployed at two nominal depths of 25 and 45 m from a slowly maneuvering ship, for the purpose of studying the depth dependence of acoustic propagation. The air gun source was fired at an interval of about 20 s. Track00 contains shots fired when the source ship was moving toward the WHOI-VLA over a distance of 900 m at a range of ~5 km. Track01 and Track02 contain shots fired when the source ship was drifting along the routes shown in Fig. 2 about 4.5 km from the array. The numbers shown in Fig. 2 are the corresponding shot numbers. The drifting distances of the source ship in Track01 were 200 m in the north-south and 350 m in the east-west directions, while in Track02 the distances were 300 m in the north-south and 400 m in the east-west directions, respectively. Table I shows the relevant information of these three data sets.

The spectrum of an air gun signal is broadband with several natural modulations in the frequency domain that arise due to the bubble pulse oscillation. The useful frequency band with adequately high signal-to-noise ratio (at least 10 dB) was from 30 to 200 Hz. This band was further separated into two subbands from 35 to 90 Hz (FB1, frequency band 1) and from 120 to 180 Hz (FB2, frequency band 2) for the inversions reported here.

B. Water column sound speed profile

CTDs (conductivity, temperature, and depth) were measured at the source position several times during the experiment between JD216 06:46:08 and JD216 07:40:41 UTC. Only one of the CTD files (c0804012.asc) was chosen to derive the sound speed profile (SSP) in the water column for the time period of our study because there was negligible change observed in the set of measurements. Although there were no CTD measurements at the receiving array position, a CTD collected by R/V OCEANUS about 3.86 km to the northwest was used as a reference.

A comparison display of the SSPs at the source and at R/V OCEANUS is shown in Fig. 3. Since both of the SSPs have very similar structure (i.e., isospeed at the surface and the bottom, large negative gradient and approximately the same thickness of the thermocline) and since there were no internal wave events observed within the time segment of the air gun experiment, we assume that the water column environment is range independent.

The SSP was further downsampled to 12 points, as shown in Fig. 3 with the circled dash-dotted line, in order to improve the computational efficiency in the acoustic field calculation. It should be mentioned that the sound speed at the water bottom was extrapolated according to the measured data.

III. GEOACOUSTIC INVERSION APPROACH

A. General

Matched field inversions of the broadband air gun data were carried out using two approaches, optimization and appraisal analysis based on Bayesian theory, in two low frequency bands with center frequencies separated by about an octave: 35–90 Hz centered at 65 Hz, and 120–180 Hz centered at 150 Hz. The range-independent normal mode acoustic propagation model ORCA[30] was used to compute the replica sound fields in both approaches. ORCA provides two computational options, i.e., real-axis option and complex-axis option. When the real-axis option is chosen, ORCA solves the eigenvalue problem on the real axis and uses a perturbation technique to approximate the imaginary part of the horizontal wave number that accounts for seabed attenu-

TABLE I. Air gun data set information.

Data set name	Number of shots	Source depth (m)	Start time	End time
Track00	28	25	06:27:16	06:37:00
Track01	80	45	06:38:00	07:06:00
Track02	103	25	07:10:00	07:46:00

ation. It is very efficient in terms of both the computation time and accuracy under the condition that the range to water depth ratio is large. When shear wave (*s*-wave) effects are included, or in a short-range propagation situation, the complex-axis option is recommended since ORCA will take improper modes into account and handle the modal attenuation in an exact manner. However, the complex-axis option is computationally intensive. In the environment (range and water depth) of this paper's study, for one forward call of ORCA, the real-axis option is 20 times faster than the complex-axis option at the frequency of 65 Hz, and 10 times faster at 150 Hz.

B. Geoacoustic model

A high resolution multilayer geoacoustic model was derived from previous geological and geoacoustic surveys in the New Jersey continental shelf region.[31–36] The structure of the seabed is believed to have an "S" reflector near the surface layer, subseafloor channels and the "*R*" reflector at depths of 8–20 m below the sea floor (bsf) that marks the change to significantly higher sound speeds of around 1900 *m*/s.[35,36] Analysis of the AMCOR (Atlantic Margin Coring Project) borehole No. 6010 suggests the presence of an additional underlying high speed layer at a deeper depth between 85 and 110 m bsf.[34]

Compressional wave (*p*-wave) speed of the sediment was derived from chirp sonar inversions,[33] the grab samples and *in situ* probes,[31,32] and shallow core measurements.[35] The distribution of the samples is shown in Fig. 4. The mass density was determined from the empirical relationship between the *p*-wave speed and the density by Richardson,[37] the estimates from chirp sonar signal by Turgut *et al.*,[33] and the compliance work of Trevorrow and Yamamoto.[34] The *p*-wave attenuation estimate was inferred first from the relationship of the *p*-wave sound speed with sediment mean grain size, along with the relationship of *p*-wave sound speed with sediment porosity, and then from mean grain size and porosity to *p*-wave attenuation. The overall result derived from these data was anomalously high, around 0.6–0.7 dB/m at 1 kHz.

Figure 5 shows the model of *p*-wave velocity in the upper 20 m bsf compared with the core data and chirp sonar inversion results. Figure 6 shows the high resolution geoacoustic model that was derived from all the available ground truth data in the region.

Based on the small variation of the bathymetry at the experimental site, we assume that the geoacoustic model is range independent, and use a simplified one-layer over half space geoacoustic model, as shown in Fig. 7, for the inversions presented in this paper. The long-range source-receiver geometry supports multiple bottom interacting paths that propagate at low angles. Since we ignore the spatial variability of the properties of the seabed, the inversion result one should expect is an average over range and depth in the sediment. The weighted average over depth depends on the frequency band used in the inversion. According to the high resolution model, the weighted *p*-wave sound speed for the four layers down to about 20 m is 1703 m/s and the mass density is 1.986 g/cm^3, while the weighted average values of

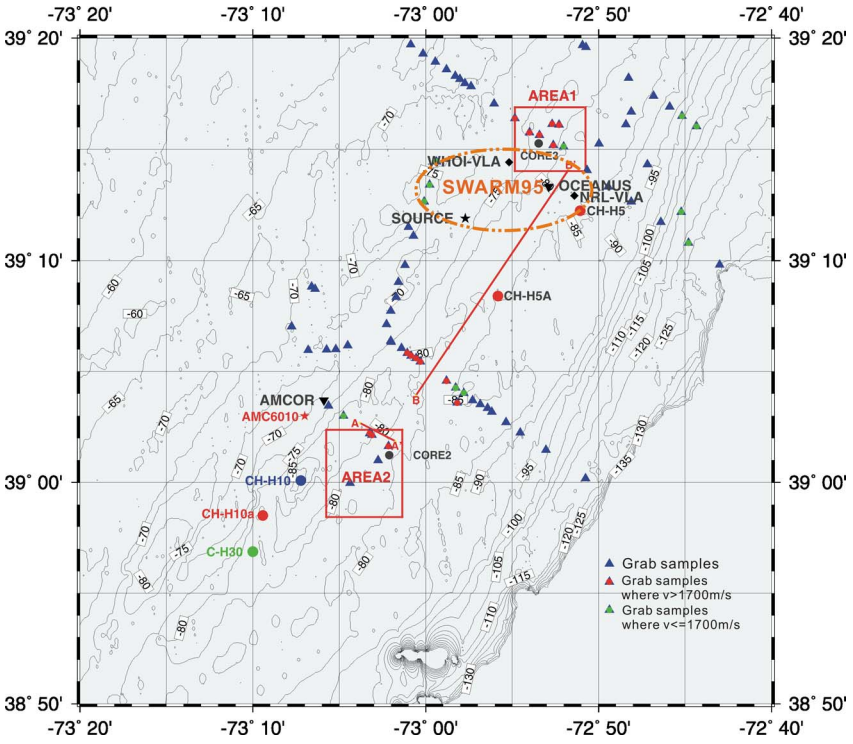


FIG. 4. (Color online) Distribution of preexisting "Ground truth" information of the New Jersey continental shelf near the experimental site of SWARM 95.

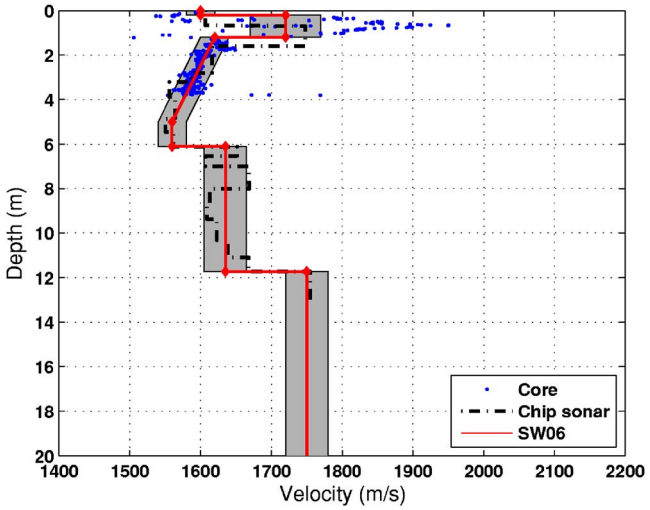


FIG. 5. (Color online) Comparison of p-wave velocity between core analysis results, estimation from chip sonar signal and the p-wave velocity model for shallow water acoustic experiment 2006 (SW06).

p-wave sound speed and mass density down to 10 m are around 1650 m/s and 1.91 g/cm^3, respectively. Simulations prior to real data inversion affirmed that a one-layer model represented the multilayer high resolution geoacoustic model very well in terms of a weighted average when the signal to noise ratio (SNR) was larger than 15 dB.

C. Parameters to be inverted

The parameters that were inverted included geoacoustic parameters for the single layer model and geometrical parameters of the experimental arrangement. The geoacoustic parameters were: sediment depth, the p-wave and s-wave speeds and attenuations, and the density, in the sediment and in the lower half space. Source depth, range, and water depth were also inverted. The long-range, low-angle propagation geometry in this experiment may increase the sensitivity to shear wave parameters, compared to a close range geometry for which the propagation is at higher angles near normal incidence. In the optimization inversions, each shot transmission was inverted with and without shear wave parameters of the seabed to investigate the resolvability of the shear wave parameters and the effect of these parameters on inversion performance in the two low frequency bands.

Table II shows the search bounds for the parameters. Due to the computationally intensive constraints of using the

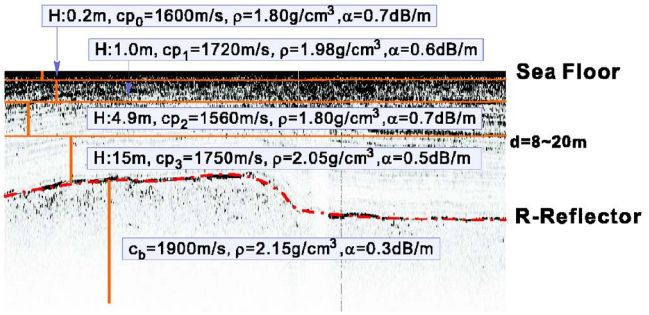


FIG. 6. (Color online) Multilayer high resolution geoacoustic model. The attenuation is the value at 1 kHz.

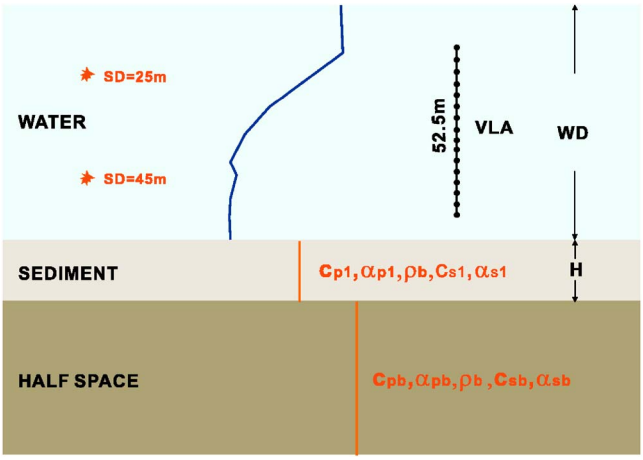


FIG. 7. (Color online) Single homogeneous sediment layer over lower half space geoacoustic model.

complex axis option of ORCA, the search bounds of some of the parameters were narrowed intentionally when including shear wave parameters in the inversions. The search bounds of source-receiver range for Track00 shots are from 4.5 to 6.5 km, and for Track01 and Track02 shots from 4.0 to 5.0 km.

D. Optimization approach

The optimization algorithm used in this study is adaptive simplex simulated annealing,[8] which is an efficient hybrid algorithm that combines the global search algorithm simulated annealing and a local optimal search algorithm downhill simplex. The typical annealing schedule that was used in this study is: starting temperature 0.3, temperature reduction factors of 0.992 and 0.994 for the inversion excluding/including shear wave parameters, respectively, and number of temperature steps 1100. Although this approach does not provide information of the uncertainties of the estimates, it does provide a useful qualitative sense of parameter sensitivity.

TABLE II. The search bounds of the model parameters to be inverted.

	Search bounds			
Parameters	Without shear		With shear	
WD(m)	[65	80]	[70	75]
H(m)	[1	80]	[1	80]
c_{p1}(m/s)	[1400	1800]	[1600	1800]
c_{pb}(m/s)	[1600	2500]	[1600	2500]
α_{p1}(dB/m kHz)	[0	0.5]	[0	0.5]
α_{pb}(dB/m kHz)	[0	1.0]	[0	1.0]
ρ_1(g/cm^3)	[1.10	2.50]	[1.10	2.50]
ρ_b(g/cm^3)	[1.50	3.00]	[2.00	3.00]
SD(m)	[1	60]	[1	60]
Range(km)	[4.0	5.0]	[4.0	5.0]
c_{s1}(m/s)	⋯	⋯	[1	500]
c_{sb}(m/s)	⋯	⋯	[1	1000]
α_{s1}(dB/m kHz)	⋯	⋯	[0	2]
α_{sb}(dB/m kHz)	⋯	⋯	[0	3]

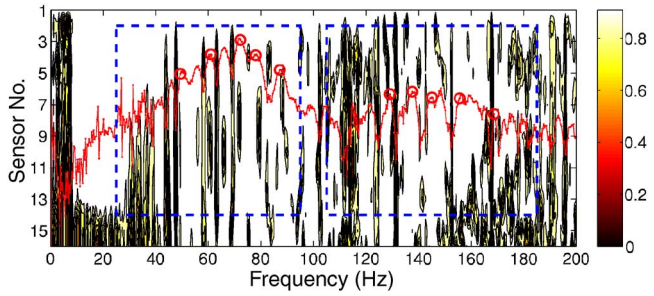


FIG. 8. (Color online) Spatial coherence coefficient for the signals from the hydrophones across the VLA with respect to the signal from top hydrophone. The curve with open circles is the spectrum of the signal from the top hydrophone. The open circles are the frequency components used in the inversion. The dashed rectangles indicate the two frequency bands that were considered in this paper.

The objective function that was applied to the data was a multifrequency mismatch function based on the spatially coherent Bartlett processor averaged incoherently over multiple frequencies:

$$E(m) = 1 - \frac{1}{N_F}\sum_{i=1}^{N_F} \frac{|\mathbf{d}_f^+(m)\mathbf{d}_f^{\text{obs}}|^2}{|\mathbf{d}_f^+(m)|^2|\mathbf{d}_f^{\text{obs}}|^2}, \tag{1}$$

where $\mathbf{d}_f^{\text{obs}}$ represents the measured (observed) complex pressures of the hydrophones on the VLA at frequency f; $\mathbf{d}_f(m)$ represents the modeled complex pressures of the hydrophones on the VLA for a set of proposed model parameters m at frequency f; the superscript + denotes conjugate transpose, and N_F is the number of frequencies, which was 5 in these inversions. Note that the comparison in Eq. (1) is between the measured and modeled fields; there is no assumption in the optimization approach on the spatial correlation and the statistics of the data error.

MFI is based on the spatial coherence information of the signals across an array at frequencies of interest. For the broadband air gun data, the spectral components used in the inversions were selected according to the following criteria: (1) high SNR (at least 10 dB); (2) enough separation in frequency; and (3) high spatial coherence coefficient of the measured pressure at the hydrophones across the array. The spatial coherence coefficient of the hydrophone signal at the depth $z_0+\Delta z$ with respect to the signal of the reference hydrophone at the depth of z_0 is defined as

TABLE III. Important properties defining parameter estimates and uncertainties.

Property	Formula	Eq. no.
Maximum *a posteriori*	$\hat{m}=\text{Arg}_{\max}\{P(m\|\mathbf{d}^{\text{obs}})\}$	(3)
Posterior mean estimate	$\bar{m}=\int mP(m\|\mathbf{d}^{\text{obs}})dm$	(4)
Marginal probability distribution	$P(m_i\|\mathbf{d}^{\text{obs}})=\int\delta(m_i'-m_i)P(m'\|\mathbf{d})dm'$	(5)
Model covariance matrix	$C^M=\int(m-\bar{m})(m-\bar{m})^TP(m\|\mathbf{d}^{\text{obs}})dm$	(6)

$$I_S(f,z_0,z_0+\Delta z) = \frac{\left|\sum_{i=-(w-1)/2}^{(w-1)/2} d_{\text{obs}}^*(f,z_0)d_{\text{obs}}(f,z_0+\Delta z)\right|^2}{\left(\sum_{i=-(w-1)/2}^{(w-1)/2}|d_{\text{obs}}(f,z_0)|^2\right)\left(\sum_{i=-(w-1)/2}^{(w-1)/2}|d_{\text{obs}}(f,z_0+\Delta z)|^2\right)}, \tag{2}$$

where $d_{\text{obs}}(f,z_0)$ is the complex pressure at the reference depth z_0 and frequency f, Δz is the difference in depth of the two hydrophones, w is the moving average window in the frequency domain, and the asterisk (*) is the conjugation operator. Values of 3 or 5 were used for w in applying this expression to the shot data. The frequency separation was arranged to obtain approximately equal separation within each frequency band. For example, the effective bandwidth of the higher frequency band was around 40% of the center frequency. Since there were five spectral components used in the inversion, the frequency separation was approximately 10% of the center frequency, i.e., around 15 Hz.

Figure 8 shows the contour plot of the spatial coherence coefficient for all the array hydrophones derived from one of the shots. The reference hydrophone in this plot is the shallowest in the water at the top of the array. The x axis of the contour plot is spectral frequency and the y axis is hydrophone number, used as an alternative representation of the hydrophone depth with smaller numbers indicating shallower depths in the water column. The red curve overlapped on top of the contour plot is the spectrum of the signal from the shallowest hydrophone, as an indication of the SNR of the spectral component. The two rectangular boxes in dashed lines confine the two frequency bands for the inversions. Figure 8 shows several regions of high spatial coherence (light colors from top to bottom in the plot) in both frequency bands. The spectral components with high spatial coherence and high SNR that were selected for both of the frequency bands for this shot are shown by the open circles on the spectrum plot.

E. Bayesian inversion approach

Bayesian geoacoustic inversion provides maximum *a posteriori* model parameter estimates and the uncertainty of the estimates, by interpreting the multidimensional posterior probability density (PPD) that summarizes the information about the model parameters in the measured data. The full Bayesian formulation of inverse theory and some successful applications incorporating matched field processing are found in Refs. 3, 20–22, 24, and 25. Table III lists the properties of the multidimensional PPD that are most commonly used to interpret the inversion estimates and their uncertainties, where $P(m|\mathbf{d}^{\text{obs}})$ is the PPD and δ is the Dirac delta function in Table III. Since geoacoustic inversion is a highly nonlinear problem with no direct analytic solution, numerical approaches are the practical alternatives to evaluate the integrals listed in Table III.

This paper adopts the Markov-chain Monte Carlo method of fast Gibbs sampling of Dosso[20] and Dosso and Nielson[21] to evaluate the integrals, and follows the method of determining the data residuals described in Ref. 24. However, we employ a different approach to estimate the data error covariance matrix. For completeness, the procedure is outlined as follows.

Assuming that the errors on the data, $\mathbf{d}_f^{\text{obs}}$, are complex, zero-mean Gaussian-distributed random variables and the separation of the frequency components to be used in the inversion is large enough that the errors are not correlated from frequency to frequency, the error correlation between the hydrophones is represented by the covariance matrix C_f at the fth frequency and the likelihood function is given by

$$L(m) = \prod_{f=1}^{N_F} \frac{1}{\pi^{N_H}|C_f|} \exp\{-[\mathbf{d}_f^{\text{obs}} - \mathbf{d}_f(m)]^+ C_f^{-1}[\mathbf{d}_f^{\text{obs}} - \mathbf{d}_f(m)]\}. \tag{8}$$

When the phase θ_f and the amplitude A_f of the source are unknown, which is common in matched-field applications, the modeled data can be written as $\mathbf{d}_f(m) = A_f e^{i\theta_f}\mathbf{p}_f(m)$, where $\mathbf{p}_f(m)$ is the modeled complex pressure on N_H hydrophones at frequency f. Substituting the expression for the modeled pressure with unknown source information into Eq. (8) and maximizing the likelihood function with respect to the unknown source phase and amplitude to solve for $A_f e^{i\theta_f}$, then the likelihood becomes

$$L(m) = \prod_{f=1}^{N_F} \frac{1}{\pi^{N}|C_f|} \exp\left\{-\left[(\mathbf{d}_f^{\text{obs}})^+ C_f^{-1}\mathbf{d}_f^{\text{obs}} - \frac{|\mathbf{p}_f^+(m)C_f^{-1}\mathbf{d}_f^{\text{obs}}|^2}{\mathbf{p}_f^+(m)C_f^{-1}\mathbf{p}_f(m)}\right]\right\}. \tag{9}$$

The corresponding misfit function is the sum of the covariance-weighted Bartlett mismatch at N_f frequencies:

$$E(m) = \sum_{f=1}^{N_F}\left[(\mathbf{d}_f^{\text{obs}})^+ C_f^{-1}\mathbf{d}_f^{\text{obs}} - \frac{|\mathbf{p}_f^+(m)C_f^{-1}\mathbf{d}_f^{\text{obs}}|^2}{\mathbf{p}_f^+(m)C_f^{-1}\mathbf{p}_f(m)}\right]. \tag{10}$$

Note that in Eq. (9), there are two different sources of errors that must be considered: measurement errors and theory errors. Measurement errors are associated with the measurement process, and include the ambient noise from the experimental environment, the error caused by the inconsistency of the complex gain (amplitude and phase) between the channels of the data acquisition system, and the uncertainties of the hydrophone positions, etc. Theory errors arise from inaccurate model parametrization and any incorrect or inaccurate assumptions in the acoustic forward model. Since we do not have access to the true model, there is always an intrinsic mismatch with the measured field for any assumed geoacoustic model.

The approach we have taken for estimating the error covariance matrix makes use of the information from the multiple shot transmissions in this experiment. By definition, the estimate of C_f is obtained by ensemble averaging the data errors over the set of inversion realizations from many different shots:

$$C_f = \langle(\mathbf{r}_f(m) - \langle\mathbf{r}_f(m)\rangle)(\mathbf{r}_f(m) - \langle\mathbf{r}_f(m)\rangle)^+\rangle, \tag{11}$$

where $\mathbf{r}_f$ is the data residuals at frequency f given by

$$\mathbf{r}_f(m) = \mathbf{d}_f^{\text{obs}} - \frac{\mathbf{p}_f^+(m)\mathbf{d}_f^{\text{obs}}}{|\mathbf{p}_f(m)|^2}\mathbf{p}_f(m). \tag{12}$$

Ensemble averaging of the residuals requires a large collection of data from similar experiments to obtain the proper knowledge of the statistics of the errors. Although there are a large number of shots in this experiment, natural decorrelation processes in the ocean constrain the actual number that can be included in the ensemble. For instance, since the acoustic channel is a slowly time varying and spatially varying system, the frequency components that are suitable for the inversion for one shot might not be suitable for other shots after a period of time. To find enough data to estimate C_f, shots with least spatial variation were chosen at first according to the position information in Figs. 2(a) and 2(b). Then, the time coherence of the signal for the hydrophone at the same depth was checked for these shots according to the definition of the time coherence coefficient:

$$I_T(f,t_0,t_0+\tau) = \frac{\left|\sum_{i=-(w-1)/2}^{(w-1)/2} d_{\text{obs}}^*(f,t_0)d_{\text{obs}}(f,t_0+\tau)\right|^2}{\left(\sum_{i=-(w-1)/2}^{(w-1)/2}|d_{\text{obs}}(f,t_0)|^2\right)\left(\sum_{i=-(w-1)/2}^{(w-1)/2}|d_{\text{obs}}(f,t_0+\tau)|^2\right)}, \tag{13}$$

where $d_{\text{obs}}(f,t_0)$ is the complex pressure at the reference time t_0 at frequency f, $d_{\text{obs}}(f,t_0+\tau)$ is the complex pressure of the signal from the same hydrophone at frequency f after time delay τ, and w is the average window in frequency domain (set equal to 3 or 5 for this application). The shots were selected from the two drifting experiments, track01 and track02, since the spatial change from shot to shot was very small. The shot positions were within 50 m in longitude and 70 m in latitude for track01, and within 90 m in latitude and 25 m in latitude for track02.

An example of this approach is shown in Fig. 9 which shows the variation of the time coherence of the acoustic pressure within the effective frequency bands for the shallowest array hydrophone for a reference shot in track01 (the display also reflects the variation of the acoustic pressure when the source position changes spatially from shot to shot as the source ship drifted). The reference shot number in Fig. 9 was shot 40. The x axis represents frequency, and the y axis is time difference represented in terms of shot number instead of seconds (the actual time difference is given by the product of the difference of shot number and transmitting interval). Similar to Fig. 8, the lighter color indicates higher time coherence in the contour plot. The estimation of C_f in our approach requires the information in the residuals determined from a large number of consecutive shots. The simple analysis of the time coherence summarized in Fig. 9 gives additional information about the frequency components that have high SNR in consecutive shots and have not faded due

to the natural decorrelation processes of the acoustic channel. These components are the ones that can be used to estimate C_f. By taking the results from all of the hydrophones into account, the number of shots that could be used to estimate C_f was 20 (from shot 31 to 50) and the corresponding time duration was approximately 400 s. Since the source position variations are relatively small, and the residuals at each depth are assumed to be stationary within this short period of time, we used an arithmetic average of the residuals for the hydrophones over the N_{shot} shots to approximate the estimate of C_f. Consequently, the expression for C_f becomes

$$C_f = \frac{1}{N_{\text{shot}}} \sum_{i=1}^{N_{\text{shot}}} \left\{ \left[\mathbf{r}_f(m) - \frac{1}{N_{\text{shot}}} \sum_{j=1}^{N_{\text{shot}}} \mathbf{r}_f(m) \right] \times \left[\mathbf{r}_f(m) - \frac{1}{N_{\text{shot}}} \sum_{j=1}^{N_{\text{shot}}} \mathbf{r}_f(m) \right]^+ \right\}. \quad (14)$$

The residuals of the data on the hydrophones across the array were obtained from the differences between the measured data and modeled data that were calculated by ORCA using the model parameter estimates from an optimization inversion for each of the 20 different shots. When estimating the data residuals, the data error covariance matrix is assumed to be a diagonal matrix with identical variances that has the form of $C_f = \sigma_f^2 I$. This is not unreasonable since there is no prior knowledge about data error correlations between the hydrophones. In our approach, the information about the error correlations contained in the experimental data is mapped directly into the residuals. The cost function for this process was obtained by substituting the approximate form for C_f into Eq. (8) and maximizing the likelihood function over the variance σ_f^2 to get the estimate of $\hat{\sigma}_f^2$. Substituting this expression for $\hat{\sigma}_f^2$ into the likelihood function obtains the misfit function:[24]

$$E(m) = N_H \sum_{i=1}^{N_F} \log_e |\mathbf{d}_f^{\text{obs}} - \mathbf{d}_f(m)|^2, \quad (15)$$

where N_H is the number of hydrophones. Equation (15) was minimized by the optimization approach to obtain the model parameter estimates $\hat{m}$ that were used to estimate the data residuals $\mathbf{r}_f$ defined by Eq. (12).

The assumption of stationarity of the residuals for the same hydrophone from the ensemble of consecutive shots that were used to estimate C_f has been examined by applying the Kolmogorov-Smirnov (KS) two-distributions test.[38] The residuals of each hydrophone are divided into two parts with equal number of data points. The null hypothesis H_0 is that the data points of the two groups are drawn from the same distribution, and the alternative hypothesis H_1 is that the data points of two groups are drawn from different distributions. The maximum difference between the cumulative distribution functions of the data points from the two groups is used to determine the decision to accept or reject H_0. The real and imaginary parts of the residuals have been examined separately at each frequency and hydrophone. The KS test was applied to four Bayesian inversion cases (two source depths, each with two frequency bands) with a significance level α of 0.05. For each case, the KS test was applied 160 times, (16 hydrophones $\times$ 5 frequency components = 80 groups of residuals, and each group of the residuals has a real part and an imaginary part), and the test results are listed in Table IV. For the worst case, the number of tests rejecting H_0 is 25 out of 160, which means for most of the cases, there is no strong evidence against the assumption of stationarity of the residuals at the significance level of 0.05.

IV. GEOACOUSTIC INVERSION RESULTS

We first present the parameter estimates from the optimization method, and discuss the sensitivities of the model parameters. For the Bayesian inversion, we compare the results obtained using the full data covariance matrix, diagonalized full covariance and identical variances diagonal matrix to illustrate the impact of data errors on the inversion. The estimated geoacoustic parameters are then discussed in the context of the ground truth information.

A. Optimization inversion

The optimization inversion was first applied to all three data sets without considering shear wave effects. Histograms of the inversion results for track01 and track02 are shown in Fig. 10 and summarized in Table V. The estimates from track00 are not included in the histogram for the 25-m shots because the number of shots is too small (only 28); however, all the results for the 25-m shots from both tracks are summarized in Table V. Since the source is moving in all three cases, the range estimates were not listed in Table V, and were excluded in Table VI for the same reason.

From the spread of values in the histograms in Fig. 10 we can infer that:

(1) Source depth, SD, water depth, WD, and sediment p-wave sound speed, c_{p1}, are very well determined with small variance. The estimates of water depth are consistent at the two different source depths and frequency bands, and the source depth estimates are consistent at the two frequency bands. The p-wave sound speed estimates at the higher frequency band are slightly lower than the estimates at lower frequency band.
(2) p-wave attenuation of the sediment is well estimated also, although the relative uncertainty is larger compared

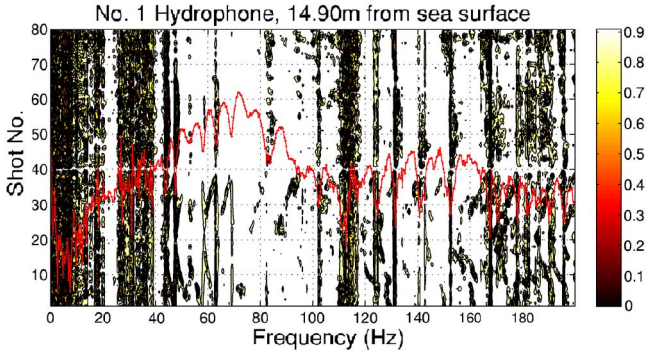


FIG. 9. (Color online) Example of time coherent analysis of the signals of all shots at the shallowest hydrophone with respect to the reference signal. In this plot, the signals are from Track01, the reference signal is from shot No. 40.

TABLE IV. KS test results for the test of stationarity of the data residuals.

	Number of cases that reject H_0 in FB1			Number of cases that reject H_0 in FB2		
	Real	Imaginary	Rejection/total	Real	Imaginary	Rejection/total
SD=45 m	8	17	25/160	4	5	9/160
SD=25 m	11	7	18/160	6	6	12/160

to that for SD, WD, and c_{p1}. The p-wave attenuation estimates for both frequency bands at the two different source depths are consistent. Moreover, the mean values (in dB/m at 1 kHz) of the estimates at FB2 are approximately twice as large as those for FB1, for both source depths.

(3) Sediment density is not as well estimated as sediment p-wave attenuation.
(4) Sediment depth is not well determined. For both source depths, at the lower frequency band the mean values of the estimate are deeper, while at higher frequency band they are shallower.
(5) p-wave sound speed, p-wave attenuation and density for lower half space are not resolvable in the inversion.

Optimization inversion was applied to the three data sets again at both of the frequency bands with the complex axis option of ORCA enabled to examine the resolvability of shear wave parameters and to determine the influence of estimating these parameters on the values estimated for the other parameters. From the results listed in Table VI, we note that:

(1) For the most sensitive parameters (i.e., SD, WD, and c_{p1}), the estimated values remain approximately the same.
(2) The standard deviations of the estimates of s-wave speed and attenuation for both the sediment and the lower-half space are large, indicating low sensitivity for all of the shear wave parameters.
(3) For all of the cases (different source depths at different frequency bands), p-wave attenuations of sediment are slightly decreased, and the density of the sediment is slightly increased. Although this result is consistent with the behavior predicted by Tindle and Zhang's[39] model of complex density, the large standard deviations of the estimates of all the shear wave parameters preclude making any definite conclusions.

B. Bayesian inversion

1. Data error covariance

The structures of the full data covariance matrices were estimated according to the procedure in Sec. III E and are shown in Fig. 11 for the five frequencies used in the inversion. The matrices are Hermitian, with nonidentical variances on the diagonal elements of the real parts of the covariance matrices and notable nonzero off-diagonal values on both real and imaginary parts of the matrices. The structure is clearly not Toeplitz, as assumed by Dosso *et al.*[24]

A qualitative way to check the assumption of randomness of the data residuals is to examine the autocorrelation of the residuals. A sharp peak with small correlation radius in the real part and close-to-zero values in the imaginary part indicate an ideal uncorrelated random process. Following Ref. 24, we first define a standardized data residual, the residual weighted by the data covariance, as $[\hat{C}_f^{-1/2}]^+\mathbf{r}_f(m)$, where $\hat{C}_f^{-1/2}$ is the inverse of the upper-triangular Cholesky decomposition of the data error covariance matrix estimated according to Eq. (14). Figure 12 shows the real and imaginary parts of the autocorrelation at the five different frequen-

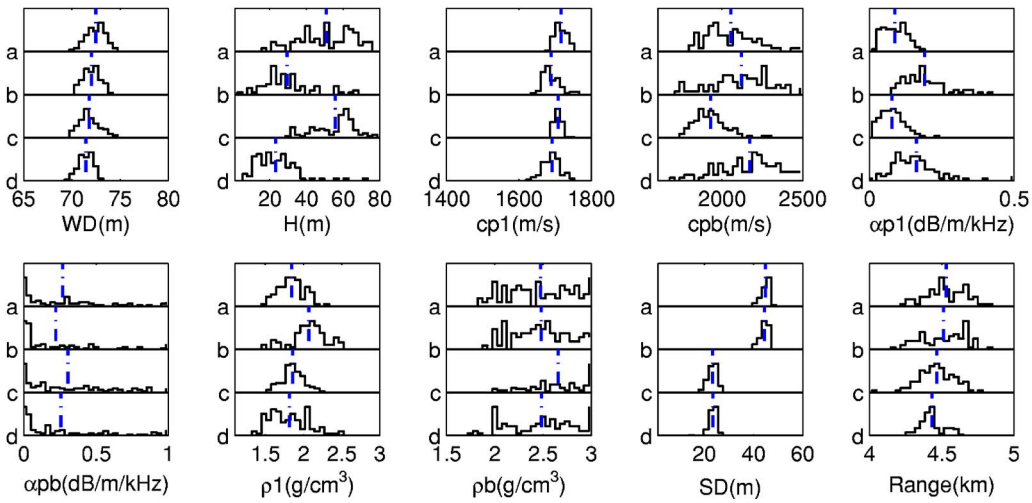


FIG. 10. (Color online) Histograms of the optimal inversion results of all of the shots from Track01 and Track02. Shear wave effects were excluded in the inversion. (a) SD=45 m, at frequency band 1; (b) SD=45 m, at frequency band 2; (c) SD=25 m, at frequency band 1; and (d) SD=25 m, at frequency band 2.

TABLE V. Summary of the means and standard deviations of the optimal estimates, excluding shear wave effects.

	Frequency band 1			Frequency band 2	
Parameters	Track00 (28 shots)	Track01 (80 shots)	Track02 (103 shots)	Track01 (80 shots)	Track02 (103 shots)
SD (m)	20.5±2.0	44.7±1.7	23.3±1.8	44.3±1.6	23.5±1.6
WD (m)	71.9±1.2	72.5±1.0	71.8±1.1	72.1±0.9	71.5±0.8
c_{p1} (m/s)	1704.2±8.6	1716.3±15.4	1708.1±10.9	1688.7±23.9	1690.6±24.0
α_{p1} (dB/m/kHz)	0.074±0.043	0.089±0.039	0.079±0.041	0.190±0.073	0.163±0.102
ρ_1 (g/cm^3)	1.92±0.17	1.84±0.17	1.86±0.15	2.07±0.23	1.81±0.26
H (m)	62.2±11.4	50.8±14.9	55.6±11.5	29.4±13.8	23.4±11.0
c_{pb} (m/s)	1964±118	2055±156	1928±99	2117±186	2170±188
α_{pb} (dB/m/kHz)	0.378±0.300	0.272±0.295	0.306±0.295	0.221±0.290	0.257±0.307
ρ_b (g/cm^3)	2.61±0.33	2.47±0.34	2.66±0.30	2.48±0.29	2.48±0.35

cies for several types of residuals: the original residuals (open circles—blue solid line); the standardized residuals weighted by the full covariance matrices (closed diamonds—red solid line), diagonalized full covariance matrices that were constructed by ignoring the covariance of the errors between the hydrophones (setting the off-diagonal elements of the full covariance matrix to zero) (open squares—black dashed line); and identical variance diagonal matrices (closed triangles—blue dotted line). Only the residuals weighted by the full covariance matrices display the behavior expected for ideal random processes. Significant spatial correlation is evident for all other types of residuals.

2. Statistical tests of the assumptions made in Bayesian inversion approach

Statistical tests were carried out to obtain more quantitative information about the assumptions of Gaussianality and randomness on the data error residuals. The KS test for normality (i.e., Gaussian distributions) and the runs test for randomness were applied to the standardized data residuals and the original residuals. Both tests used a significance level of $p=0.05$ for hypothesis evaluation. There are a total of 40 series of residuals being examined (2 source depths $\times 2$ frequency bands $\times 5$ frequency components at each frequency band, and each has real and imaginary parts). For the KS test, all of the original series rejected the null hypothesis that the residuals were Gaussian, but only 3 of the 40 standardized residuals series rejected the null hypothesis with a p value between 0.03 and 0.05. For the runs test, 29 of the 40 original series rejected the null hypothesis that the residuals were random, but only 2 of the 40 standardized residuals series rejected the null hypothesis with a p value between 0.04 and 0.05. The results from both tests indicate that there is no strong evidence against the assumptions that the standardized data residuals weighted by full data error covariance matrix are Gaussian distributed and spatially uncorrelated at the significance level of 0.05.

The assumption that the residuals are uncorrelated between the frequencies was examined qualitatively by examining the cross correlation of the covariance weighted residuals series from different frequencies, on both real and imaginary parts.

3. Impact of correlated data errors

In order to demonstrate the effect of data error correlations on the parameter uncertainties, Bayesian inversion was applied to the same shot from track01 at the lower frequency band using the three different approaches of estimating data error covariance matrices described earlier.

The comparison of the Bayesian inversion results is shown in Fig. 13. The red dash-dotted line indicates the MAP estimates and the blue dashed line indicates the mean values. The uncertainty estimates obtained by using full covariance matrices (row a) are consistently smaller than those of the other two approaches, and, overall, the qualitative information obtained about parameter sensitivity in the optimization inversions is confirmed by the Bayesian results. The approach taken in this inversion of selecting spectral components with high SNR and high spatial coherence (to ensure data quality) has improved the overall performance obtained using diagonalized full covariance matrices and identical variance diagonal matrices.

It is important to note that the objective in deriving the error correlation information for use in the inversion is not to

TABLE VI. Summary of the means and standard deviations of the optimal estimates, including shear wave parameters.

	Frequency band 1		Frequency band 2	
Parameters	Track01 (80 shots)	Track02 (103 shots)	Track01 (80 shots)	Track02 (103 shots)
SD (m)	44.7±1.4	23.6±1.6	44.0±3.0	23.6±1.4
WD (m)	72.5±1.0	72.0±0.9	72.0±0.8	71.2±0.8
c_{p1} (m/s)	1719.0±13.9	1705.3±11.3	1686.6±22.9	1680.9±27.5
α_{p1} (dB/m/kHz)	0.076±0.034	0.055±0.037	0.194±0.080	0.150±0.093
ρ_1 (g/cm^3)	1.89±0.20	1.98±0.18	2.10±0.20	2.08±0.45
H (m)	63.6±12.6	64.1±9.0	23.5±8.3	21.8±6.8
c_{pb} (m/s)	2136±165	2032±182	2161±172	2168±186
α_{pb} (dB/m/kHz)	0.334±0.251	0.321±0.299	0.356±0.239	0.399±0.234
ρ_b (g/cm^3)	2.51±0.29	2.56±0.326	2.56±0.24	2.61±0.26
c_{s1} (m/s)	147.3±105.7	126.2±98.6	100.6±101.6	222.3±139.7
c_{sb} (m/s)	473.1±213.0	447.9±254.4	598.7±458.8	1051.7±500.9
α_{s1} (dB/m/kHz)	0.597±0.461	0.525±0.471	0.763±0.473	0.744±0.504
α_{sb} (dB/m/kHz)	1.065±0.665	1.077±0.683	1.043±0.592	0.670±0.581

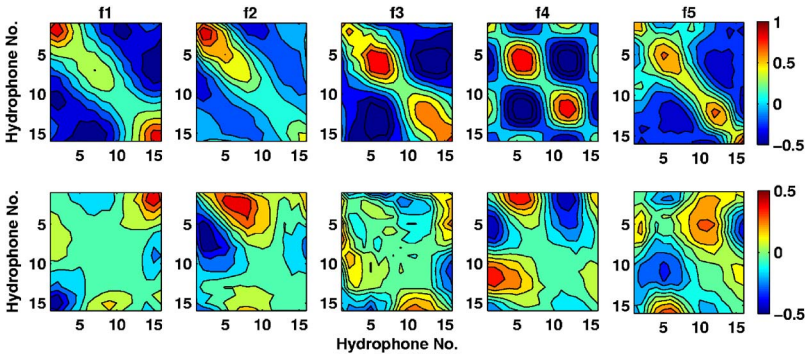


FIG. 11. (Color online) Example of estimated full data error covariance matrices at five frequencies in the lower frequency band. The data are from Track01 and the number of shots used was 20. The upper row is the real parts and the lower row shows the imaginary parts of the matrices.

generate smaller uncertainties (i.e., narrower one-dimensional marginal distributions). Instead, the issue is that the use of appropriate information about the data errors will generate more reliable (and likely more accurate) data uncertainty estimates, and for highly spatially correlated data errors, our results show that properly estimating the off-diagonal elements of covariance matrix is crucial to the uncertainty estimates of the parameters.

As discussed previously, relevant statistical tests should always be used to test the assumptions that are made about the data error residuals in deriving the error covariances. Here, KS and runs tests were applied to the standardized residuals weighted by the three different types of covariance information. There were ten standardized residuals series (five frequencies, each has real and imaginary parts) in each case to be examined. For the full covariance matrices case none of the tests rejected the null hypothesis that the standardized residuals were Gaussian distributed and spatially uncorrelated, at the significance level of 0.05. However, all of the series rejected the null hypothesis with very strong evidence for the diagonalized full covariance matrix and identical variance matrix cases.

4. Bayesian inversion results

In this section, we present the Bayesian inversion estimates for each of the two frequency bands for one shot from track01 and one from track02. Based on the results of the optimization inversions, shear wave parameters were not included. The marginal one-dimensional distributions for the most sensitive parameters are shown in Fig. 14, and all the estimated values are summarized in Tables VII(a) and VII(b), in terms of MAP estimates and their 95% highest probability density interval (HPD). The results shown were obtained using the full data covariance matrices.

The estimates of sediment p-wave sound speed and attenuation in each frequency band are consistent for the two source depths. The attenuation is frequency dependent, and the estimates from both of the frequency bands approximately fit in the simple relationship of p-wave attenuation and the frequency $\alpha_p = 0.80 f^{1.84}$, where the unit of frequency f is kHz and α_p is in dB/m. This result is different from Zhou's model only in the value of the constant factor.[40] The p-wave sound speed is slightly higher at the lower frequency band. This is most likely due to the deeper penetration depth of the lower frequency sound wave, and consequently samples the higher sound speeds in the deeper layers. Overall, the estimates from Bayesian inversion are consistent with the results from optimization approach presented in Sec. IV A.

5. Inter parameter correlations

The Bayesian inversion provides valuable information about the interparameter correlations. Interparameter correlations were quantified by normalizing the model covariance matrix C^M defined by Eq. (6), to give the model correlation matrix, R^M:

$$R_{ij}^M = C_{ij}^M / \sqrt{C_{ii}^M C_{jj}^M}, \tag{16}$$

where $i, j = 1, \ldots, N_M$, and N_M is the number of parameters to be inverted. Figure 15 shows the correlation matrix for the parameters that were inverted for the data from track01 at the higher frequency band (Fig. 14, case b). Similar correlations between the parameters can be found for the other three cases shown in Fig. 14. The main features are as follows:

(1) Strong positive correlation between the water depth and the source range, along with notable positive correlation between the water depth and source depth. These correlations indicate the effect of water depth mismatch on

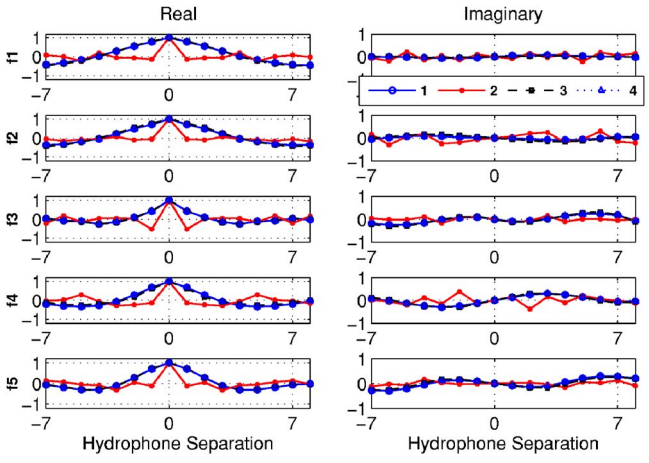


FIG. 12. (Color online) Autocorrelation of original residuals (1, open circles solid line) and the standardized residuals weighted by full covariance matrices (2, closed diamonds solid line), diagonalized full covariance matrices (3, open squares dashed line), and identical variance diagonal matrices (4, closed triangles dotted line) at the five different frequencies. Note that curves 1, 3, and 4 are superimposed.

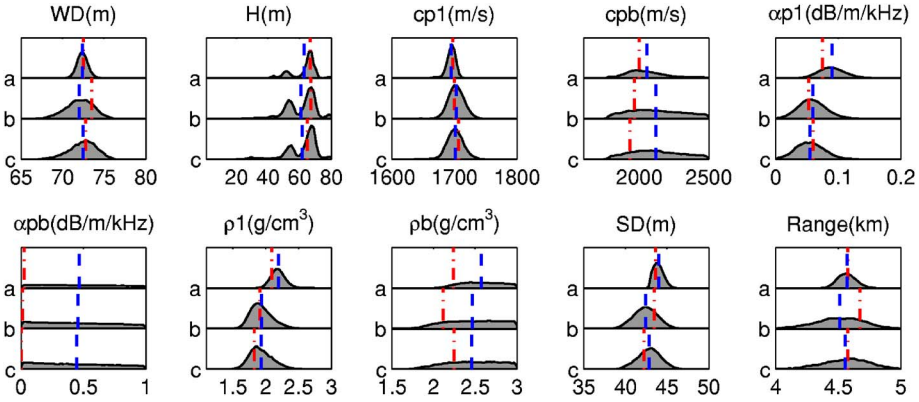


FIG. 13. (Color online) Comparison of one-dimensional marginal distribution between the Bayesian inversions using different data error covariance matrices. The dash-dotted lines indicate MAP estimates; the dashed lines indicate the mean values of the estimates. (a) Full data covariance matrices, (b) diagonalized full data covariance matrices, and (c) identical variance diagonal matrices.

source localization, and are consistent with previous studies by D'Spain *et al.*,[41] Harrison and Siderius,[42] and Shang.[43]

(2) Positive correlation is evident between the sediment depth and half space sound speed, and negative correlations between the sound speed and the density of the lower half space, and between the attenuations of the sediment and the lower half space. Although the correlations are evident, the sensitivity to all these parameters is very weak and there are no strong conclusions about physical relationships. The behavior suggests that the inversion creates convenient relationships to prevent reflection from the interface between the sediment and the lower half space.

(3) Notable positive correlation between the sediment sound speed and attenuation. This relationship suggests that the inversion tries to balance the effects of these parameters in order to match the sound field in the water. Increasing the sediment sound speed increases the impedance at the sea bottom interface, increasing the reflectivity; this increase can be balanced in the inversion by increasing the attenuation. The weak negative correlation between sediment sound speed and density can be understood in terms of the acoustic impedance at the sea floor. The inversion creates a balance between these two parameters in order to control the amount of energy returned to the water.

The information in the parameters' correlations can also be displayed quantitatively by joint marginal PPDs. The joint marginal distributions of selected pairs of sensitive parameters are shown in Fig. 16, from which we can get the information of the level of the correlation between the parameter pairs as well as the numerical ranges of the uncertainties.

6. Discussion of the estimated geoacoustic profile

The long-range experimental geometry supports multiple bottom interacting paths that propagate at low angles ($<10°$) for both source depths. The estimated geoacoustic profile is consistent with the expected behavior for a shallow water environment bounded by a high speed interface at the sea floor in which the sound field is confined within a few meters of the interface. The critical angle for the estimated sediment sound speed is around 28°, so the bottom interacting paths are critically reflected. Sensitivity to the model parameters is consequently greatest for the sea floor parameters, because the sound field is evanescent in the bottom. The estimated values of sediment sound speed and density for both frequency bands are very close to the weighted averages of these parameters from the high resolution geoacoustic model, averaged over the first 20 m bsf.

The most significant features of the underlying sediment structure that affect the results of this inversion are the high

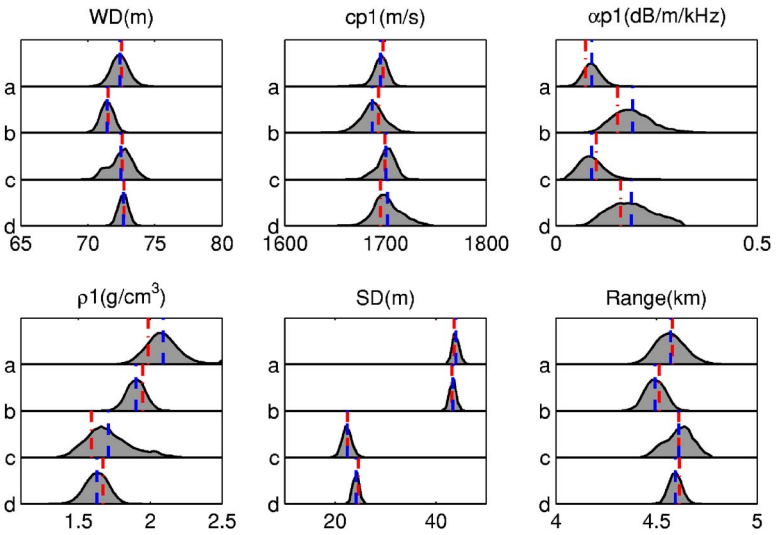


FIG. 14. (Color online) Comparison of selected Bayesian inversion results from two data sets (Track 01 and Track 02). Dash-dotted lines indicate MAP estimates; dashed lines indicate the mean values of the estimates. (a) SD=45 m, frequency band 1; (b) SD=45 m, frequency band 2; (c) SD=25 m, frequency band 1; and (d) SD=25 m, frequency band 2.

TABLE VII. Summary of MAP estimates and 95% HPD interval from Bayesian inversion [left bound MAP right bound], excluding shear wave effects. (a) Lower frequency band results and (b) higher frequency band results.

(a) Lower frequency band results

Parameters	Frequency band 1					
	Track01			Track02		
SD (m)	[42.9	43.6	45.0]	[20.5	22.5	24.2]
WD (m)	[71.2	72.5	73.7]	[70.8	72.5	73.9]
Range (km)	[4.43	4.58	4.72]	[4.480	4.613	4.739]
c_{p1} (m/s)	[1680.6	1697.4	1708.2]	[1680.7	1699.4	1716.4]
α_{p1} (dB/m/kHz)	[0.049	0.074	0.131]	[0.029	0.088	0.147]
ρ_1 (g/cm^3)	[1.846	1.986	2.325]	[1.444	1.593	2.053]
H (m)	[44.3	66.6	71.5]	[38.7	38.8	79.1]
c_{pb} (m/s)	[1777.4	1997.1	2370.2]	[1875.2	2108.8	2461.7]
α_{pb} (dB/m/kHz)	[0.000	0.023	0.938]	[0.000	0.012	0.929]
ρ_b (g/cm^3)	[2.188	2.240	2.998]	[1.652	2.863	2.999]

(b) Higher frequency band results

Parameters	Frequency Band 2					
	Track01			Track02		
SD (m)	[42.2	43.2	44.4]	[23.1	24.6	25.1]
WD (m)	[70.5	71.5	72.4]	[71.9	72.7	73.3]
Range (km)	[4.40	4.51	4.59]	[4.533	4.614	4.653]
c_{p1} (m/s)	[1661.8	1693.2	1712.9]	[1675.7	1695.1	1732.4]
α_{p1} (dB/m/kHz)	[0.103	0.153	0.296]	[0.090	0.160	0.295]
ρ_1 (g/cm^3)	[1.773	1.947	2.027]	[1.477	1.674	1.784]
H (m)	[16.0	23.5	79.9]	[11.2	26.0	80.0]
c_{pb} (m/s)	[1686.4	2182.4	2455.5]	[1727.2	1937.3	2480.9]
α_{pb} (dB/m/kHz)	[0.055	0.691	0.999]	[0.040	0.150	0.983]
ρ_b (g/cm^3)	[1.956	2.667	2.999]	[1.603	2.291	2.995]

speed transition layer between medium or fine sand and more consolidated clay within a meter of the sea floor, and the deeper R reflector. Most of the ground truth information for this region is based on the *in situ* probe data from the high speed transition layer of medium to fine sand. Since this layer is much less than the sound wavelengths, the low frequency averaged estimates from this inversion strongly depend on the depth of the R reflector that separates the lower speed semiconsolidated clay layer with a significantly faster layer with sound speeds of ~1900 m/s. The inversion for the higher frequency band data suggests that the depth of this reflector is around 20 m. Although the sensitivity to the subsea-floor parameters is very weak, it is most likely that the depth is not less than 20 m over most of the track.

The estimated value of sediment attenuation at the lower frequency band (~0.08 dB/m/kHz) is about half the value estimated at the higher frequency band (~0.16 dB/m/kHz). Although this result translates to a nonlinear frequency dependence for attenuation at these low frequencies, the result should be interpreted in terms of a range and depth average of losses due to many effects in addition to intrinsic attenuation. The depth average over the first 20 m bsf includes an average over a layered structure consisting of several different materials with different particle sizes, porosities, and permeabilities. Since the attenuation generally decreases with depth in the sediment, the result for the higher frequency band may also reflect the changes due to shallower penetration below the sea floor. Overall, the estimated values are consistent with the values in the high resolution model, and are generally higher than the values compiled for sea floor sediments by Zhou.[40]

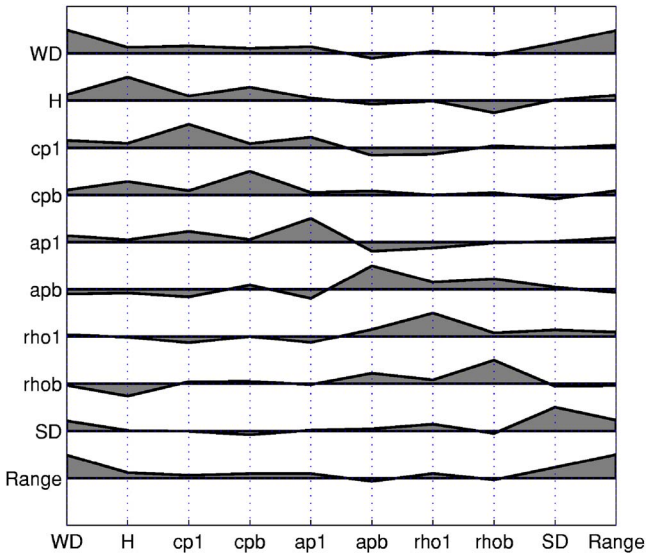


FIG. 15. (Color online) Interparameter correlation matrix for data from Track01 at the higher frequency band.

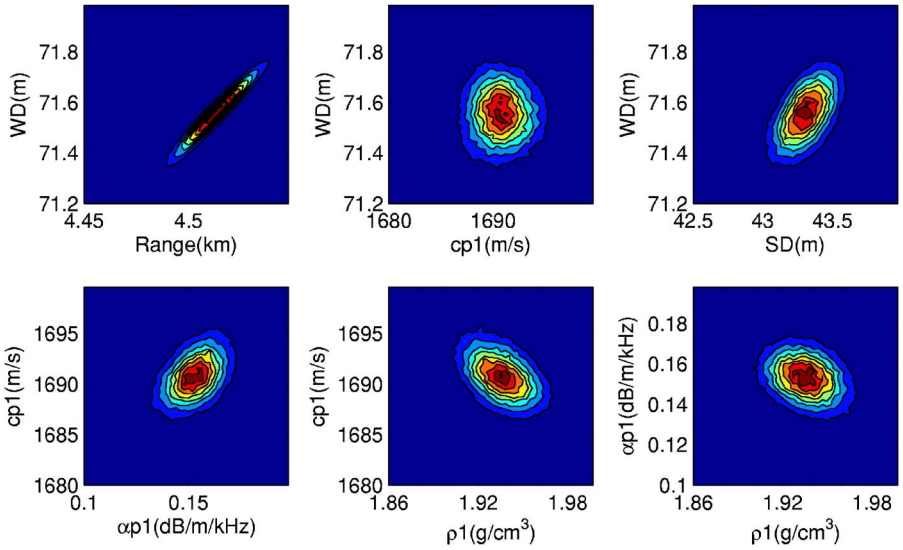


FIG. 16. (Color online) Selected joint marginal distributions of the Bayesian inversion result for data from Track01 at the higher frequency band.

The slightly lower value of sediment sound speed for the higher frequency band may also be due to the shallow penetration depth, and interaction with lower sound speed layers just below the sea floor. From Fig. 6, the sound speed decreases between 2 and 12 m, beneath the high speed layer at the sea floor.

V. SUMMARY

This paper applied matched field optimal and Bayesian matched field inversion techniques to air gun data from the SWARM'95 experiment to estimate an equivalent geoacoustic model for the site. The data were collected by a vertical hydrophone array at long ranges from the air gun. The geoacoustic model consisted of a two-layer approximation to a multilayer high resolution model derived from extensive ground truth information. The inversions used multiple frequencies that were selected from the air gun shot spectrum on the basis of high SNR and high spatial coherence across the array. The frequencies spanned two low frequency bands centered at 65 and 150 Hz.

The effect of correlated errors on the data was examined by comparing the Bayesian inversion results for three different approaches to account for the spatially correlated errors. A straightforward method using an average of temporally coherent spectral components from an ensemble of air gun shots was adopted for calculating the full covariance matrix. Statistical tests showed that there was no contrary evidence to the underlying assumption that the standardized data residuals weighted by the full data error covariance matrix for the spectral components were temporally stationary and spatially uncorrelated and Gaussian distributed.

The most sensitive geoacoustic parameters were the *p*-wave sound speed, attenuation, and density at the sea floor. The estimated values of sediment sound speed and density were consistent with the weighted average of the high resolution model over the first 20 m bsf. The long-range experiment was also effective for estimating sediment attenuation in this environment. The attenuation dependence on frequency in the two low frequency bands is nonlinear. However, the frequency dependence is interpreted partly as a depth average over the different types of sediment material in the layered structure of the ocean bottom. Overall, the estimated model is consistent with the results expected for a shallow water environment bounded by a high speed interface at the sea bottom in which the sound field in the bottom is constrained within a wavelength of the sea floor.

ACKNOWLEDGMENTS

This work is supported by Office of Naval Research. The authors would like to thank Dr. A. Song from University of Delaware for making the data available for further analysis. Y.-M. J. appreciates the helpful discussions on Bayesian inversion theory with Dr. S. E. Dosso, Dr. M.J. Wilmut, and Dr. J. Dettmer at the University of Victoria.

[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]S. E. Dosso, M. L. Yeremey, J. M. Ozard, and N. R. Chapman, "Estimation of ocean-bottom properties by matched-field inversion of acoustic field data," IEEE J. Ocean. Eng. **18**, 232–239 (1993).
[3]P. Gerstoft, "Inversion of seismo-acoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[4]P. Gerstoft, "Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach," J. Acoust. Soc. Am. **97**, 2181–2190 (1995).
[5]M. I. Taroudakis and M. G. Markaki, "On the use of matched field processing and hybrid algorithm for vertical slice tomography," J. Acoust. Soc. Am. **102**, 885–895 (1997).
[6]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).
[7]M. R. Fallat and S. E. Dosso, "Geoacoustic inversion via local, global and hybrid algorithms," J. Acoust. Soc. Am. **105**, 3219–3230 (1999).
[8]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).
[9]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).
[10]M. Siderius, P. Gerstoft, and P. Nielsen, "Broadband acoustic inversion from sparse data using genetic algorithms," J. Comput. Acoust. **6**, 117–134 (1998).
[11]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).

[12]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).
[13]A. Tolstoy, "Using matched-field processing to estimate shallow water bottom properties from shot data taken in the Mediterranean Sea," IEEE J. Ocean. Eng. **21**, 471–479 (1996).
[14]Z. H. Michalopoulou, "Matched impulse response processing for shallow water localization and geoacoustic inversion," J. Acoust. Soc. Am. **108**, 2082–2090 (2000).
[15]D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).
[16]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).
[17]Z. H. Michalopoulou and U. Ghosh-Dastidar, "Tabu for matched-field source localization and geoacoustic inversion," J. Acoust. Soc. Am. **115**, 135–145 (2004).
[18]C. W. Holland, J. Dettmer, and S. E. Dosso, "Remote sensing of density and velocity gradients in the transition layer," J. Acoust. Soc. Am. **118**, 163–177 (2005).
[19]P. Gerstoft and Mecklenbräuker, "Ocean acoustic inversion with estimation of *a posteriori* probability distributions," J. Acoust. Soc. Am. **104**, 808–819 (1998).
[20]S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).
[21]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).
[22]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Uncertainty analysis in matched-field geoacoustic inversions," J. Acoust. Soc. Am. **119**, 197–207 (2006).
[23]M. K. Sen and P. L. Stoffa, "Bayesian interference, Gibbs; sampler and uncertainty estimation in geophysical inversion," Geophys. Prospect. **44**, 313–350 (1996).
[24]S. E. Dosso, P. Nielsen, and M. J. Wilmut, "Data error covariance in matched-field geoacoustic inversion," J. Acoust. Soc. Am. **119**, 208–219 (2006).
[25]D. Tollefsen, S. E. Dosso, and M. J. Wilmut, "Matched-field geoacoustic inversion with a horizontal array and low-level source," J. Acoust. Soc. Am. **120**, 221–230 (2006).
[26]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Data error covariance matrix for vertical array data in an ocean waveguide," J. Acoust. Soc. Am. **119**, 3272 (2006).
[27]J. R. Apel *et al.*, "An overview of the 1995 SWARM shallow-water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500 (1997).
[28]A. C. Kibblewhite, "Attenuation of sound in marine sediments: A review with emphasis on new low-frequency data," J. Acoust. Soc. Am. **86**, 716–738 (1989).
[29]http://www.ngdc.noaa.gov/mgg/gdas/gd_designagrid.html
[30]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).
[31]B. J. Kraft, L. A. Mayer, P. Simpkin, P. Lavoie, E. Jabs, and J. A. Goff, "Calculation of in situ acoustic wave properties in marine sediments" *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer Academic, Dordrecht, 2002), pp. 123–130.
[32]L. A. Mayer, B. J. Kraft, P. Simpkin, P. Lavoie, E. Jabs, and E. Lynskey, "In-Situ determination of the variability of seafloor acoustic properties: An example from the ONR geoclutter area," in Ref. 31, pp. 115–122.
[33]A. Turgut, D. Lavoie, D. J. Walter, and W. B. Sawyer, "Measurements of bottom variability during SWAT New Jersey shelf experiments," in Ref. 31, pp. 91–98.
[34]M. V. Trevorrow and T. Yamamoto, "Summary of marine sedimentary shear modules and acoustic speed profile results using a gravity wave inversion technique," J. Acoust. Soc. Am. **90**, 441–455 (1991).
[35]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P.S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Pet. Geol. **209**, 147–172 (2004).
[36]W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey continental shelf," IEEE J. Ocean. Eng. **20**, 321–336 (1995).
[37]M. D. Richardson and K. B. Briggs, "Relationships among sediment physical and acoustic properties in siliciclastic and calcareous sediment," ECUA2004, Delft, Vol.**2**, pp. 659–664.
[38]W. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, 1992), pp. 617–622.
[39]C. T. Tindle and Z. Y. Zhang, "An equivalent fluid approximation for a low shear speed ocean bottom," J. Acoust. Soc. Am. **91**, 3248–3256 (1992).
[40]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**, 2494 (2005).
[41]G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265 (1999).
[42]C. H. Harrison and M. Siderius, "Effective parameters for matched field geoacoustic inversion in range-dependent environment," Indian J. Chem. **28**, 432–445 (2003).
[43]E. C. Shang and Y. Y. Wang, "Environmental mismatching effects on source localization processing in mode space," J. Acoust. Soc. Am. **89**, 2285–2290 (1991).

Effects of ocean thermocline variability on noncoherent underwater acoustic communications

Martin Siderius,[a] Michael B. Porter, Paul Hursky, Vincent McDonald, and the KauaiEx Group
HLS Research Corporation, 12730 High Bluff Drive, San Diego, California 92130

(Received 16 December 2005; revised 14 December 2006; accepted 30 December 2006)

The performance of acoustic modems in the ocean is strongly affected by the ocean environment. A storm can drive up the ambient noise levels, eliminate a thermocline by wind mixing, and whip up violent waves and thereby break up the acoustic mirror formed by the ocean surface. The combined effects of these and other processes on modem performance are not well understood. The authors have been conducting experiments to study these environmental effects on various modulation schemes. Here the focus is on the role of the thermocline on a widely used modulation scheme (frequency-shift keying). Using data from a recent experiment conducted in 100-m-deep water off the coast of Kauai, HI, frequency-shift-key modulation performance is shown to be strongly affected by diurnal cycles in the thermocline. There is dramatic variation in performance (measured by bit error rates) between receivers in the surface duct and receivers in the thermocline. To interpret the performance variations in a quantitative way, a precise metric is introduced based on a signal-to-interference-noise ratio that encompasses both the ambient noise and intersymbol interference. Further, it will be shown that differences in the fading statistics for receivers in and out of the thermocline explain the differences in modem performance. [DOI: 10.1121/1.2436630]

PACS number(s): 43.30.Re, 43.30.Zk, 43.30.Cq, 43.60.Dh [DRD] Pages: 1895–1908

I. INTRODUCTION

It is generally accepted that the ocean environment impacts acoustic communication performance. What is largely unknown is which factors have the strongest impact on performance (e.g., noise, sea state, mixed layer depth, etc.), keeping in mind that they may be signaling method dependent. Also not well known is how to best adapt communication parameters to minimize the impact of these factors. Unlike line-of-sight communications, the shallow-water acoustic ocean channel often gives rise to significant multipath energy caused by reflected and refracted propagation paths between the source and receiver. These time-dispersed multipaths cause replicas of previously transmitted symbols to interfere in the detection of the current symbol and consequently, if strong enough, will result in bit errors.

There are two purposes of this research: (1) to determine, through measurements and modeling, the impact of source/receiver geometry and various environmental factors on shallow-water communications performance and (2) to demonstrate that modeling, with sufficient environmental information, can be used for precise, quantitative performance predictions.

A channel simulator with an embedded high-fidelity acoustic model is used to reproduce both the multipath structure and ultimately the communications performance. Accurate modeling allows the results to be generalized to other sites and environmental conditions, as well as the determination of optimal source/receiver placement. Determining the environmental impact on performance is important for predicting when and where underwater communication systems suffer degradation and to what extent. A high-fidelity channel simulator allows for virtual experiments in any desired environment or configuration. For example, suppose a communication link is desired between two underwater vehicles deployed in an area that had previously shown good communications performance for source and receivers moored near the seabed. Can it be assumed that there will be good performance even though the vehicles are moving and operating at different depths and possibly in a different season?

Since many environmental factors that influence performance can either be measured *in situ* or obtained through archival data, as oceanographic and acoustic models improve, so will acoustic communication system performance prediction and enhancement (e.g., recommending preferred source/receiver operating depths for communications). Cognizance of environmental factors that cause communication system degradation will influence how, when, and/or where a system is deployed.

A study of environmental factors that impact communication performance requires experimentation with careful measurement of channel properties (e.g., ocean sound-speed structure, surface roughness, and currents) while simultaneously transmitting communication signals. Further, these experiments must be conducted over a statistically significant time, and measurements must be designed to isolate the environmental parameters of interest. The Kauai Experiment (KauaiEx) was designed exactly for this purpose and took place off the northwest coast of Kauai, HI in June and July of 2003.[1] The experiment was designed to measure the environ-

[a]Electronic mail: siderius@hlsresearch.com

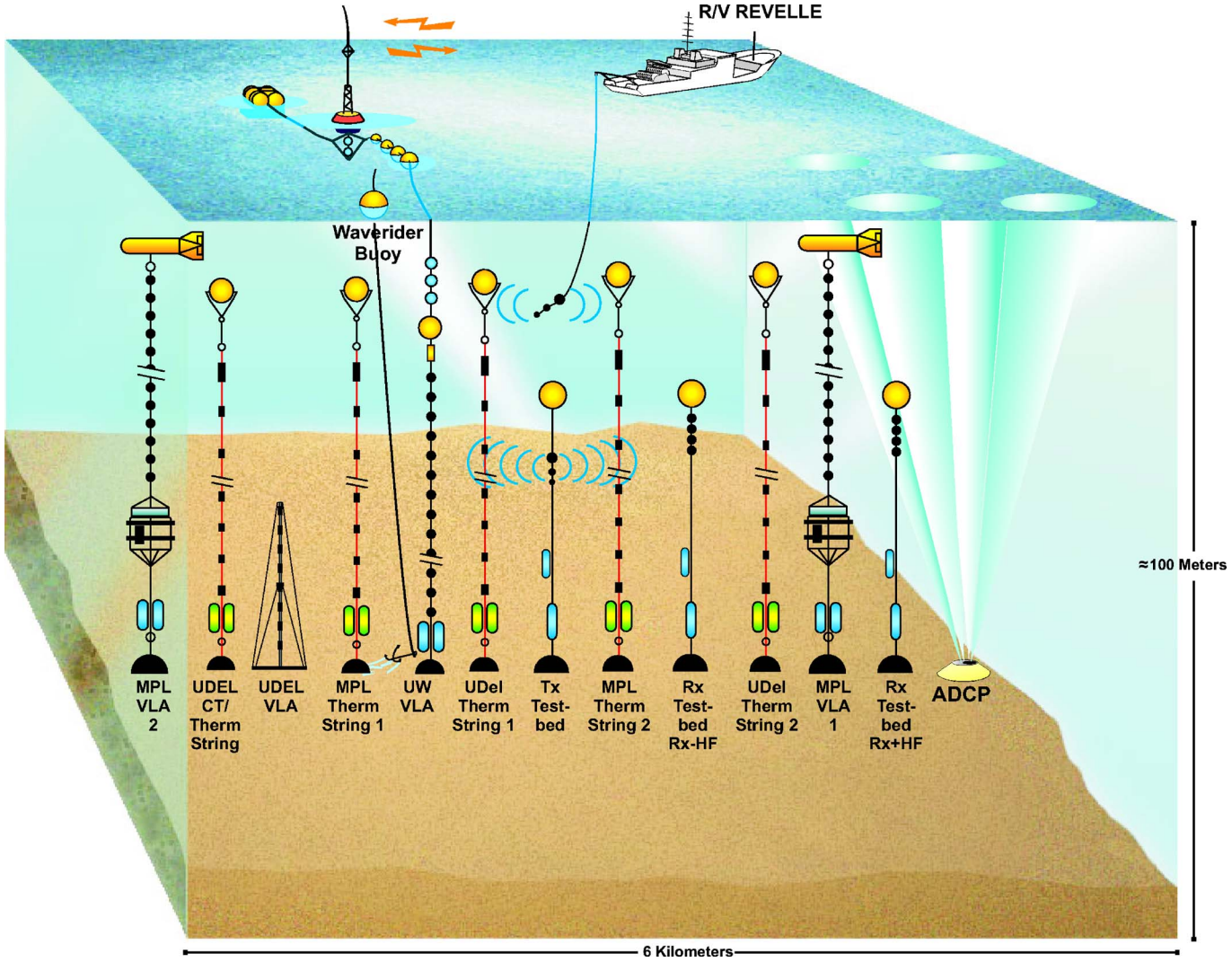


FIG. 1. (Color online) Geometry for the second deployment of KauaiEx. Acoustic source is labeled as Tx Testbed and the data analyzed in this paper was received on array MPL VLA2. The towed source transmissions from the *R/V* Revelle were interleaved with the moored source for independent analysis.

ment and simultaneously transmit acoustic communication waveforms over several days. This type of experiment has been of interest in recent years as applications for underwater communications have increased.[2–4]

The KauaiEx experiment site presented dynamic oceanographic conditions, typical of shallow-water littorals. To fully capture these conditions, full water-column measurements of both the acoustics and oceanographic properties were conducted simultaneously. Data were collected over many days to obtain significant statistics. Although there were a variety of phases in KauaiEx, this paper focuses on a fixed source and fixed, vertical, receive-array configuration. This provides a much better isolation of the time-varying channel characteristics without having to unravel performance differences caused by changing bathymetry or Doppler effects, as an example. Further, the full-water column vertical array allows different receiver depths to be analyzed simultaneously using a single transmission source. These types of careful, simultaneous, environmental and communication measurements are also required to validate models that simulate the channel characteristics and predict communication performance.

This paper analyzes performance of communication signals using frequency-shift-key modulation (FSK)[5] that is often implemented in commercial modems because of its robustness and implementation simplicity, especially in the receiver. A general review of various underwater acoustic communications techniques and performance can be found in an overview article by Kilfoyle and Baggeroer.[6] Because of their implementation in undersea networks,[7] FSK modems are of practical interest as well. Coherent methods, such as quadrature phase shift keying offer higher spectral efficiencies than noncoherent (i.e., FSK for one) methods. If point-to-point data throughput is an important design consideration, then coherent techniques should be considered for the highly, band-limited underwater acoustic channel. Bandwidth limitations are determined by absorption and are approximately 1 dB/km for the 8–13 kHz band considered in this paper. The absorption roughly increases with frequency squared. In addition, the high resonant quality of electro-acoustical transduction equipment also limits available bandwidth. However, bandwidth-efficient coherent methods come at the price of processing complexity at the receiver needed to overcome channel variability, and generally require a

higher signal-to-noise ratio (SNR). In addition to being valuable in its own right, the simple and robust nature of FSK signaling makes its performance a useful yardstick for other methods to compare against.

The balance of this paper is organized in the following manner. Section II describes the Kauai experiment and the data used for analysis. The environmental measurements and the transmitted acoustic communications signals are described. Section III presents the measured performance and that expected assuming both fading and nonfading channel models. This section illustrates the communication system performance impact of source/receiver geometry, water-column temperature structure, and wind speed. In Sec. IV a channel simulator with embedded ocean acoustic model is used to replicate and explain measured performance and confirm the channel fading statistics.

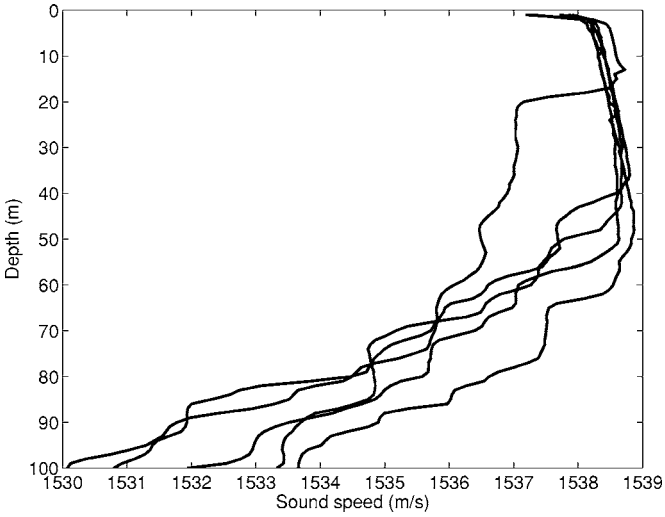


FIG. 2. Measured sound speed profiles taken on 1 July 2003 near the experimental area. Note the change in depth of the mixed layer.

II. THE KAUAI EXPERIMENT

Details of all experiments that comprise KauaiEx are described by Porter *et al.*[1] This paper is based on data collected during the second deployment, 30 June to 3 July 2003. The instruments and their locations are shown in Fig. 1. The towed and moored sources transmitted nonoverlapping acoustic waveforms. Data analyzed here are from the moored Telesonar Testbed only (indicated as Tx Testbed near the middle of the track in Fig. 1) with the source located about 5 m from the seabed. The Telesonar Testbed is a versatile, wideband, acoustic communication research instrument that has been the centerpiece of many acoustic communication experiments.[8] The moored Testbed used a subsurface float to maintain the position of the sound projector. Receptions were recorded on multiple arrays but here the data are analyzed from the MPL-VLA2 (Marine Physical Laboratory) receiver array located 3 km from the source. The vertical line receive array (VLA) was moored and configured with 16 hydrophones spaced 5 m apart with the first hydrophone about 8.5 m from the seabed.

As can be seen from Fig. 1, there were extensive environmental measurements including five strings of either thermistors or CTD (conductivity, salinity, depth) sensors to measure water column properties along the acoustic propagation path between transmitter and receiver. In addition, a waverider buoy measured wave heights, and an acoustic Doppler current profiler (ADCP) measured the volumetric water currents. Other geophysical measurements such as grab samples, seismic profiling, and multibeam mapping were also made to help characterize the seabed.

From an acoustic propagation point of view, the bathymetry and seabed along the acoustic propagation path were fairly benign. The acoustic path was over an area believed to be a submerged beach. The grab sample analysis and subsequent visual observation showed it to be mostly a medium grain coral sand with bits of larger coral mixed in. Away from the track, the bathymetry dropped off to several kilometers depth and near the shore decreased to less than 10 m. However, the bathymetry along the propagation path was almost uniform at 100 m.

In contrast to the simple bathymetry and seabed, the oceanographic conditions were relatively complex and dynamic over the experiment duration. The water sound-speed, vertical profile depicted a region near the surface with a high degree of mixing due to the often windy conditions; however, the depth of this mixed layer typically varied from approximately 10 to 60 m but was sometimes deeper. This can be seen from the CTD casts made on 1 July 2003 shown in Fig. 2. Here, the sound speed shows the general trend to decrease with water depth but the depth where the mixed layer ends and the thermocline begins varied with location and time. In these five CTD casts, the mixed layer depth is between 40 and 50 m for four of the profiles and decreases to about 20 m for one.

In some locations around the world's oceans, the sound speed near the surface is highly variable due surface warming effects; however, at the KauaiEx site, the wind-driven mixing causes the water near the surface to be more uniform with most of the variability occurring at greater depths. These sound speed profiles give a sense of the structure and variability, but the thermistor strings give a time history for a particular location. In Fig. 3 the data from the thermistor string nearest MPL-VLA2 are shown (it is labeled "UDel CT/Therm. String" in Fig. 1 and is about 500 m from the VLA). There were 13 thermistors located at depths between 4 and 82 m. There is a clear, regular pattern evident in the thermistor data and shows the time-dependent, thermocline depth variability. It can be seen that in some cases, the thermocline depth is quite shallow and at other times the water column is much more uniform. The impact of these variations on the acoustic communication signals will be discussed in following sections.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The FSK signals considered here use 128 frequency components spaced 40 Hz apart in the 8–13.2 kHz band. The upper and lower 4 tones are reserved for pilot tones to compensate for Doppler. The information is passed using a subset of the 128 frequencies that are modified every 25 ms.

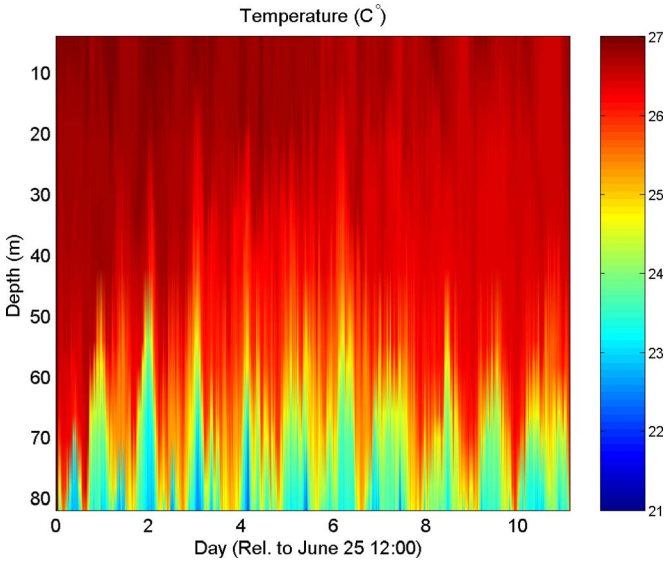


FIG. 3. (Color online) Time history of the ocean temperature during the experiment. This was from the same location on the moored UDel CT/Therm. string located near MPL-VLA2. Sound speed is mostly a function of the water temperature with a slight dependence on salinity.

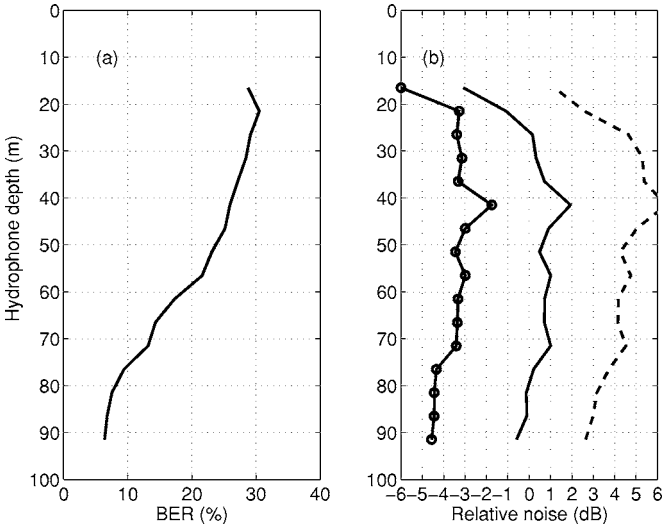


FIG. 4. (a) The depth-dependent percent bit errors averaged over about 1 day of transmissions. (b) The ambient noise as a function of depth averaged over the same time period in three frequency bands: 8–13.2 kHz (solid line), 5–8 kHz (dashed line), and 13.2–16.2 kHz (circles).

The FSK modulation scheme uses 30 instantiations of 1 of 4 coding (i.e., M-ary FSK with $M=4$).[5] This means that 1 of 4 tones activated is used to encode 2 bits of data, i.e., 0-0, 0-1, 1-0, or 1-1. A practical decoding advantage is gained by requiring the receiver to simply determine which of 4 tones is loudest. This method is less problematic than on/off keying where the decoder decides if a tone is a 1 (on) or 0 (off). This requires thresholding which is very sensitive to channel fading.

Thirty blocks of 4 tones are transmitted simultaneously producing 60 bits in 25 ms, or 2400 bits per second (bps). At the receiver, a spectrogram is taken of the FSK payload (i.e., excluding pilot tones and acquisition components of the transmission packet) using a nonoverlapping boxcar window of 25 ms. The strongest tone in each block of 4 tones is then determined. The ocean, of course, acts like an echo chamber producing multipath spread. To combat multipath spread, the tone duration is increased and the energy over the longer interval is accumulated before conjecturing which tone from the group of four was transmitted. This in turn means a data rate loss. For instance, increasing the tone duration to 50 ms (by adding two 25 ms blocks to maintain frequency separation) yields a data rate of 1200 bps.

Another component of the modem design is the acquisition process used for initial symbol alignment. There are many ways to do this with pros and cons for each. For these data a set of *m*-sequences proceeding the data payload were match filtered at the receiver to provide symbol time alignment.

Error correction coding at the transmitter is an effective way to reduce errors in fading channels.[5,9] Although errors can be significantly reduced by coding, this paper considers raw bit errors to reduce the time period over which significant statistics can be developed for adequate analysis. Finally, it should be noted that Doppler effects due to source/receiver motion are typically important for mobile platforms. However, the focus here is on fixed networks so such effects will not be addressed.

A. Measured FSK performance

One of the most notable observations during KauaiEx was the performance improvement with hydrophone depth. This can be seen in Fig. 4(a) where the bit errors as a function of depth are averaged over about 1 day for 2400 bit/s transmissions. The deepest hydrophone at approximately 91.5 m shows an average of about 5% bit errors while the most shallow at roughly 16.5 m averages about 30%. One thought might be the lower bit errors are a result of increased SNR due to a decrease in the ambient noise level with depth (transmit level was held constant). However, this is not the case. Shown in Fig. 4(b) is the ambient noise averaged over the same period for the in-band (8–13.2 kHz), below-band (5–8 kHz), and above-band (13.2–16.2 kHz) frequency regions. The figure is on a relative dB scale and shows the roughly 6 dB decrease in noise as frequency is doubled. Also, there is little evidence that the noise conditions are improved at lower depths; actually, the shallowest hydrophone shows the lowest noise level. Nevertheless, the data show that, in general, the noise field is relatively homogeneous, vertically. It turns out that the improvement in performance with depth is partly due to higher signal level rather than lower ambient noise levels (another factor is the multipath which is analyzed in detail in later sections). The higher signal levels are caused by the thermocline trapping acoustic communication signals near the sea floor; this will be discussed further in Secs. III E and III F.

There is also an interesting difference in the temporal variability of performance at different receiver depths observed over a 24 hour period which captures diurnal oceanographic and wind cycles. Figure 5(a) shows time history of the average bit errors for the four shallowest hydrophones at 16.5, 21.5, 26.5, and 31.5 m for 2400 bit/s rate. In Fig. 5(b)

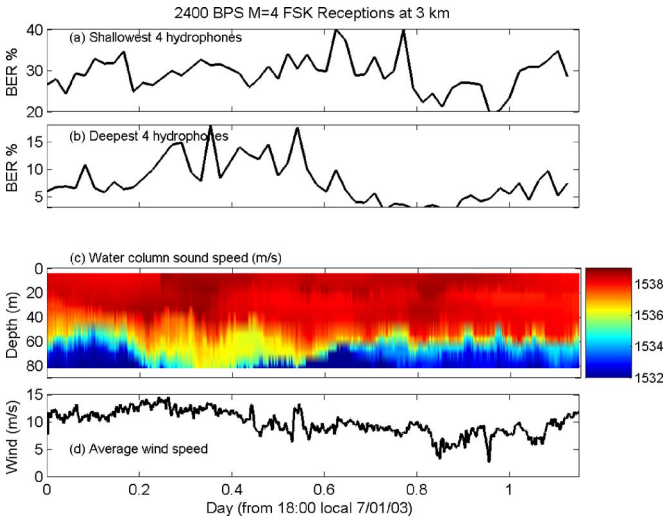


FIG. 5. (Color online) (a) The averaged percent bit errors for the shallowest four hydrophones for 2400 bps transmissions over about 1 day. In (b), average bit errors are shown for the deepest four hydrophones [note the different y axis from (a)]. In (c), the water column sound speed is shown and in (d) the average wind speed during the same period. The x axis is slightly more than 1 day referenced to 18:00 on 1 July 2003 (local time).

the average bit errors for the four deepest hydrophones at 91.5, 86.5, 81.5, and 76.5 m are shown [note the change in the y axis compared to (a)]. The period of the day between about 0.2 and 0.6 shows a marked increase in bit errors for the deepest hydrophones followed by a decrease. Figure 5(c) shows the water column sound speed and Fig. 5(d) shows the wind speed during the same time period. The water sound speed measurement was made by a thermistor string located about 500 m from the receivers.

A few observations can be made from Fig. 5. First, the period corresponding to the bit error increase for the deepest hydrophones corresponds to the period where the water column is more mixed and the thermocline presence is reduced. The link between bit errors and thermocline position is weaker for the shallower hydrophones. Second, there are more errors overall for the shallow hydrophone ($\sim 30\%$) compared to the deep hydrophone ($\sim 5\%$). The resultant temperature-dependent channel conditions account for this phenomenon. The percent bit errors for the deepest hydrophones increase when the water column becomes more mixed due to upper, warmer water moving deeper. This creates a sound-speed profile without a thermocline and therefore without the mechanism to trap acoustic signals. This is more evident when looking at modeling results in Sec. IV and at the impulse response measurements described in Sec. III F.

Wind speed is fairly constant during the 0.2–0.6 period; however, in the 0.8–1.0 time period the wind speed decreases below 5 m/s with a corresponding decrease in bit errors for both the shallow and deep hydrophones. The winds obviously affect both the ambient noise and the surface scatter loss. It is reasonable to wonder whether surface scatter will attenuate the steeper angle paths and thereby reduce multipath leading to a corresponding reduction in inter symbol interference (ISI). One might then also wonder if better modem performance can be expected with surface scatter. However, surface scatter is a complicated subject. Surface losses can be due to (1) a static surface that scatters energy in different directions, ultimately leading to higher losses through bottom absorption and (2) a perfectly flat but dynamic surface that distorts the FSK tones and produces a weaker coherent signal. The effects of the latter depend also on the tone duration since a short tone effectively freezes the surface. In fact, both these mechanisms are in play; however, in acoustic modeling these distinct mechanisms are often treated vaguely as one. Although not conclusive, the measurements here do not appear to support an argument that increased roughness improves modem performance. Actually, the performance improves during periods with *calmer* seas and *lower* wind speed. However, in these data the variations in the sound speed profile seem to dominate.

B. Predicted FSK performance

The theory of FSK performance for a nonfading channel in the presence of additive noise has been well developed.[5] As background to the discussions here, a review of those results is presented with the notation closely following the derivation presented in Proakis.[5] This analysis generally holds for FSK transmissions but concentrates specifically on the $M=4$ FSK that was used during KauaiEx.

1. Nonfading channel model

The signal sinusoidal tones are "on" for the bit duration T_b and are expressed as

$$s(t) = \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos(2\pi ft + 2\pi(m-1)\Delta ft), \quad m = 1,2,3,4, \tag{1}$$

where f is the frequency for $m=1$ and $f+(m-1)\Delta f$ for the neighboring tones. The amplitude is expressed in terms of the energy per bit $\mathcal{E}_b$. In a nonfading channel, the received signal is

$$r(t) = \sqrt{\frac{2\mathcal{E}_b}{T_b}} \cos(2\pi ft + 2\pi(m-1)\Delta ft + \phi) + n(t), \quad m = 1,2,3,4, \tag{2}$$

where ϕ is the phase shift due to the transmission delay and $n(t)$ is additive white Gaussian noise. To obtain the amplitude of the received signal regardless of the phase, $r(t)$ is correlated with the quadrature carriers, $\sqrt{1/T_b}\cos(2\pi ft + 2\pi(m-1)\Delta ft)$ and $\sqrt{1/T_b}\sin(2\pi ft + 2\pi(m-1)\Delta ft)$. The detector selects the largest tone by computing the envelope of the correlations, or

$$r_m = \sqrt{r_{mc}^2 + r_{ms}^2}, \tag{3}$$

where r_{mc} and r_{ms} are the correlation outputs from the cosine and sine components for the mth tone.

For the nonfading channel the on tone components will be denoted as $m=1$ and are simply

$$r_{1c} = \sqrt{\mathcal{E}_b}\cos\phi_1 + n_{1c} \tag{4}$$

and

$$r_{1s} = \sqrt{\mathcal{E}_b} \sin \phi_1 + n_{1s}. \tag{5}$$

The "off" tones ($m=2,3,4$) have: $r_{mc}=n_{mc}$ and $r_{ms}=n_{ms}$ with the noise components n being mutually, statistically independent, zero-mean Gaussian variables with equal variance $\sigma^2 = N_0/2$. For notational convenience, the random variable R_m is defined as

$$R_m = \frac{\sqrt{r_{mc}^2 + r_{ms}^2}}{\sigma}. \tag{6}$$

Next, consider the probability distributions of the amplitudes for the on and off tones. The on tone amplitude probability distribution is Ricean and written as

$$p_{R_1}(R_1) = R_1 \exp\left[-\frac{1}{2}\left(R_1^2 + \frac{4\mathcal{E}_b}{N_0} \right) \right] I_0\left(\sqrt{\frac{4\mathcal{E}_b}{N_0}} R_1 \right), \tag{7}$$

where I_0 is the zeroth-order modified Bessel function of the first kind. For the $m=2,3,4$ terms the probability distributions are Rayleigh,

$$p_{R_m}(R_m) = R_m \exp\left(-\tfrac{1}{2} R_m^2 \right). \tag{8}$$

A correct decision will be made if $R_1 > R_m$ or

$$P_c = \int_0^\infty [P(R_1 > R_m | R_1 = x)]^3 p_{R_1}(x) dx, \tag{9}$$

where

$$P(R_1 > R_m | R_1 = x) = \int_0^x p_{R_m}(r_m) dr_m = 1 - e^{-x^2/2}, \tag{10}$$

and the power of 3 arises from the fact that for $m=2,3,4$ the random variables are statistically independent and identically distributed so the joint probability factors into a product,

$$P_c = \int_0^\infty (1 - e^{-x^2/2})^3 p_{R_1}(x) dx. \tag{11}$$

The general solution to Eq. (11) is given in Proakis[5] and for $M=4$ is

$$P_c = \sum_{n=0}^{3} (-1)^n \frac{3!}{n!(3-n)!(n+1)} \exp\left[\frac{-2n\mathcal{E}_b}{(n+1)N_0} \right]. \tag{12}$$

Finally, the probability of a bit error is

$$P_{\mathrm{BE}} = \tfrac{2}{3}(1 - P_c). \tag{13}$$

The factor of $\frac{2}{3}$ provides the additional reduction in errors that accounts for only using one out of four tones to convey 2 bits.

2. Fading channel model

A fading channel model can cause an average decrease in received signal-to-noise ratio (over that of the nonfading channel) which leads to higher bit errors. In addition, the fading causes the amplitudes of the on and off bits to vary in such a way that additional bit errors are produced. In the presence of multipath, ISI causes a bleed which may make the off bits appear to be on. In other words, the multipath causes the signal to appear in adjacent time bins which is effectively another noise mechanism. Further, both the noise and signal amplitudes vary due to the multipath interference. Following Proakis,[5] for the slowly fading channel, the received signal for the on bits is attenuated by a factor of α, that is, $r(t)=\alpha s(t)+n$. The energy per bit to noise is effectively $\gamma_b = \alpha^2 \mathcal{E}_b / N_0$. The probability of an error can be predicted by modifying Eq. (13) to incorporate the new SNR. But, for the fading channel, α is random so the previous estimate for the probability of an error needs to be averaged over the probability density function of $\gamma_b, p_{\gamma_b}(\gamma_b)$. Thus, to estimate the number of bit errors for the fading channel, P_{BEF}, the error probability for the nonfading channel [given by Eq. (13)] is averaged over $p_{\gamma_b}(\gamma_b)$,

$$P_{\mathrm{BEF}} = \int_0^\infty P_{\mathrm{BE}}(\gamma_b) p_{\gamma_b}(\gamma_b) d\gamma_b. \tag{14}$$

For Rayleigh fading, α is Rayleigh distributed. This means α^2 has a chi-square distribution and so too does γ_b. Thus,

$$p_{\gamma_b}(\gamma_b) = \frac{1}{\bar{\gamma}_b} e^{(-\gamma_b/\bar{\gamma}_b)}, \tag{15}$$

where $\bar{\gamma}_b$ is the average energy per bit to noise ratio,

$$\bar{\gamma}_b = \frac{\mathcal{E}_b}{N_0} E(\alpha^2), \tag{16}$$

with $E(\alpha^2)$ being the average of α^2.

This discussion clearly assumes that the multipath structure is sufficient to generate Rayleigh fading. Another fading model uses the Nakagami-m distribution.[5] This is useful since it allows for a family of bit error probabilities that have fading better than and worse than Rayleigh. Nakagami-m fading has a fading figure parameter m that can be less than 1 for fading situations worse than Rayleigh, equal to one for exactly Rayleigh, and greater than 1 when fading is more favorable than Rayleigh.[5] The data analysis will show that the Rayleigh fading model is good much of the time, but not universally.

C. Measured signal-to-noise ratio

The probability of bit errors, for both fading and nonfading channels is, in part, dependent on the actual or effective SNR (or energy-per-bit to noise ratio). In underwater acoustics, SNR has historically been calculated using the sonar equation,[10]

$$SNR = SL - TL - N, \tag{17}$$

where SL, TL, and N are the source level, transmission loss, and noise levels measured in decibels. This approach works well for sonar applications where signals are often integrated for periods much longer than the duration of multipath. However, for computing SNR levels and consequently predicting communication system performance, the sonar equa-

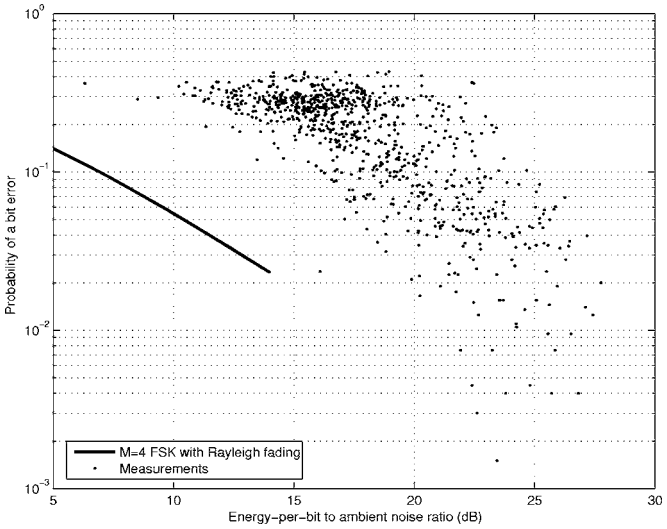


FIG. 6. Measured percent bit errors as a function of the energy-per-bit to ambient noise in decibels (black dots). Also shown is the expected bit error probability for Rayleigh fading with $M=4$ FSK (solid line).

tion approach can be very misleading. To illustrate, the KauaiEx data SNR was calculated this way using the following procedure. The background noise N was measured just below and above the transmission band (8–13.2 kHz). To obtain an in-band noise estimate, these bracketing, but out-of-band levels were interpolated on a log scale. The interpolation on a log scale is used since ambient noise in this frequency band is approximately linear on a log scale[10] [refer to Fig. 4(b) showing measured ambient noise below, in, and above the transmission band]. The signal-plus-noise ($S+N$) was measured in the band and SNR calculated as

$$10 \log\left\{\frac{(S+N)-N}{N}\right\}. \tag{18}$$

The apparent signal and noise level is obviously reduced when processed through a passband filter. Thus, the signal and noise were scaled to compensate for different bandwidths, and further adjusted by 3 dB to convert from energy per symbol to energy per bit. The measured SNR (i.e., energy-per-bit to background ambient noise ratio) is plotted against the probability of a bit error in Fig. 6. The measured SNR calculated in this way shows a wide range of bit errors for a given SNR. For instance, at 20 dB there is a spread in error probability from $<2\%$ to around 40%. Furthermore, the predicted bit errors using a Rayleigh fading channel model [Eq. (14)] is shown as a solid line in Fig. 6. This would imply a severely fading channel much worse than even a Rayleigh model described in Sec. III B. The measured SNR in Fig. 6 is the energy-per-bit to background ambient noise ratio. This is different from the effective noise that includes multipath, which has the biggest impact on bit errors as will be described in the next section.

D. Measured energy-per-bit to noise ratio with multipath included

Due to significant multipath that exists in many shallow water environments, the concept of "signal plus multipath" has been introduced.[11] The idea being that the "true noise" is a combination of ambient noise plus multipath and "true signal" also contains many arrivals. If there is more than one arrival, i.e., direct and surface bounce paths, they will interfere producing a tone stronger or weaker than either of the individual arrivals (again, remembering that it is assumed that Doppler is not significant in these discussions). In addition, subsequent arrival(s) may "bleed" (ISI) into the next symbol's time slot which may cause an error. The duration and stability of the multipath controls the fading of the shallow water channel.

Theoretical FSK performance for both fading and nonfading channels was shown to be a function of SNR or, more specifically, the energy-per-bit ($\mathcal{E}_b$) to noise (N_0) ratio. The SNR calculation presented for the KauaiEx data in Sec. III C does not show typical SNR-dependent performance, due in part to multipath interference effects. In fact, the background ambient noise level has little to do with the actual performance in these data. An effective $\mathcal{E}_b$ and N_0 can be measured that also contains the multipath. The energy-per-bit-with-multipath ($\mathcal{E}_{bM}$) to noise-with-multipath ratio (N_{0M}) is the effective $\mathcal{E}_b/N_0$ needed to compare with the theoretical predictions and to determine the channel fading statistics. To obtain $\mathcal{E}_{bM}/N_{0M}$ the same procedure is used as for decoding the data as described in the introduction to Sec. III. An *m-sequence* matched filter is applied to the received signal for aligning symbol timing. A spectrogram with a T_b-length boxcar window is then computed. This will produce the on-tone amplitudes and the amplitude of bins that should be off. The average of all the on bins and all the off bins is a direct measurement of $\mathcal{E}_{bM}$ and N_{0M} and this is used to compare with predicted fading models.

The measured probability of bit error is shown as a function of measured $\mathcal{E}_{bM}/N_{0M}$ in Fig. 7. Also shown are the theoretical performance for $M=4$ FSK in a nonfading [Eq. (13)] and Rayleigh fading channel [Eq. (14)]. As can be seen, the data roughly fall on the curve predicted for a Rayleigh fading channel; however, the fading is slightly worse at low values of $\mathcal{E}_{bM}/N_{0M}$ and slightly better at high $\mathcal{E}_{bM}/N_{0M}$. Further, the figure has a color coding which shows the dependency on hydrophone depth and indicates that the $\mathcal{E}_{bM}/N_{0M}$ is lower for the shallow hydrophones and the bit errors are higher.

Interestingly, in Fig. 7, the best performance at the highest $\mathcal{E}_{bM}/N_{0M}$ ratios has a fading characteristic that is much better than the Rayleigh model yet not quite as good as a nonfading channel. These, isolated data, were fit to a Nakagami fading model with a 1.5 fading factor. This indicates times when the channel had a dominant arrival and results were closer to the nonfading channel. Results in Fig. 7 indicate that better performance is achieved through larger $\mathcal{E}_{bM}/N_{0M}$ but this does not imply higher source level. The performance degradation occurs because of multipath and not the ambient noise level so only a more favorable geom-

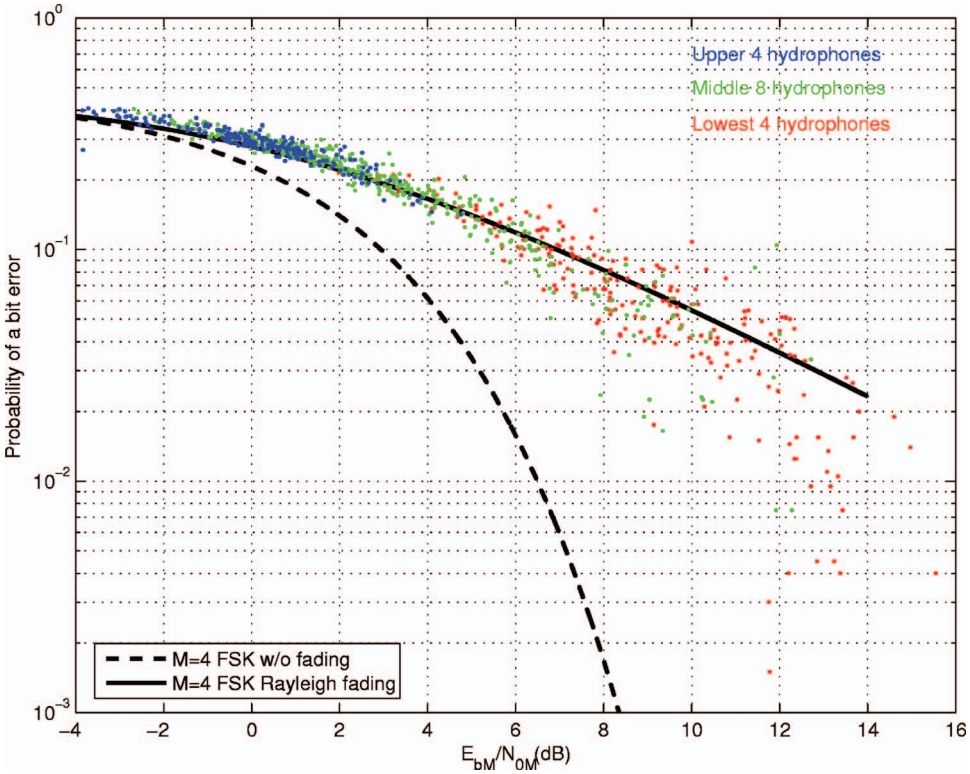


FIG. 7. Probability of bit error for transmissions during 1 day of KauaiEx vs $\mathcal{E}_{bM}/N_{0M}$. The solid line is the theoretical performance for a Rayleigh fading channel and the dashed line is for a nonfading channel. The colored dots indicate receiver depth within the water column. The blue dots are the shallowest hydrophones, the green dots are in the middle of the water column, and the red dots are the deepest hydrophones.

etry (or different environmental conditions) can improve performance. This is indicated by observing the color coded points in Fig. 7 showing different $\mathcal{E}_{bM}/N_{0M}$ at different receive depths with the same source level. This point will be demonstrated further in Sec. IV.

E. Measured and predicted channel fading

The envelope of the measured amplitude distributions for shallow and deep hydrophones taken over the 1 day experiment is shown in Fig. 8 as solid black and gray lines. The distributions are also fit to the best Rayleigh curves and those are shown as dashed lines. Figure 8 shows the shallow hydrophone has a distribution that is worse than Rayleigh. That is, the amplitudes are smaller being clustered more tightly near zero. For hydrophones at these depths a more general fading model that uses, for example, the Nakagami-*m* distribution (with an *m* parameter less than 1 for more severe fading than Rayleigh) may provide a better prediction. However, the intermediate and deep hydrophone are fit very well to a Rayleigh curve. These distributions are consistent with the bit error probabilities, that showed slightly worse than Rayleigh fading on the shallow hydrophones and nearly Rayleigh fading on the deep hydrophones.

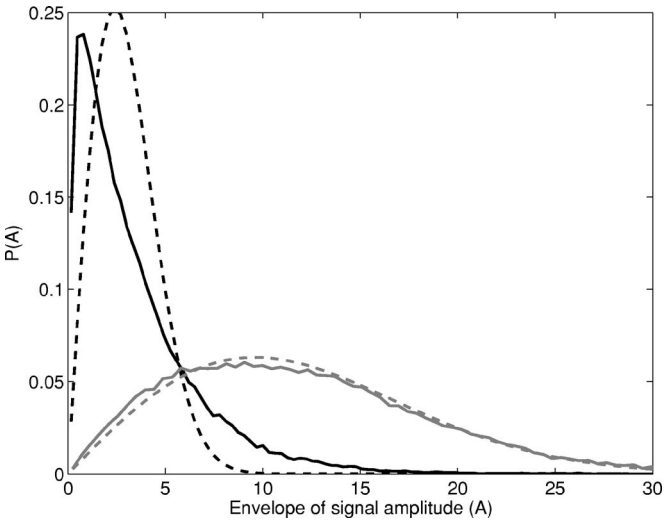


FIG. 8. The envelopes of the amplitude distributions for shallow (16.5 m, black solid curve) and deep (86.5 m, gray solid curve) hydrophones averaged over the 1 day experiment. Also shown are the curves that best fit using a Rayleigh distribution (dashed lines).

Although the distribution for the deep hydrophones is approximately Rayleigh when averaged over the 1 day experiment, there is less agreement when considering shorter time scales. Figure 9 depicts two curves corresponding to the envelope of the amplitude distribution when signaling on one of the deep hydrophones. Each curve represents three 0.8 s transmissions occurring at two different times: (1) the gray dashed line corresponds to a time frame when the water column is well mixed, nearly no thermocline, and incidentally also corresponds to higher error rates and (2) the black dashed line corresponds to a time frame when the thermocline is well established, and also corresponds to lower error rates. The solid lines are the best-fit curves and demonstrate that during the mixed period with high error rates, the distribution is closely fit to Rayleigh; on the other hand, a Ricean (Sec. III B 1) curve fits best to the period with a well-defined thermocline and corresponding lower error A rates.

To summarize, the Rayleigh fading model is useful to explain much of the data; however, it is not universally applicable. In particular, it is not applicable when there is a dominant arrival as was the case for deep receivers with a

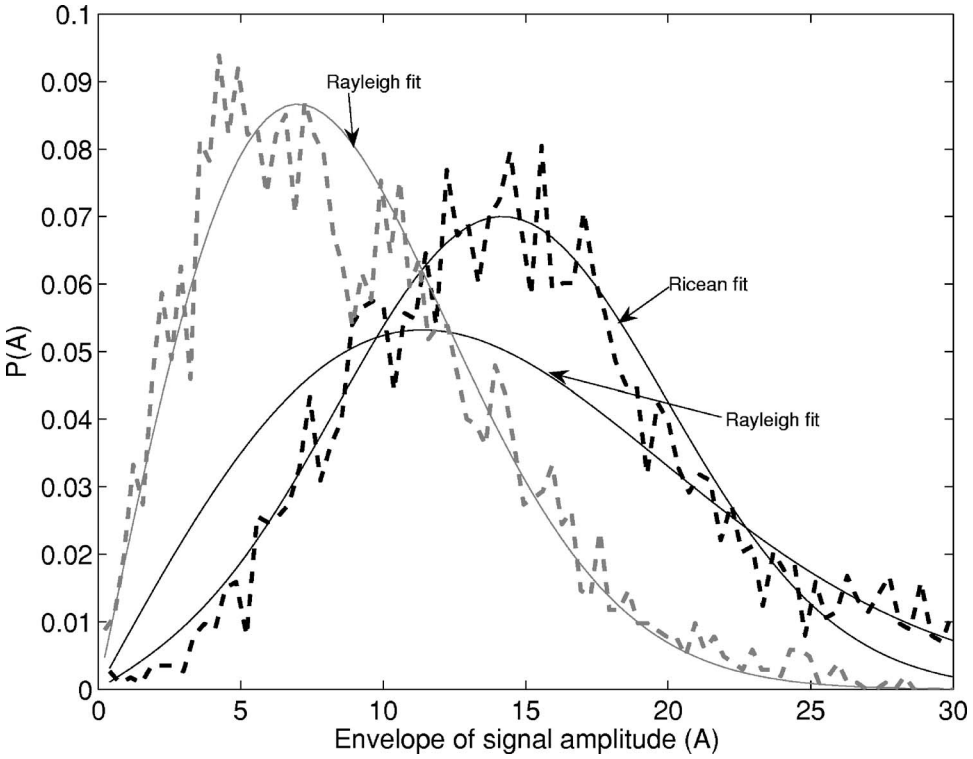


FIG. 9. Envelopes of the amplitude distributions on a short time scale (data are taken from 3, 0.8 s transmissions) for a deep hydrophone at 86.5 m. During the period with the water column mixed, the errors increased and the amplitude distribution is represented by the gray dashed curve, with the solid gray being the best fit to a Rayleigh distribution (higher BER). For the period with a strong thermocline, there were much fewer errors and the amplitude distribution for that time is shown as a black dashed line with the best fit Rayleigh and Ricean curves as solid black lines (lower BER).

well-established thermocline. Further, it is important to consider the time scale over which a statistic is desired. Modem performance on a very long time scale is more likely to average through periods where there is a dominant multipath. The statistics over the long time will then be characterized by the Rayleigh fading model. In contrast, predicting performance over a short period may require taking into account a situation with a single dominant multipath. The short-time scenario might be an AUV deployed for a few hours, while the long-time scenario might be for a fixed networked deployed for many months.

F. Impulse response measurements

The channel impulse response is one of the most important measurements for understanding propagation physics. Snapshot impulse response measurements reveal instantaneous multipath structure, duration, and strength; taken over time, these measurements often show the impulse response time-variability due to changing environmental conditions. These time-stacked impulse responses can be used to understand the amplitude distributions and how they impact performance.

A matched filter was applied to 50 ms, 8–16 kHz, linear frequency modulated (LFM) probe signals that were transmitted during KauaiEx to provide an equivalent, band-limited impulse response. Matched filtering was implemented for each receiver on the vertical array. In Fig. 10, an example is shown of measured impulse responses for periods corresponding to times with low (a) and high (b) bit errors. Figure 10 clearly shows multipath arrivals with duration of 50–100 ms, which is greater than the FSK symbol length of 25 ms. The figures have a 30 dB dynamic range scaled by the largest values. For the data in Fig. 10(a), there is a region around 0.025 s on the deeper hydrophones that shows a much larger amplitude arrival relative to the others. In Fig. 10(b) note that there is no one dominant arrival.

IV. PERFORMANCE PREDICTION USING A CHANNEL SIMULATOR

In recent years, advances have been made in using physics based, propagation modeling to simulate the channel impulse response and communications performance.[11,12] However, there have been very few experiments with simultaneous acoustic and environmental measurements to the extent taken during KauaiEx. These simultaneous measurements are needed for model validation and model/data comparisons. The simulation tool used for comparing measured data with modeled results is based on the Gaussian-beam tracing code BELLHOP[13] with an added feature to allow for moving platforms (i.e., Doppler effects).[14] This added feature, which produces different Doppler on each propagation path, is not exploited here since the source and receiver are stationary.[14] This simulator can be used with any communications signal in environments that vary volumetrically. That is, variable bathymetry and seabed properties, and depth- and range-dependent sound speed can be included for both coherent and noncoherent simulated transmissions.

Only the static case (i.e., simulations with source and receivers in fixed positions) will be described here. In this case, the BELLHOP model produces a set of arrivals each with the appropriate time delay and a complex amplitude. To describe the process of obtaining simulated communication transmissions, begin by noting that the complex pressure field, $P(\omega)$, can be represented as a sum of K arrival amplitudes $A_k(\omega)$ and delays $\tau_k(\omega)$ according to

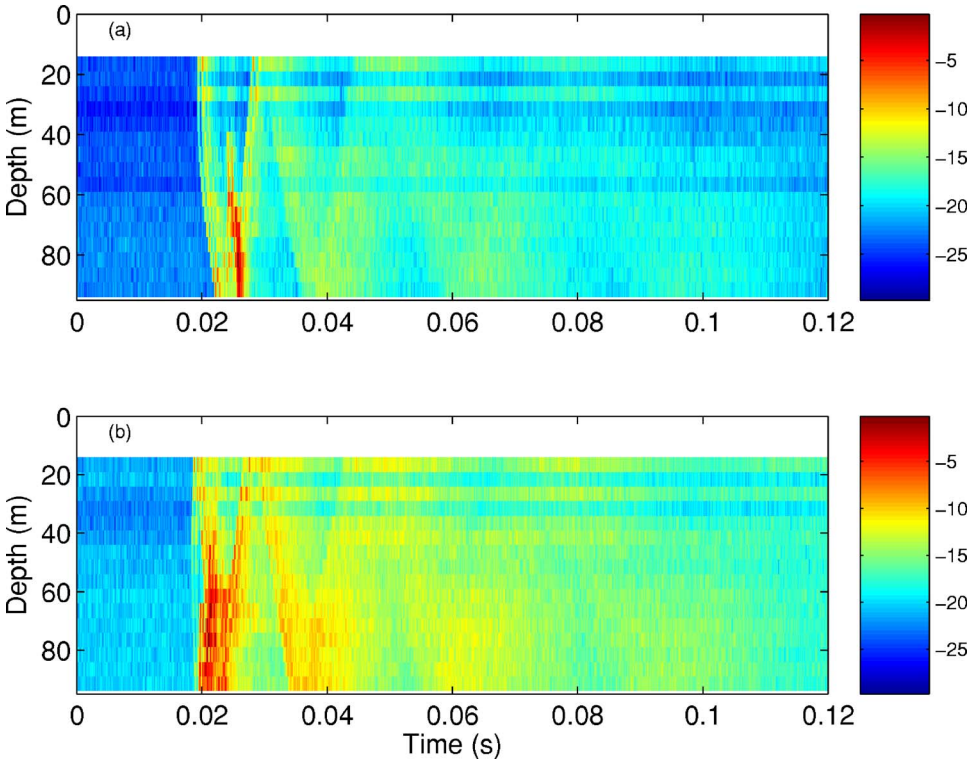


FIG. 10. (Color online) Measured impulse responses (magnitude) from matched-filtered LFM transmissions at times with low and high bit errors. Each panel is on a relative scale of 30 dB with the highest overall peak set to 0 dB. Panel (a) corresponds to a period with bit errors less than 5% (day 0.15) and panel (b) to errors close to 20% (day 0.35). The notable feature in (a) is the focusing of arrivals at the deepest hydrophones that produces a peak, in time, much larger than the other paths.

$$P(\omega) = S(\omega)\sum_{k=1}^{K} A_k e^{-i\omega\tau_k}. \tag{19}$$

The convolution theorem states that a product of two spectra is a convolution in the time domain. This leads to the corresponding time-domain representation for the received wave form, $p(t)$, which is often written as

$$p(t) = \sum_{k=1}^{K} A_k s(t-\tau_k), \tag{20}$$

where $s(t)$ is the source wave form. This representation is very intuitive, showing the sound that is heard as a sum of echoes with various amplitudes and delays. However, the amplitudes are complex to account for the interactions with the seabed and the additional time delays introduced. A more careful application of the convolution theorem considers the complex amplitudes and the conjugate symmetry of $P(\omega)$ which is necessary to guarantee a real received wave form. The proper result is then

$$p(t) = \sum_{k=1}^{K} \mathrm{Re}\{A_k\}s(t-\tau_k) - \mathrm{Im}\{A_k\}s^{+}(t-\tau_k), \tag{21}$$

where $s^{+}=\mathcal{H}(s)$ is the Hilbert transform of $s(t)$. The Hilbert transform is a 90° phase shift of $s(t)$ and accounts for the imaginary part of A_k. Equation (21) states that any arbitrary phase change can be understood as a weighted sum of the original wave form and its 90° phase-shifted version. The weighting controls the effective phase shift which occurs at bottom reflections and can yield arbitrary phase shifts. Additionally, paths that refract within the water column can be distorted in a similar way as the waves pass through caustics. It should be noted that for these simulations the seabed is treated as an infinite half-space which is reasonable since in the communications frequency band there is minimal penetration into the seabed. The half-space representation allows for a single ray trace to be used when constructing the broadband time series. This allows for rapid calculation of these high-frequency, broad band transmissions.

Simulations for KauaiEx. During the second major deployment during KauaiEx, the source was located at 95 m depth and the 16-element receiver array was 3 km away at depths of 16.5–91.5 m in 5 m increments. For the simulations, the seabed properties used were compressional sound speed of 1600 m/s, attenuation of 0.5 dB/λ, and density of 1.8 g/cm^3. The ray traces between the source and the deepest hydrophone are shown in Fig. 11 for the two time periods previously discussed, that is, low and high bit errors. The bit error for just the deepest hydrophone at 91.5 m is shown in the top panel of Fig. 11 and the middle panel depicts the water-column sound speed structure during the same period. During the first period, the thermocline existed well above the hydrophones and during the second, the thermocline was absent (or nearly).

Impulse response simulations. Fifty millisecond LFM transmissions from 8 to 16 kHz were simulated and matched filtered in the same way the measured impulse responses were processed. These impulse response plots are shown in Fig. 12 and can be compared with the measured impulse responses shown in Fig. 10. In these cases, the water column sound-speed profile measured near the VLA was used for times closest to the measured impulse responses shown in Fig. 10. Note the strong focused region for the deeper hydrophones and the similarity to the measurements.

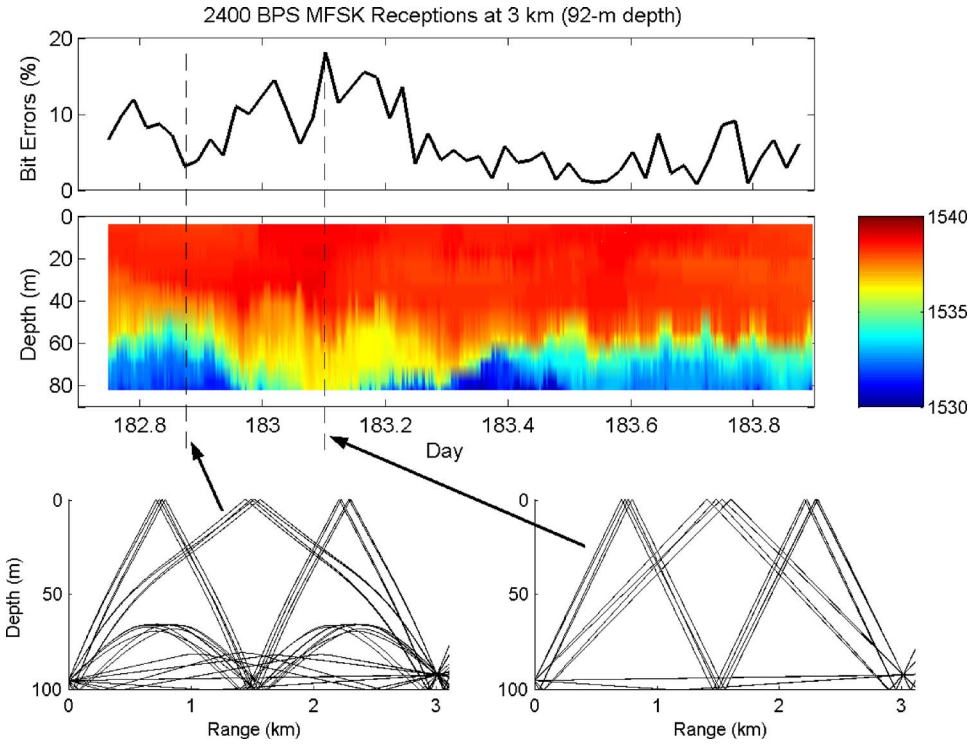


FIG. 11. (Color online) Top panel shows the percent bit errors for 2400 bps transmissions over about 1 day using the deepest hydrophone at 91.5 m. The middle panel shows the corresponding water column sound speed profile. The lowest left panel shows the ray trace that corresponds to the time with a thermocline (lower error) and the lower right panel for mixed water column (higher errors).

Simulations of FSK fading and performance. Communication signals as input to simulators can provide insight into observed performance behavior and predict optimal geometries and/or performance under different environmental conditions. During KauaiEx, extensive oceanographic measurements were made and can be used to improve simulation fidelity. The thermistor-array data provided a sound-speed profile measurement every minute and could be assumed to be representative of oceanographic conditions between the source and receiver. Simulations for computing impulse responses were conducted for each profile and showed strong agreement with the measured responses. A model-generated impulse response was generated with each new sound speed profile measurement. The FSK signals were then convolved with the simulated channel impulse response, and demodulated by a virtual receiver. This was done for each of the 16 hydrophones in the vertical array for each sound-speed-profile measurement time step. The bit error percentage as a

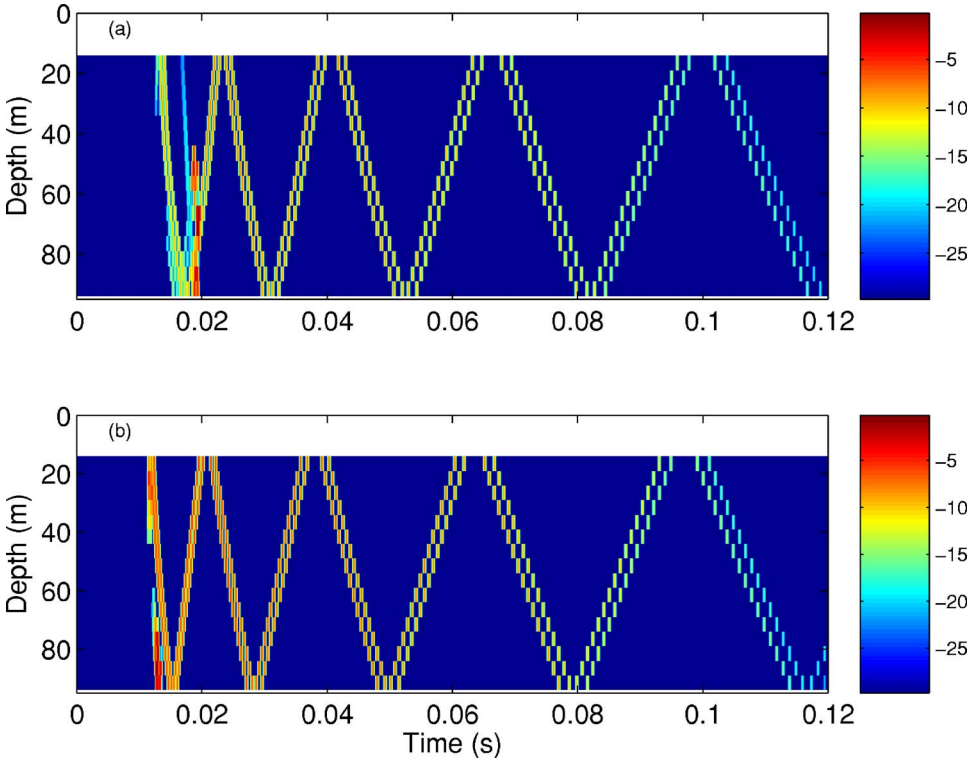


FIG. 12. (Color online) Simulated impulse responses from matched-filtered LFM transmissions at times with low and high bit errors. Each panel is on a relative scale of 30 dB with the highest overall peak set to 0 dB. Panel (a) corresponds to period with low bit errors [measured impulse response at that time is shown in panel (a) in Fig. 10] and (b) to the period with high bit errors [corresponding to (b) of Fig. 10]. As with the measured impulse responses, the notable feature in (a) is the focusing of arrivals at the deepest hydrophones that produces a peak in time much larger than the other paths.

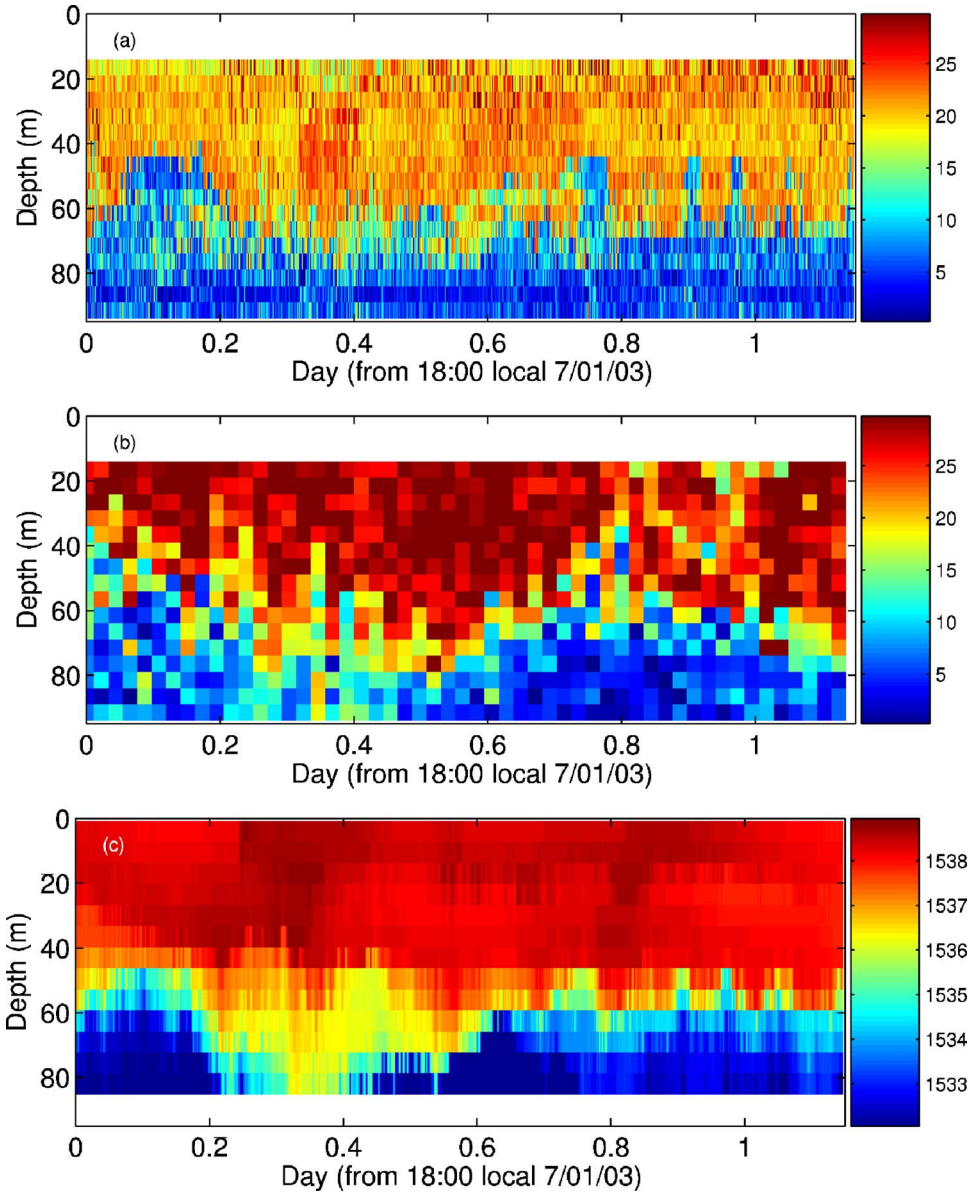


FIG. 13. (Color online) Panel (a) is the simulated percent bit errors at the 16 hydrophone depths over the 1 day experiment (percent error). Panel (b) is the same for the measured data (percent error). Panel (c) is the sound speed profile (m/s).

function of time is shown in Fig. 13 along with the measurements of the same quantity. Also shown, for comparison, is the sound-speed profile during the same period. From Fig. 13, several points can be made. First, the simulations and measurements are very similar in both time and space. Second, the upper and lower portions of the water column show very different bit errors in both the measurements and simulation. Third, both appear to track changes in the oceanography in a similar way. Last, it is interesting to note that the overall best performance was observed for the second deepest hydrophone (at about 86.5 m) and this was duplicated in the simulation.

The simulator-computed fading characteristics were calculated in a similar manner to that discussed in Sec. III E. In Fig. 14, the envelope of the amplitude distributions for the same hydrophone (Sec. III E, second deepest) is shown for both high and low bit-error periods corresponding to when the thermocline was weak (mixed water column) and when it was strong, respectively. The distributions computed by the simulator and directly from experimental data show good agreement. A strong thermocline results in a shifted distribution toward higher amplitudes and an approximate Ricean curve fit. A more uniformly mixed water column results in lower amplitudes and an approximate Rayleigh curve fit.

Finally, the model-predicted, bit-error probabilities are compared with the theoretical performance curves for $M=4$ FSK signaling; the results are shown in Fig. 15. The points on the figure are color coded showing the upper 4 hydrophones in blue, the middle 8 hydrophones in green, and the deepest 4 hydrophones in red. As was the case for the measurements, the performance improves with depth as $\mathcal{E}_{bM}/N_{0M}$ increases. As was the case for the measurements, at the low end of $\mathcal{E}_{bM}/N_{0M}$ the errors are slightly worse than the Rayleigh-fading prediction. At the high $\mathcal{E}_{bM}/N_{0M}$ end, the performance is much better than Rayleigh but not quite reaching the nonfading performance curve. It is important to

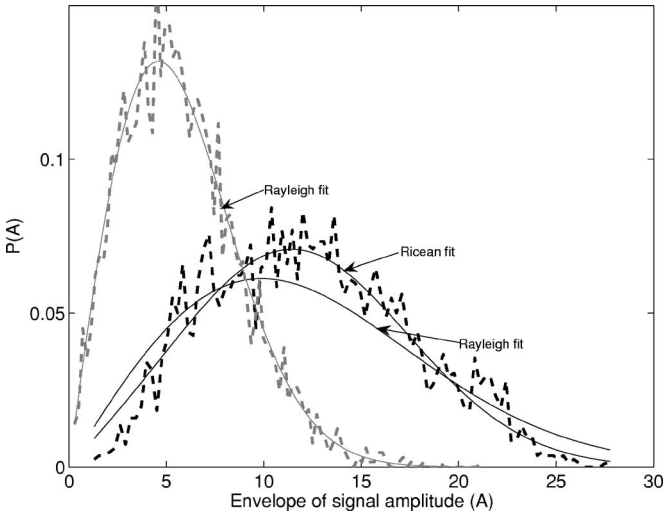


FIG. 14. Simulation: Envelopes of the amplitude distributions on a short time scale (data are taken from 3, 0.8 s transmissions) for just the deep hydrophone at 86.5 m. During the period with the water column mixed, the errors increased and the amplitude distribution is shown by the gray dashed curve with the solid gray being the best fit to a Rayleigh distribution (higher BER). For the period with a strong thermocline, there were much fewer errors. The amplitude distribution is shown as a black dashed line with the best fit Rayleigh and Ricean curves as solid black lines (lower BER).

note here that this performance calculation was done *without* added noise. The performance is nearly the same as that measured indicating the background ambient noise has little to do with the performance in this regime. This was verified by adding background ambient noise (equal to that for the measurements) and there was no significant change in the simulated results shown in Fig. 15.

V. DISCUSSION AND CONCLUSIONS

The communication performance dependence on source-receiver geometry and oceanographic conditions have been described for FSK transmissions over distances of 3 km in the 8–13.2 kHz band for an experimental site near Kauai, HI. Since the received signal level was well above the ambient noise level, the limiting factor in the performance was the multipath interference. Thus, key factors in the modem performance were the source/receiver geometry and the oceanography. Using measured sound-speed profiles, simulations were made to mimic the measured data collection over the 1 day experiment. Results showed a simulated performance very similar to that measured. This held true even in the absence of added ambient noise in the channel. This is somewhat counterintuitive but it implies that once the channel is no longer ambient-noise limited, increasing the source level has no impact on performance.

The greatest improvements in performance were achieved by changing the receiver depth (the transmitter depth was fixed). The communication signals from receivers in the middle of the water column showed fading consistent with a Rayleigh-fading model over much of the experiment duration. The shallowest hydrophones exhibited slightly worse fading characteristics and the deepest hydrophones were slightly better. During the most favorable periods when there was a strong thermocline, the deepest hydrophones ap-

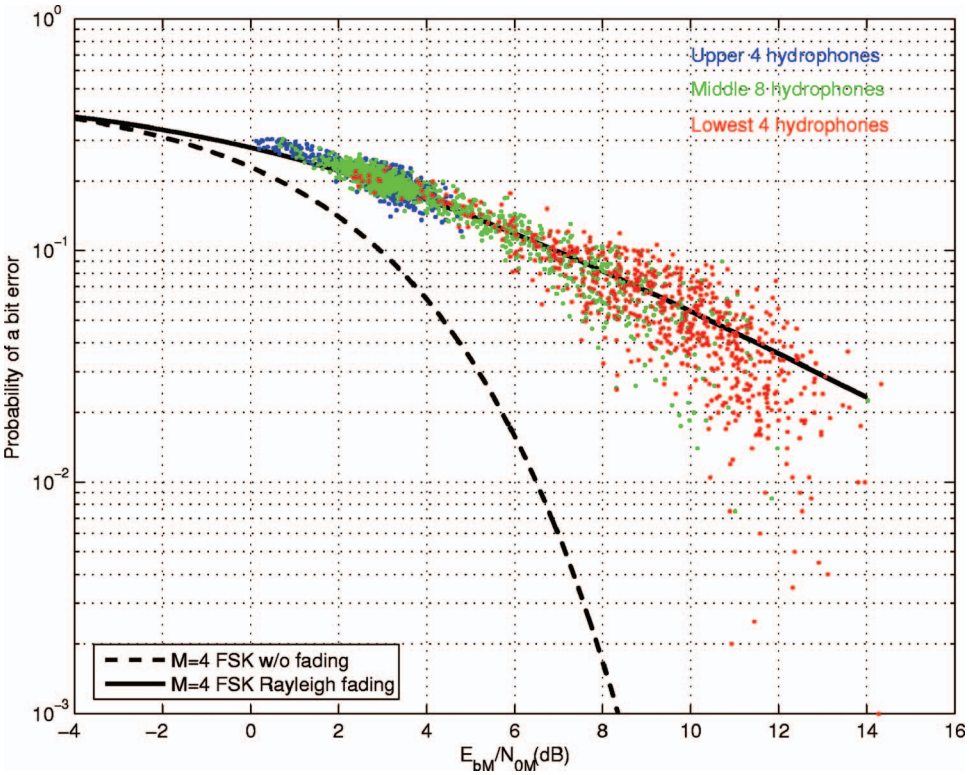


FIG. 15. Simulation: Probability of bit error for transmission during 1 day of KauaiEx vs $\mathcal{E}_{bM}/N_{0M}$. The solid line is the theoretical performance for a Rayleigh fading channel and the dashed line is for a nonfading channel. The colored dots indicate depth within the water column. The blue dots are the shallowest hydrophones, the green dots are in the middle of the water column, and the red dots are the deepest hydrophones. This simulation was done *without* any ambient noise added.

proached the characteristics of a nonfading channel. The thermocline varied significantly during the 24 h measurement period and there was a period when the thermocline nearly disappeared and the water column was entirely mixed (iso-speed). At this time, the deeper hydrophones lost their favorable conditions and error rates increased significantly.

Modeling was used to show how acoustic energy is trapped due to the thermocline giving rise to the observed favorable arrival structure in the lowest hydrophone depths. The favorable arrival structure is characterized by a very large amplitude arrival (or group of arrivals) that is not the earliest arrival(s). When the water column is mixed, the thermocline is gone and the lowest hydrophones show a similar impulse response to the shallower hydrophones and performance is similar.

Oceanographic conditions like these are common. Summer conditions often give rise to a strong thermocline, while winter conditions usually show more mixing resulting in iso-speed profiles. These results, together with the modeling, show how the environment can play a significant role in underwater acoustic communications performance. In situations similar to those during KauaiEx, a strategy for optimizing performance might include avoiding transmission times when the water column is mixed and concentrating assets near the seabed as opposed to near the surface. While this is not a general rule since environmental conditions vary at different locations, the modeling results show how predictions can be made if sufficient environmental knowledge exists.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. We would like to express particular appreciation to the team from the Marine Physical Laboratory at the University of California, San Diego, William Hodgkiss, Jeff Skinner, and Dave Ensberg for the vertical array data used here. The authors also gratefully acknowledge the University of Delaware team, led by Mohsen Baidey, for the CTD and thermistor data used for this analysis. We would also like to thank Naval Research Enterprise Internship Program (NREIP) student Laura Meathe and SPAWARSYSCEN, San Diego employee Leo Ghazikhanian for their assistance with operating the Telesonar Testbed instrument. Additionally, we would like to acknowledge Joe Rice for the Telesonar Testbed concept and for his support during its development. The KauaiEx Group consists of: Michael B. Porter, Paul Hursky, Martin Siderius (HLS Research), Mohsen Badiey (University of Delaware), Jerald Caruthers (University Southern Mississippi), William S. Hodgkiss, Kaustubha Raghukumar (Scripps Institute of Oceanography), Daniel Rouseff, Warren Fox (University of Washington), Christian de Moustier, Brian Calder, Barbara J. Kraft (University of New Hampshire), Keyko McDonald (SPAWARSSC), Peter Stein, James K. Lewis, and Subramaniam Rajan (Scientific Solutions).

[1]M. B. Porter and the KauaiEx Group, "The Kauai experiment," in *High-Frequency Ocean Acoustics* (AIP, Melville, NY, 2004), pp. 307–321.
[2]M. Siderius, M. B. Porter, and the KauaiEx Group, "Impact of thermocline variability on underwater acoustic communications: Results from KauiEx," in *High-Frequency Ocean Acoustics* (AIP, Melville, NY, 2004), pp. 358–365.
[3]M. B. Porter, V. K. McDonald, P. A. Baxley, and J. A. Rice, "Signalex: Linking environmental acoustics with the signaling schemes," in *Proceedings of MTS/IEEE Oceans00* (IEEE, New York, 2000), pp. 595–600.
[4]N. M. Carbone and W. S. Hodgkiss, "Effects of tidally driven temperature fluctuations on shallow-water acoustic communications at 18 kHz," IEEE J. Ocean. Eng. **25**, 84–94 (2000).
[5]J. G. Proakis, *Digital Communications*, 3rd ed. (McGraw-Hill, New York, 1995).
[6]D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[7]J. Rice *et al.*, "Evolution of seaweb underwater acoustic networking," in *Proceedings of MTS/IEEE OCEANS'00 Conference* (IEEE, New York, 2000), pp. 2007–2017.
[8]V. K. McDonald, P. Hursky, and the KauaiEx Group, "Telesonar testbed instrument provides a flexible platform for acoustic propagation and communication research in the 8–50 kHz band," in *High-Frequency Ocean Acoustics* (AIP, Melville, NY, 2004), pp. 336–349.
[9]J. G. Proakis, "Coded modulation for digital communications over Rayleigh fading channels," IEEE J. Ocean. Eng. **16**, 66–73 (1991).
[10]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
[11]A. Zielinski, Y. H. Yoon, and L. Wu, "Performance analysis of digital acoustic communication in a shallow water channel," IEEE J. Ocean. Eng. **20**, 293–299 (1995).
[12]C. Bjerrum-Niese, L. Bjorno, M. Pinto, and B. Quellec, "A simulation tool for high data-rate acoustic communication in a shallow-water time-varying channel," IEEE J. Ocean. Eng. **21**, 143–149 (1996).
[13]M. B. Porter and H. P. Bucker, "Gaussian beam tracing for computing ocean acoustic fields," J. Acoust. Soc. Am. **82**, 1349–1359 (1987).
[14]M. Siderius and M. B. Porter, "Modeling techniques for marine mammal risk assessment," IEEE J. Ocean. Eng. **31**, 49–60 (2006).

Bi-static sonar applications of intensity processing

Nathan K. Naluai[a] and Gerald C. Lauchle
The Pennsylvania State University, Graduate Program in Acoustics, 217 Applied Science Building, University Park, Pennsylvania 16802

Thomas B. Gabrielson
The Pennsylvania State University, Applied Research Laboratory, P.O. BOX 30, State College, Pennsylvania 16804-0030

John H. Joseph
NAVAIR - NAWCAD, 22347 Cedar Point Road, Unit 6, Patuxent River, Maryland 20670-1161

(Received 27 April 2006; revised 16 January 2007; accepted 17 January 2007)

Acoustic intensity processing of signals from directional sonobuoy acoustic subsystems is used to enhance the detection of submerged bodies in bi-static sonar applications. In some directions, the scattered signals may be completely dominated by the incident blast from the source, depending upon the geometry, making the object undetectable by traditional pressure measurements. Previous theoretical derivations suggest that acoustic vector intensity sensors, and the associated intensity processing, are a potential solution to this problem. Deep water experiments conducted at Lake Pend Oreille in northern Idaho are described. A large, hollow cylindrical body is located between a source and a number of SSQ-53D sonobuoys positioned from 5 to 30 body lengths away from the scattering body. Measurements show changes in the acoustic pressure of less than 0.5 dB when the scattering body is inserted in the field. However, the phase of the acoustic intensity component formed between the acoustic pressure and particle velocity component orthogonal to the direction of incident wave propagation varies by as much as 55°. This metric is shown to be a repeatable and strong indicator of the presence of the scattering body. [DOI: 10.1121/1.2642243]

PACS number(s): 43.30.Vh, 43.20.Fn, 43.30.Zk [LLT] Pages: 1909–1915

I. INTRODUCTION

In bi-static sonar, acoustic energy is transmitted from a source, propagates through the medium, and scatters from nearby objects. The time-delayed direct and any scattered signals are received by remotely positioned receivers. For most receiver, source, and object positions, the scattered signal is distinguished from the direct path signal because it arrives at a later time. This time delay, however, becomes essentially zero when the object is on the line of sight (LOS) between the receiver and source. In this scenario, the incident and scattered acoustic components arrive at the receiver at the same time and the received signal is dominated by the incident blast from the source.

Many authors have discussed the forward-scattering problem in connection with imaging, radar, and sonar systems.[1–8] These investigations emphasize calculations—or measurements—of the scalar field magnitude. Such scalar field measurements in acoustics are equivalent to determining the acoustic pressure field in the region of interest. Recent studies suggest that acoustic intensity processing of the signals received by acoustic vector intensity sensors may present a solution to the forward-scattering problem. Theoretical computations of acoustic scattering from a rigid prolate spheroid insonified by single frequency plane waves have recently been completed.[9–12] These studies have suggested that the pressure gradients created within interference regions of the total (incident plus scattered) field cause significant variations in both the reactive intensity and the phase angle between pressure and particle velocity. The forward-scattered region is composed of a mixture of incident and scattered wave components; therefore, it is a region rich in acoustic pressure gradients and intensity phase variations.

The present paper attempts to experimentally verify the theoretical study in a practical underwater scenario. The fundamental acoustic intensity relations of the proposed theoretical model of the scattered intensity field are briefly reviewed. Experiments designed to determine the existence of a scattering body in the acoustic field, particularly in the forward-scattering direction, are presented.

II. SCATTERED ACOUSTIC INTENSITY

This section summarizes some of the basic equations of relating acoustic pressure and particle velocity to the acoustic intensity. Detailed derivations for the intensity concepts that are available in other references[11,13,14] are not presented here. Previously developed theory of acoustic intensity field scattered from a rigid elliptical body is described.

A. Fundamental equations for acoustic intensity

The classic definition of the time-averaged real acoustic intensity vector at a point **r** in space is given as one half of

[a]Author to whom correspondence should be addressed. Electronic mail: nathan.naluai@navy.mil

the real part of the product of the complex acoustic pressure and the conjugate of the complex particle velocity

$$\mathbf{I}(\mathbf{r}) = \tfrac{1}{2}\,\mathrm{Re}\{p(\mathbf{r},t)\mathbf{u}^*(\mathbf{r},t)\}. \quad (1)$$

This quantity is often called the ordinary active intensity and represents the net transport of acoustic energy.

For acoustic fields composed of multiple interfering acoustic waves, it is often helpful to consider the concept of complex acoustic intensity. The complex intensity is defined as one half the product of the complex pressure and the conjugate of the complex particle velocity,

$$\mathbf{I}_c(\mathbf{r}) = \tfrac{1}{2}p\mathbf{u}^* = \mathbf{I}(\mathbf{r}) + i\mathbf{Q}(\mathbf{r}). \quad (2)$$

The real part of the complex intensity, **I**, is identical to the ordinary active intensity of Eq. (1). The imaginary part, **Q**, is commonly referred to as the reactive intensity. It is of interest in this study because it represents the presence of local oscillatory energy flow. Local oscillatory energy is required to support the existence of any spatial heterogeneity in the acoustic pressure field.

The relative phase between the pressure and particle velocity, $\phi_{p\mathbf{u}}$, is equal to the complex intensity phase. For pure plane waves, the pressure and particle velocity are everywhere in phase ($\phi_{p\mathbf{u}} \equiv 0$), and the reactive intensity is zero. If acoustic pressure and velocity are in quadrature, the active intensity is zero but the reactive intensity is maximized (e.g., standing wave field). In the case of wave fronts having curvature or for scattered acoustic fields, both **Q** and $\phi_{p\mathbf{u}}$ will be nonzero due to the presence of local gradients in the pressure amplitude. The measurement of reactive intensity and/or intensity phase may therefore be a viable metric in locating these gradients in the scattered acoustic field.

B. Theoretical scattered intensity model

When a scattering object is insonified by an acoustic wave, the scattered and incident acoustic fields will interfere and produce pressure gradients. At very low frequencies, where the wavelength of sound is large compared to the typical dimensions of the scattering body, the incident sound passes by the body with little distortion; scattering is virtually absent in this so-called Rayleigh region. At wavelengths comparable to the dimensions of the scattering body, the object scatters energy specularly (the angle of the incident acoustic signal to the scattering body being approximately equal to the angle of the predominant scattered intensity).

As the frequency of the incident wave increases still further and the wavelength becomes comparable to the smallest radii of curvature of the scattering body, acoustic diffraction effects appear on the far side of the body. In this direction, acoustical interference patterns form regardless of the incident sound pulse length, or angle of incidence. The incident wave and the scattered wave interfere over essentially the same time scale because there is virtually no difference in path length between the two fields. Then, from the standpoint of the acoustic source, it makes no difference whether the acoustic excitation is steady state or impulsive.

Exact theoretical formulations and computations for the complex intensity field scattered by an object have recently been presented.[9–12] In these studies, the intensity field resulting from plane waves incident upon a rigid prolate spheroid is considered, where the spheroid has a 10:1 fineness ratio, defined as the ratio of the major to minor axis of the ellipsoid. The acoustic energy is incident on the spheroid at an angle of 60° from the major axis. Theoretical predictions of the equivalent plane wave intensity indicate that the resulting scattered field is dominated by the incident field. A small perturbation of the total acoustic pressure, observable in the forward-scattered direction, differs by less than 0.5 dB from the incident pressure level.

The computed intensity field, however, indicates features not observable in the total scalar field. While the complex intensity phase component in the direction parallel to the wave propagation vector is shown to be zero in the forward-scattered direction, the existence of phase gradients is predicted for the intensity component that is orthogonal to the incident wave vector. These gradients are particularly severe in the forward-scattered region, and may produce measurable variations in reactive intensity amplitude and in the complex intensity phase, $\phi_{p\mathbf{u}}$.

In their analysis, Rapids and Lauchle[12] predict a complex intensity phase shift on the order of 5°. This phase angle is shown to be proportional to frequency, and inversely proportional to the distance from the scattering body. Further, the calculations predict the occurrence of primary sidelobes in the phase angle at roughly $\psi = \pm 5°$ off the forward-scattered direction angle. These sidelobes appear to diminish with increasing frequency. These studies provide theoretical evidence that intensity-based processing may allow for the determination of the presence of a scattering body in the forward-scattered direction through measurement of variations in the phase angle of the complex acoustic intensity for the total field.

III. EXPERIMENTAL MEASUREMENTS

An experiment involving the use of SSQ-53D sonobuoys to verify the forward-scattering model theory is described in this section. The test was conducted in May 2005 at the Naval Surface Warfare Center, Lake Pend Oreille Acoustics Research Detachment, at Bayview, ID. The objective is to use intensity processed vector sensors to determine the presence of a large scattering body in a configuration of source, scatterer, and receiver in which it is typically not possible to make such a determination from measurements of the acoustic pressure alone. Intensity processing techniques are applied to the acoustic receivers, with the source operating over a broad range of frequencies, and with several receivers placed at various distances from the body of interest.

A. Experimental test design

Experimental verification of the theoretical intensity predictions proposed by Rapids and Lauchle[12] is conducted on the intermediate scale measurement system (ISMS) range, located approximately 11 miles north of Bayview, ID. Lake Pend Oreille is a freshwater lake which provides a deep (350 m) and quiet body of water simulating a free-field ocean-like environment. The ISMS range consists of a large

anchored barge, which serves as the main measurement platform, and a small remote platform that is positioned roughly 670 m to the north of the ISMS barge.

In order to test acoustic intensity scattering using the SSQ-53D sonobuoy, the dimensions of the scattering object must be much larger than the incident acoustic wavelengths. Given the frequency limits of the sonobuoy (2.4 kHz), a large metal liquid-storage tank was chosen as the scatterer. The tank is a hollow cylindrical body with hemispherical end caps which measures $L=19.2$ m in length and $d=2.44$ m in diameter. The submerged tank was oriented horizontally and positioned at a depth of 123 m.

This scattering body represents a departure from Rapids' theoretical model in several respects. A long cylindrical body oriented orthogonal to the direction of wave propagation will have a larger scattering cross section than the prolate spheroid oriented at 60° off the major axis, which is assumed in the model predictions. Also, the tank is not a rigid body, but is an air-filled container. Structural vibrations may cause reradiation of additional pressure waves that will add to the acoustic pressure waves created by scattering alone. It appears sensible that the pressure gradients generated in the forward direction will be larger and more severe. Thus, a significant increase in the values of the cross-spectral phase angles is expected from those predicted for a simple rigid prolate spheroid.

The acoustic source is an ITC 4141 spherical source driven with white noise in the primary operating range of the SSQ-53D sonobuoy (500–2400 Hz frequency band). It was secured from the ISMS barge at the test depth of 123 m and approximately 30 m from the scattering body. At this range and at the frequencies involved, the acoustic waves incident on the scattering body are considered planar.

Four SSQ-53D sonobuoys were prepared for use as intensity vector sensors. These sonobuoys are passive, directional underwater sensors used as part of the United States Navy's DIFAR acoustic subsystem. The SSQ-53Ds measure acoustic pressure and two orthogonal components of particle velocity directly in two magnetic compass corrected directions (a North-South and an East-West component). The three measured signals are combined into a composite signal which is transmitted and received via rf transmissions.

The sonobuoys were positioned at a fixed depth of 123 m, and at distances of 5, 10, 20, and 30 L from the body, where the L is the length of the scattering body. The sonobuoys, or receivers, are labeled A through D, with A being the closest to the body. These were tethered to a support line which ran from the ISMS test barge to the remote platform, positioned approximately due north of the test barge. This alignment allows the radial (north) and the transverse (east) components of acoustic intensity along this line to be measured directly, without correction for sensor orientation.

The sonobuoys were deployed early in the experiment and remained in the water for the duration of the test week. However, due to shifting winds and water currents, the absolute position of the sonobuoys relative to the magnetic north line varied from day to day. The positional offsets were not anticipated to introduce errors in the measurements significant to warrant the difficult and time-consuming adjustments of the receiver and remote platform locations. Alternatively, for each day's testing, the location of the sonobuoy surface float was noted via global positioning system (GPS) monitoring.

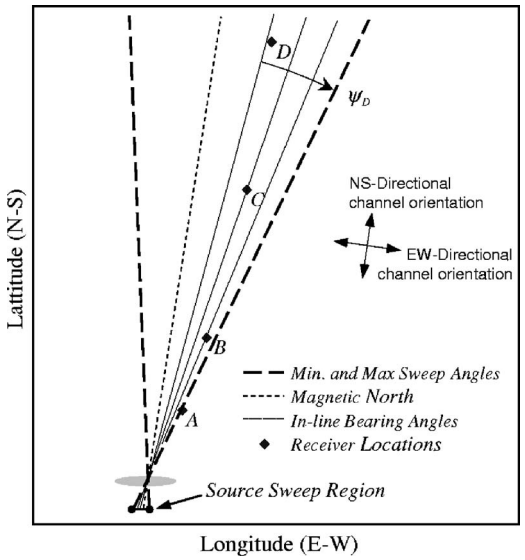


FIG. 1. Diagram of the lines of sight (LOS) to each receiver on the first of two trial days. The receivers' bearing to the LOS is given as ψ_i, illustrated for receiver D above. The sweep region is due to the variation of the source position along the ISMS barge in the E-W direction (not to scale). For this trial set (day 1), receiver A is outside the sweep region and does not cross the forward-scattered line.

On board the ISMS barge, a 31-channel ARR-75 sonobuoy receiver system and a bank of Sparton sonobuoy de-multiplexers were used for the reception of the rf sonobuoy signals and subsequent separation into the component omni-hydrophone and directional sensor outputs. The received signals are decomposed, the complex intensity components are determined directly by calculating the cross spectra between the pressure signal and each directional signal using a dynamic signal analyzer.

The position of the source was systematically varied laterally along the ISMS barge (east to west) to affect a change in the incidence angle, simulating a motion of the receiver array through the object's forward-scattered path. As such, this test configuration would match the format of the theoretical predictions presented by Rapids and Lauchle.[12] The locations of the scattering body and each top-side buoy, in addition to the 15 source positions considered, were determined from GPS readings. The net result of this variation is an effective sweep of the receiver array over a distinct range of forward-scatter bearings and across the direct line of sight (LOS) and in the plane of the major features of the scattering body. This bearing angle is denoted here as ψ_i for the ith receiver, where $\psi_i=0°$ corresponds to the line containing the source, scattering body and receiver. A plan view diagram of several of the resulting lines of sight is illustrated in Fig. 1.

Data measurements are acquired and recorded for all of the receivers at each of the source positions. The received signals of each sonobuoy are demultiplexed and the resulting component signals are processed using Agilent 35670A Dynamic Signal Analyzers. The relevant auto- and cross spectra

for a finite number of spectral averages are calculated and stored on a standard PC. Real-time recordings of the raw sonobuoy signals are also digitally captured using an Alesis ADAT HD24 data recorder. This process is repeated for each of the 15 source positions. Finally, the entire test procedure was conducted on consecutive days to verify repeatability.

It should be noted that, given this experimental design, the angle of incidence to the scattering body also changes slightly as the source is moved. However, the theoretical predictions show strong phase changes in the forward-scattered direction, regardless of incidence angle.[12] Hence, no attempt was made to correct any of the data for the small change in incidence angle. Further, the LOS and sweep angles differ slightly between the trial days due to shifting wind and surface currents. This difference is accounted for in the data analysis.

B. Data analysis and results

Analysis of the recorded data is relatively straightforward due to the simple relationship between the complex intensity spectrum and the cross spectrum.[13–17] Various cross- and auto spectra were determined among the pressure and the two particle velocity sensor outputs for each of the four sonobuoys using Agilent four-channel Dynamic Signal Analyzers (Model 35670A). This process involves spectral averaging of the computed finite Fourier transformed time records collected from the receivers at each combination of source-body-receiver location to generate the desired auto- or cross-spectral values. These spectral magnitudes and phases are then compared to the theoretical predictions. Results can also be averaged over the operating frequency range of the sonobuoys (1.0–2.4 kHz) to simplify presentation of the data. These are generally presented as a function of the ith receiver's angle to the LOS, ψ_i.

We note here that no attempt is made to determine an absolute calibration for each of the sonobuoys. Thus, the measured cross spectra can only be said to be proportional to the complex intensity spectra. Actual values for active and reactive intensity components cannot be determined. However, in this study we are interested only in relative changes in the intensity magnitude and phase as the receiver traverses the LOS. The data will accurately reflect any changes in these variables provided that calibrations of a given sonobuoy do not change over the measurement period. Therefore, an absolute calibration of each sonobuoy is not necessary.

For our clarity in presentation, it is sufficient to add a constant phase offset to each complex cross-spectral measurement such that in the absence of the scattering body, the measured complex cross-spectral phase is roughly zero. This constant offset is determined by averaging the measured phase for a given buoy over several points well outside the expected location of the direct LOS and subtracting this phase from all computed phase values for that particular buoy. This process is repeated for each buoy.

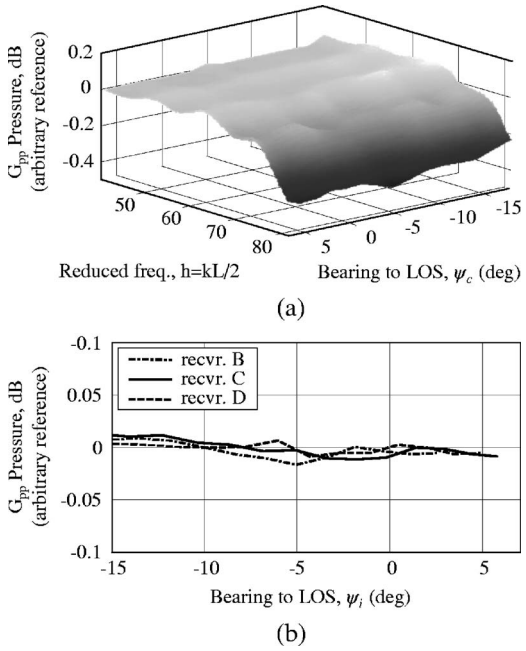


FIG. 2. Normalized acoustic pressure level (SPL) measurement (in dB with an arbitrary reference) which indicates no apparent change in level due to forward-scattering from the object. (a) Example pressure spectrum for receiver C plotted as a function of the reduced frequency, $h=kL/2$ and the bearing angle to the LOS, ψ_C; (b) SPL of other receivers in day 1 trials averaged over the operating frequency range of the SSQ-53D (1.0–2.4 kHz). Measured SPL values across the LOS for each buoy indicate variations of less than 0.2 dB. Similar results are found for day 2 trials and, hence, are not presented here. Bearing angles are approximated to account for wind and water currents.

1. Acoustic pressure

The un-calibrated acoustic pressure is measured as the auto-spectral level of the sonobuoy's omni-hydrophone output voltage, corrected for the transmit voltage response of the ITC 4141 source. The measured spectral level is converted to decibels with an arbitrary reference point. The reference point for this study is the average level measured in the absence of the scatterer. An example of the calculated pressure for receiver C is presented in Fig. 2(a) as a function of the receiver's angle to the LOS, ψ_C, and a modified non-dimensional frequency, h. This frequency relates the acoustic wave number, k, to the scattering object length, L, and is defined as

$$h = \frac{kL}{2}. \tag{3}$$

From this figure, it is readily apparent that the pressure level remains relatively constant regardless of the receiver's bearing angle. A small perturbation of less than 0.2 dB results as the receiver traverses the forward-scattered region. This is consistent with the other intensity sensors in the array, as is indicated in Fig. 2(b). As predicted by theory, there is no significant change in any of the sensors at $\psi_i=0°$, corresponding to the line-of-sight connecting source, scattering object, and receivers. For these trials, the LOS for receiver A was completely outside the sweep region of the source and was not crossed. Thus, data results for this receiver are not

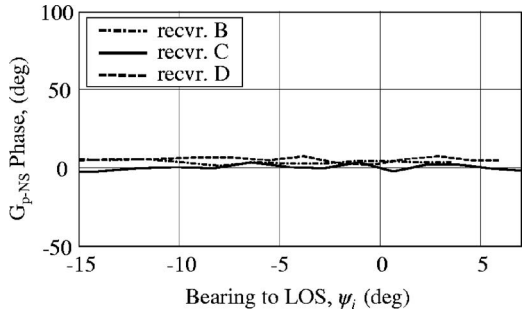


FIG. 3. Normalized intensity phase, $\phi_{p\mathbf{u}}$, in direction of incident wave (North-South) during Day 1 trials. Displayed data are averaged over the operating frequency range of the SSQ-53D. The standard deviations of the measured phase curves (on the order of 2°) indicate no significant phase shifts across the LOS. Day 2 trials match these trends, and hence, are not presented here.

presented here. These results support the theory presented by Rapids and Lauchle[12] and clearly demonstrate the problem of measuring a forward-scattered acoustic signal using a single acoustic pressure sensor.

2. Complex intensity phase

As was illustrated in theoretical model, the complex intensity phase is predicted to contain significant information regarding the acoustic field scattered by an object particularly along the forward LOS. For the complex intensity spectra, the phase of the cross spectra is measured between the pressure hydrophone and the N-S and E-W components of the particle velocity, for each SSQ-53D sonobuoy. Figure 3 shows the phase of the radial component of the complex intensity averaged over the operating frequency range of the sonobuoy as a function of the receivers' angle to the direct line of sight, ψ_i. In the current configuration, this component of the intensity phase corresponds to the complex intensity phase, $\phi_{p\mathbf{u}}$, in the direction of propagation of the incident wave (the N-S direction).

The measured data reveal only very small phase variations with a standard deviation of less than 2.0° as the receiver passes through the LOS. These changes are attributed to distant interfering sources, surface wave motion, lake reverberation effects, as well as random phase fluctuations associated with the self-noise of the receiver. Variation on this order are to be expected since accurately quantifying very small shifts in phase angles is difficult due to these extraneous effects. Notwithstanding, no significant phase changes can be observed; the cross-spectral phase is essentially constant for all angles relative to the LOS. This conclusion also supports completely the trends predicted for this component of scattered intensity.[9–12]

Particularly interesting features can be seen in the cross-spectral phase of the transverse component of the complex intensity, measured between the pressure and the E-W particle velocity component. The resulting phase measurements for each sonobuoy are averaged across its useful range and are presented as a function of the sonobuoys' angle to the LOS in Fig. 4.

The complex intensity phase data from Day 1 testing, plotted in Fig. 4(a), clearly indicate strong phase variations

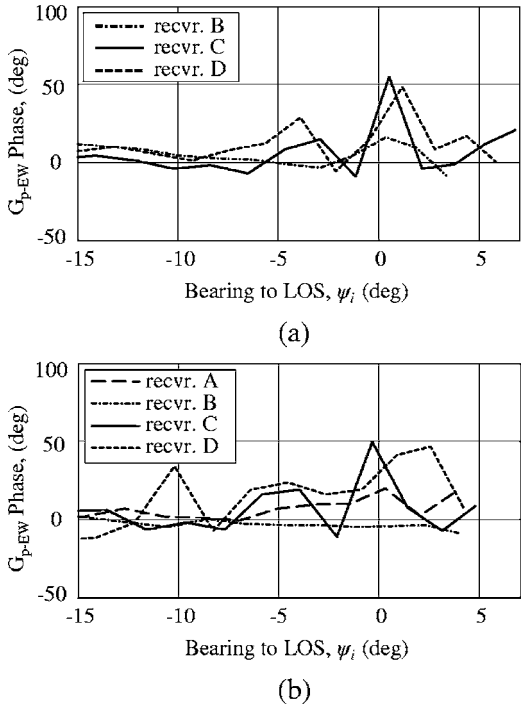


FIG. 4. Normalized phase for the transverse complex intensity component for all receivers averaged over frequency and plotted as a function of bearing angle, ψ_i. (a) Day 1 trials for all receivers which cross the LOS show significant and measurable shifts in phase of nearly 55° at ψ_i=0° as well as sidelobes around $\psi_i \approx \pm 5°$, as predicted, see Ref. 12. (b) Day 2 trials provide verification of these repeatable shifts in phase for several receivers. Bearing angles are approximated to account for wind and water currents.

for all of the sensors as they approach and cross over the direct line of sight, ψ_i=0°. Further, primary sidelobe detections are also evident at $\psi_i \approx \pm 5°$, as predicted. The experiment was duplicated the following day to confirm the previous day's analysis and verify repeatability. These results are illustrated in Fig. 4(b), which when compared to Fig. 4(a) show good agreement and indicate strong, repeatable phase shifts for both the direct LOS and the adjacent sidelobes.

For the Day 2 trials, an additional phase shift is apparent in receiver *D* at an angle of $\phi_D \approx -10°$, the result of a large external noise contribution during data acquisition for that location. We also note that the predicted shift in phase for receiver *B* is suspiciously absent. This is most likely due to the discrete position sampling used along the transverse path of the source. In this example, the direct LOS may have been inadvertently missed because of the severe weather conditions that existed on Day 2. Separate analysis of the time captured data stored by the Alesis data recorder can be used to verify this possibility.

Fourier transforms of the recorded data for sonobuoys *B* and *C* are calculated for a small region corresponding to the direct LOS for each. Receiver *C* is used to establish validity of the processing routines. Figure 5 shows the cross-spectral phase measured between the pressure channel and the E-W particle velocity channel of receiver *C* in much higher resolution than shown in Fig. 4. The progression of the receiver across the LOS reveals a uniform increase in phase as the receiver bearing approaches 0°. Likewise, it decreases as the

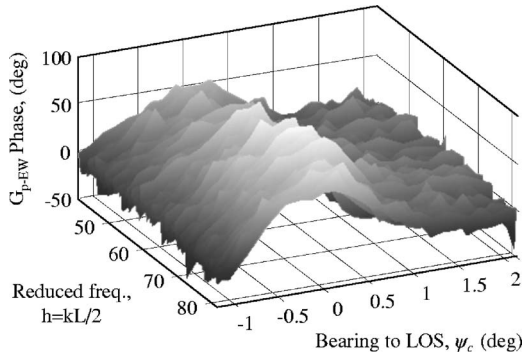


FIG. 5. High resolution view of the progression of receiver *C* across the line of sight as a function of the reduced frequency, $h=kL/2$, and the bearing angle to the LOS, ψ_C. A clear indication of the shifting phase is shown, with a peak shift of 55° at $\psi_C=0°$, corresponding well to results in Fig. 4.

receiver moves away from that line. The peak level of the measured cross-spectral phase is roughly 55°, which agrees well with Fig. 4(b).

A similar analysis can be completed for receiver *B* using data collected from the second day's trial to determine if discrete sampling was to blame for the missing phase shift at $\psi_B=0°$. From GPS data, it is determined that the sampled bearing angles for receiver *B* around the LOS include $[\ldots, -3.0°, -1.3°, 0.7°, 2.2°, \ldots]$. A higher resolution view of the transverse phase responses of buoys *B* and *C*, averaged over the sonobuoy frequency range, is shown in Fig. 6. Here it can be seen that the sampled locations at −1.3° and 0.7° do indeed span a region which exhibits a clear phase shift with a magnitude that agrees well with the shift measured in the previous day's trials. Thus, it is apparent that discrete sampling in combination with weather conditions was the most likely cause of the resulting phase measurements for receiver *B* in Fig. 4(b).

IV. CONCLUSIONS

The use of SSQ-53D sonobuoys as underwater acoustic intensity sensors has been demonstrated to detect the presence of a submerged body in a zone where the received scalar pressure signal is dominated by the incident blast from the source. In these experiments, performed in a deep water lake, the intensity receivers were located from 5 to 30 body lengths away from a large (relative to acoustic wavelength) scattering body. The source was driven by continuous broadband random noise; thus, all of the effects of lake reverberation are included in the findings.

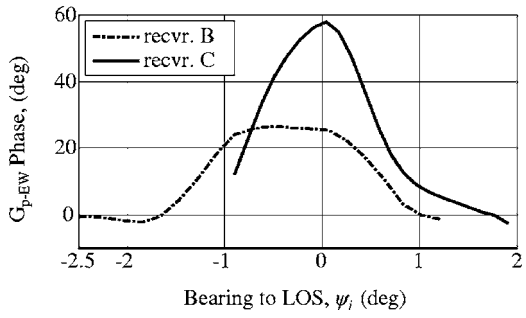


FIG. 6. High resolution view of the progression of receivers *B* and *C* across the line of sight, averaged over the operating frequency range of the SSQ-53D. The shifting phase in receiver *C* from Fig. 5 is shown, for reference. Similar processing of receiver *B* data also indicates phase shifts associated with the presence of the scatterer. The discrete sampling used in Fig. 4 at −1.3° and 0.7° resulted in an inadvertent missing of this phase shift.

The scalar pressure signal changed by less than 0.5 dB when the body passed through the line of sight between the source and receiver. However, the cross-spectral phase measured between the pressure and the acoustic particle velocity component in the direction that is perpendicular to the direction of incident wave propagation revealed significant variations (as much as 55°).

The observation that the largest effect is in the phase of the orthogonal velocity/pressure cross spectrum appears sensible. For a plane traveling wave, the relative phase, $\phi_{p\mathbf{u}}$, is zero. The scatterer makes small changes in the magnitude and direction of the field components. Compared to the incident direction, these changes are small in that the along-axis propagation is still primarily traveling. But perpendicular to the axis, there is no component in the absence of the scatterer. The scatterer perturbs the direction of the wave—to the right for parts of the wave that pass to the right of the scatterer and to the left for parts of the wave that pass to the left of the scatterer. The transverse components lead to standing-wave-like fields, but only in the transverse direction.

The measured results support previously developed and published theory that pressure gradients in the forward-scattered zone cause measurable perturbations in the intensity phase. The transverse component of intensity phase is therefore a viable indicator of the presence of a scattering object in the forward-scattered zone.

ACKNOWLEDGMENTS

We are grateful for the support of the Office of Naval Research, Code 321MS (Dr. James McEachern and Michael Wardlaw; Contract No. N00014-00-G-0058). Assistance and support were provided by the Naval Air Warfare Center, Aircraft Division, Patuxent River, MD; our sincere thanks are given to Frank Mitchell, Jason Ho, and Jason Payne. The effort given by the Naval Surface Warfare Center, Acoustics Research Detachment, Bayview, ID support team, under the direction of Steve Finley, is also sincerely appreciated. And finally, but by no means least appreciated, are the discussions, and help provided by Dr. Christopher Barber of ARL Penn State and Michael Higgins of RDA, Inc., Doylestown, PA.

[1]G. C. Lauchle, "Short-wavelength acoustic diffraction by prolate spheroids," J. Acoust. Soc. Am. **58**, 568–575 (1975).
[2]N. Willis, *Bistatic Radar* (Artech House, Boston, 1991).
[3]H. M. Nussenzveig, *Diffraction Effects in Semiclassical Scattering* (Cambridge University Press, Cambridge, Great Britain, 1992).
[4]A. Sarkissian, C. F. Guammond, and L. R. Dragonette, "T-matrix implementation of forward scattering from rigid structures," J. Acoust. Soc. Am. **94**, 3448–3453 (1993).
[5]B. Gillespie, K. Rolt, G. Edelson, R. Shaffer, and P. Hursky, "Littoral target forward scattering," in *Acoustic Imaging*, edited by S. Lees and L. A. Ferrair (Plenum Press, New York, 1997), Vol. **23**, pp. 501–506.
[6]G. S. Sammelmann, D. H. Trivett, and R. H. Hackman, "High-frequency scattering from rigid prolate spheroids," J. Acoust. Soc. Am. **83**, 46–54 (1998).
[7]P. Ratilal and N. C. Makris, "Extinction theorem for object scattering in a stratified medium," J. Acoust. Soc. Am. **110**, 2924–2945 (2001).

[8]H. Song, W. A. Kuperman, W. S. Hodgkiss, T. Akal, and P. Guerrini, "Demonstration of a high-frequency acoustic barrier with a time-reversal mirror," IEEE J. Ocean. Eng. **28**, 246–249 (2003).
[9]B. R. Rapids and G. C. Lauchle, "Processing of forward scattered fields with intensity sensors," in *Proceedings of Oceans 2002*, Biloxi, MS, 2002, pp. 1911–1914.
[10]B. R. Rapids and G. C. Lauchle, "Acoustic intensity measurements involving forward scatter from prolate spheroids," J. Acoust. Soc. Am. **116**, 2528 (2004).
[11]B. R. Rapids, "Acoustic intensity methods in classical scattering, Ph.D. thesis, The Pennsylvania State University (2004).
[12]B. R. Rapids and G. C. Lauchle, "Vector intensity field scattered by a rigid prolate spheroid," J. Acoust. Soc. Am. **120**, 38–48 (2006).
[13]F. J. Fahy, *Sound Intensity*, 2nd ed. (Elsevier, London, 1995).
[14]J. A. Mann, J. Tichy, and A. J. Romano, "Instantaneous and time-averaged energy transfer in acoustic fields," J. Acoust. Soc. Am. **82**, 17–30 (1987).
[15]P. J. Westervelt, "Acoustical impedance in terms of energy functions," J. Acoust. Soc. Am. **23**, 347–348 (1951).
[16]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic impedance, I. Theory, II. Experiment," J. Acoust. Soc. Am. **68**, 907–921 (1978).
[17]T. K. Stanton and R. T. Beyer, "Complex wattmeter measurements in a reactive acoustic field," J. Acoust. Soc. Am. **65**, 249–252 (1979).

Finite-bandwidth Kramers-Kronig relations for acoustic group velocity and attenuation derivative applied to encapsulated microbubble suspensions

Joel Mobley[a)]
Department of Physics and Astronomy, National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 9 September 2006; revised 9 January 2007; accepted 10 January 2007)

Kramers-Kronig (KK) analyses of experimental data are complicated by the conflict between the inherently bandlimited data and the requirement of KK integrals for a complete infinite spectrum of input information. For data exhibiting localized extrema, KK relations can provide accurate transforms over finite bandwidths due to the local-weighting properties of the KK kernel. Recently, acoustic KK relations have been derived for the determination of the group velocity (c_g) and the derivative of the attenuation coefficient (α') (components of the derivative of the acoustic complex wave number). These relations are applicable to bandlimited data exhibiting resonant features without extrapolation or unmeasured parameters. In contrast to twice-subtracted finite-bandwidth KK predictions for phase velocity and attenuation coefficient (components of the undifferentiated wave number), these more recently derived relations for c_g and α' provide stricter tests of causal consistency because the resulting shapes are invariant with respect to subtraction constants. The integrals in these relations can be formulated so that they only require the phase velocity and attenuation coefficient data without differentiation. Using experimental data from suspensions of encapsulated microbubbles, the finite-bandwidth KK predictions for c_g and α' are found to provide an accurate mapping of the primary wave number quantities onto their derivatives. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2535616]

PACS number(s): 43.35.Bf, 43.20.Hq [RR] Pages: 1916–1923

I. INTRODUCTION

Fundamentally rooted in causality, Kramers-Kronig (KK) relations provide linkages between the physical properties that govern the response of matter and materials to external stimuli. Due to their general foundations[1] KK relations have proven to be adaptable to, and applicable in, a wide array of tasks which include measuring fundamental material parameters, establishing the consistency of laboratory data, and building causally consistent physical models. One complication in adapting KK relations for the analysis of data is the knowledge gap that exists between the infinite bandwidth required by the KK integrals and the inherently bandlimited measurements. The impact of this gap on KK calculations depends on many factors, both general and system-dependent. In some approaches the gaps are filled in via extrapolation[2] while in others the potential influences of the unknown bands are reduced through subtractions.[2,3] One can also estimate analytically admissible approximations to a conjugate KK parameter given finite data.[4]

In previous work using a twice subtracted form, the consistency of finite-bandwidth acoustic KK relations between the ultrasonic attenuation coefficient $\alpha(\omega)$ and phase velocity $c_p(\omega)$ has been demonstrated using data from suspensions of encapsulated microbubbles that exhibit an isolated resonance.[5] Using only information from within the measurement spectrum, finite-bandwidth KK relations provided for accurate transformations between the two quantities, $\alpha(\omega)$ and $c_p(\omega)$, with the proper selection of the subtraction frequency. The choice of subtraction frequency ω_0 determines the slope and intercept of a linear contribution to the calculation that is critical to the accuracy of the prediction. In the ideal case, in which the entire spectrum is available for the calculation, the results are independent of the subtraction frequency. However, for the finite bandwidth case the result is highly sensitive to this choice of ω_0. Furthermore, it is not clear that objective criteria can be established to favor the best-fitting choices for ω_0 over other values. As a result, the subtraction frequency is effectively a tuning parameter which makes this approach somewhat unsatisfactory. In order to avoid this limitation, once-subtracted KK relations have been derived for the direct determination of the inverse group velocity $1/c_g(\omega)$ and frequency derivative of attenuation $\alpha'(\omega)=d\alpha(\omega)/d\omega$, which are the real and imaginary components of the first derivative (with respect to frequency) of the acoustic complex wave number. These relations were also shown to be consistent with data from microsphere suspensions exhibiting well-resolved resonant features.[6] For finite bandwidth analysis, an important general property of these KK relations for c_g and α' is that the shapes of the predicted curves are independent of the subtraction frequency. As a result these relations can serve as a more stringent test of causal consistency for dispersive acoustic data than the twice-subtracted KK relations for the primary quantities c_p and α. Another feature of this technique is that the KK integrals are formulated in terms of c_p and α, and thus do not

[a)]Electronic mail: jmobley@olemiss.edu

 0001-4966/2007/121(4)/1916/8/$23.00

require any differentiated quantities as inputs. In this work, the KK relations for c_g and α' are applied to encapsulated microbubble data, and the KK links in the data are clearly established without shape-altering factors. This is in contrast to the previous KK work predicting c_p and α in these systems, where the accuracies achieved were strongly dependent on the subtraction frequency.[5]

In the following section, the theoretical foundations of the relations are described. Next, a brief discussion of the encapsulated microbubble suspensions themselves and the methods used to determine the required quantities is given. Following that, the KK predictions for $1/c_g(\omega)$ and $\alpha'(\omega) = d\alpha(\omega)/d\omega$ are compared with the experimental data. In the discussion, the issues related to these results and methods are addressed, including expectations for their accuracy, comparisons with earlier KK analysis of similar data, and the applicability of these relations for nonresonant data. The consistency between the data and predictions for group velocity and derivative of attenuation under the KK relations described here clearly establishes the causal link between attenuation and dispersion for these data over a finite bandwidth without extrapolation or shape-altering factors.

II. THEORY

The transfer function for a passive, linear isotropic medium can be written

$$H(\omega,d) = \exp[iK(\omega)d], \tag{1}$$

where

$$K(\omega) = \omega/c_p(\omega) + i\alpha(\omega) \tag{2}$$

is the complex wave number, $\alpha(\omega)$ is the attenuation coefficient, $c_p(\omega)$ is the phase velocity, and d is the thickness. [A suspension of randomly dispersed spheres can be considered statistically homogeneous and isotropic; $K(\omega)$ represents the suspension's properties in the ensemble-averaged sense.[7]] The transfer function $H(\omega,d)$ is the Fourier transform of a causal, square-integrable function [i.e., the impulse response $h_d(t)$], which implies via Titchmarsh's theorem[8] that its real and imaginary parts form a Hilbert transform pair. Since $h_d(t)$ is real, the components of $H(\omega,d)$ exhibit definite parity, which in turn permits the mapping of the negative frequency components of the Hilbert integrals to positive frequencies. The resulting transforms are labeled as Kramers-Kronig relations. Using the method of subtractions, one can also derive KK relations for the components of $K(\omega)$. Based on both empirical and analytic evidence[5,6,9–11] two subtractions appear to be sufficient for establishing a Hilbert transform pair from the acoustic complex wave number. The twice subtracted relations in the expanded form are

$$\frac{\omega}{c_p(\omega)} = \frac{\omega_0}{c_p(\omega_0)} + (\omega-\omega_0)\frac{d}{d\omega}\frac{\omega}{c_p(\omega)}\bigg|_{\omega=\omega_0} + \lim_{\substack{\sigma\to 0\\ \Omega\to\infty}}\left[I_\alpha(\omega,\sigma,\Omega) - I_\alpha(\omega_0,\sigma,\Omega) - (\omega-\omega_0)\frac{d}{d\omega}I_\alpha(\omega,\sigma,\Omega)\bigg|_{\omega=\omega_0}\right], \tag{3}$$

where

$$I_\alpha(\omega,\sigma,\Omega) = \frac{1}{\pi}\int_\sigma^\Omega \frac{\alpha(x)-\alpha(\omega)}{x-\omega}dx - \frac{1}{\pi}\int_\sigma^\Omega \frac{\alpha(x)-\alpha(\omega)}{x+\omega}dx, \tag{4}$$

and

$$\alpha(\omega) = \alpha(\omega_0) + (\omega-\omega_0)\alpha'(\omega_0) + \lim_{\substack{\sigma\to 0\\ \Omega\to\infty}}\left[I_c(\omega,\sigma,\Omega) - I_c(\omega_0,\sigma,\Omega) - (\omega-\omega_0)\frac{d}{d\omega}I_c(\omega,\sigma,\Omega)\bigg|_{\omega=\omega_0}\right], \tag{5}$$

where

$$I_c(\omega,\sigma,\Omega) = -\frac{1}{\pi}\int_\sigma^\Omega \frac{x/c_p(x)-\omega/c_p(\omega)}{x-\omega}dx - \frac{1}{\pi}\int_\sigma^\Omega \frac{x/c_p(x)+\omega/c_p(\omega)}{x+\omega}dx, \tag{6}$$

and ω_0 is the subtraction frequency. By evaluating the integrals in Eqs. (3) and (5) before taking the limit of $\Omega\to\infty$, the divergences of the individual integrals will cancel. Although it is conventional to combine the remapped negative frequency contribution [the second integrals in Eqs. (4) and (6), respectively] with the positive frequency part, keeping them separate has some advantages for computational and analytical work. Note that the local variations in the quantities $\omega/c_p(\omega)$ and $\alpha(\omega)$ on the left-hand sides of Eqs. (3) and (5) are largely generated by the first terms [i.e., $I_\alpha(\omega,\sigma,\Omega)$ and $I_c(\omega,\sigma,\Omega)$] on the respective right-hand sides. The remaining terms in both relations define linear contributions whose slopes and intercepts are functions of ω_0.

The group velocity and the derivative of the attenuation coefficient are components of the differentiated complex wave number,

$$\frac{d}{d\omega}K(\omega) = \frac{d}{d\omega}\frac{\omega}{c_p(\omega)} + i\frac{d}{d\omega}\alpha(\omega) \tag{7a}$$

$$= 1/c_g(\omega) + i\alpha'(\omega), \tag{7b}$$

As detailed in previous work,[6] KK relations of the once-subtracted type for $1/c_g(\omega)$ and $\alpha'(\omega)$ can be derived. The group velocity relation is

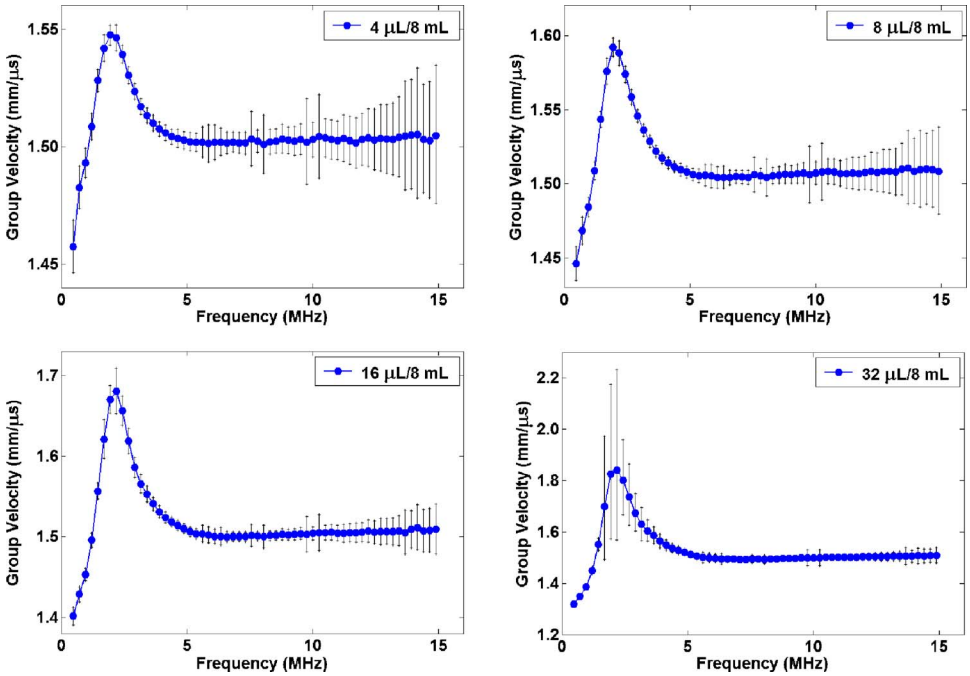


FIG. 1. (Color online) The experimentally determined group velocity data for the four encapsulated microbubble suspensions. The legend in each graph refers to the ratio of Albunex volume to the total suspension volume for each sample.

$$1/c_g(\omega) = 1/c_g(\omega_0) + \lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} [I_\alpha^{(1)}(\omega,\sigma,\Omega) - I_\alpha^{(1)}(\omega_0,\sigma,\Omega)], \tag{8}$$

where

$$I_\alpha^{(1)}(\omega,\sigma,\Omega) = \frac{1}{\pi} P \int_\sigma^\Omega \frac{\alpha(x)-\alpha(\omega)}{(x-\omega)^2} dx + \frac{1}{\pi}\int_\sigma^\Omega \frac{\alpha(x)-\alpha(\omega)}{(x+\omega)^2} dx. \tag{9}$$

As discussed in previous work,[6] in addition to the integral pair shown in Eq. (9) there are two more sets that can be similarly derived. The three sets are all equivalent in the limits $\sigma=0, \Omega\to\infty$, but are distinct in their predictions over finite integration intervals. As in the earlier work, the form shown in Eq. (9) is found to provide the most accurate predictions. For the derivative of the attenuation coefficient, the relation is

$$\alpha'(\omega) = \alpha'(\omega_0) + \lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} [I_c^{(1)}(\omega,\sigma,\Omega) - I_c^{(1)}(\omega_0,\sigma,\Omega)], \tag{10}$$

where

$$I_c^{(1)}(\omega,\sigma,\Omega) = -\frac{1}{\pi} P \int_\sigma^\Omega \frac{x(1/c_p(x)-1/c_p(\omega))}{(x-\omega)^2} dx + \frac{1}{\pi}\int_\sigma^\Omega \frac{x(1/c_p(x)-1/c_p(\omega))}{(x+\omega)^2} dx. \tag{11}$$

As discussed earlier, there are alternate forms for the integrals in Eq. (11), but the best accuracies are achieved with the integral pair shown here. These relations have two important features. First, the integrals themselves contain no differentiated quantities, that is, they are formulated purely in terms of the frequency variables and the primary quantities $c_p(\omega)$ and $\alpha(\omega)$. Second, and most significantly, the shapes of the $1/c_g(\omega)$ and $\alpha'(\omega)$ curves are completely independent of the choice of subtraction frequency ω_0.

III. EXPERIMENTAL METHODS

The data examined in this work are from transmission measurements of agitated suspensions of protein-encapsulated microbubbles diluted in saline. The encapsulated microbubbles are the primary content of the pharmaceutical product Albunex, which was one of the first generation of commercially available contrast agents for enhancing cardiac sonograms. Details of the laboratory methods and acquisition procedures for these experiments are provided in earlier publications.[12,13] The data presented here were calculated from the same set of rf waveforms used to determine published results for the phase velocity[13] and attenuation coefficient.[5] The sample suspensions consist of various volumes of Albunex (ranging from 4 to 32 μL) diluted into 8 mL of Isoton. Most of the encapsulated microbubble volume is provided by bubbles between 2 and 5 μm in diameter. In spite of their noted instability in ultrasonic fields, with proper laboratory procedures these suspensions have exhibited remarkedly stable and repeatable[14] behavior, with phase velocities, attenuation coefficients, and scattering properties that scale linearly over 5 octaves of volume concentration.[5,12,13] The group velocity and derivative of attenuation data, including error estimates, for the samples examined in this work are shown in Figs. 1 and 2. The original attenuation coefficient and phase velocity data used in the

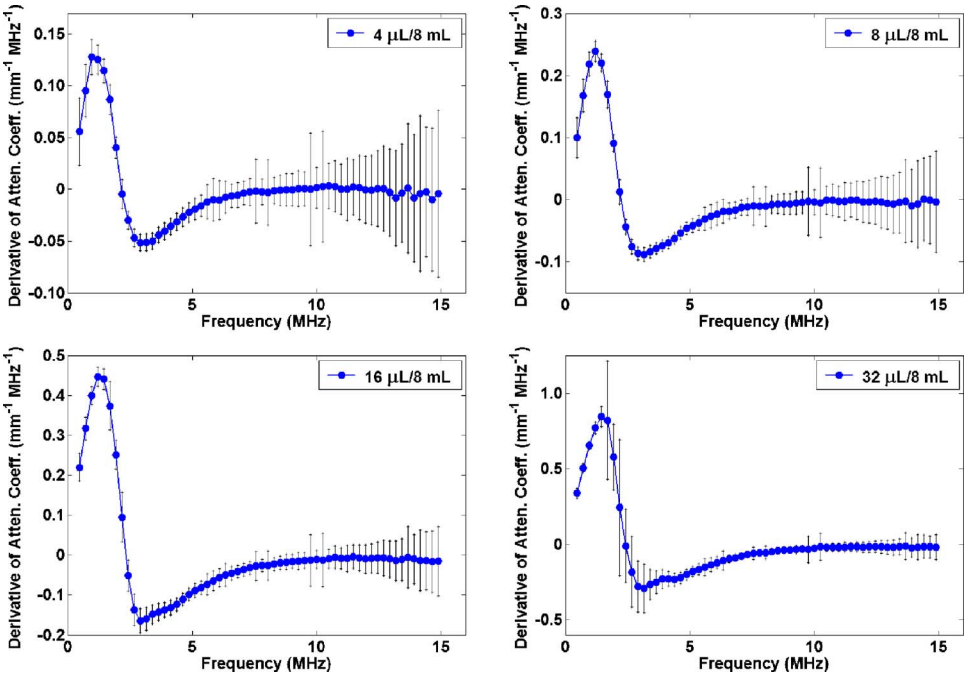


FIG. 2. (Color online) The experimentally determined derivative of the attenuation coefficient data for the four encapsulated microbubble suspensions. The legend in each graph refers to the ratio of Albunex volume to the total suspension volume for each sample.

KK calculations are shown in Fig. 3.

IV. RESULTS

The four panels of Fig. 4 show the finite-bandwidth KK predictions of $1/c_g(\nu)$, the inverse group velocity (i.e., the group slowness), using the subtraction frequency $\nu_0 = 10.5$ MHz for all four suspensions. (Note that $\nu = \omega/2\pi$.) The calculations were performed using discrete analogs of Eqs. (8) and (9), with the exception that the limits of integration are fixed to finite values. The integrals were approximated as Riemann sums over the spectrum from $\sigma/2\pi = 0.5$ MHz to $\Omega/2\pi = 15$ MHz with a spectral step size of approximately 0.244 MHz [i.e., $(4.096\ \mu s)^{-1}$]. In all four cases the agreement is readily apparent as the shapes track one another quite closely, although the global minimum in each data set exceeds that of the KK prediction.

In Fig. 5, the finite-bandwidth KK predictions for $\alpha'(\nu)$, the derivative of the attenuation coefficient, are plotted with the experimental data for the four suspensions. The calculations used discrete analogs of Eqs. (10) and (11) with the limits fixed to 0.5 and 15 MHz, respectively. As before, the integrals were performed as Riemann sums. The subtraction frequency used in all of these was $\nu_0 = 10.5$ MHz. The KK predictions track the shape of the data fairly well, although not to the same degree as the group slowness curves. It should also be noted that if all of the terms in Eqs. (10) and (11) containing the subtraction frequency are ignored, there is no practical difference in the level of agreement for $\alpha'(\nu)$ between the KK predictions and the data.

V. DISCUSSION

A. Expectation of finite-bandwidth artifacts

The KK relations used in this work demonstrate a novel method for applying causal methods to experimental data. The only underlying requirement for their use is that the data exhibit some localized resonance-related behavior in its spectral measurement window. The resonance-related structures in the microbubble response were well resolved in the data, although partly cut off at the low frequency end. To

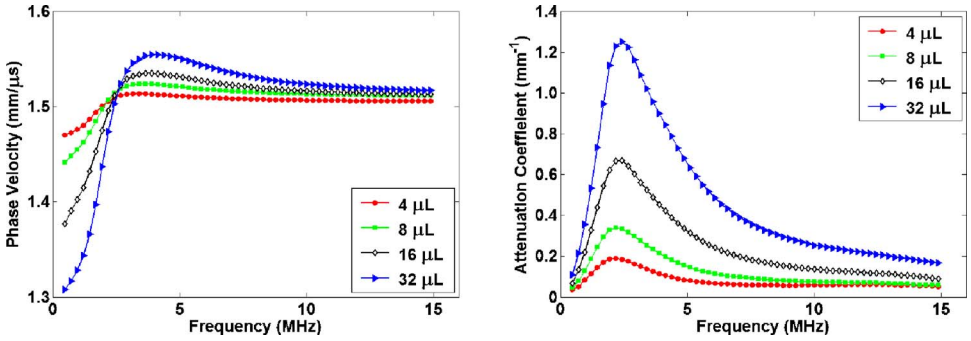


FIG. 3. (Color online) The experimentally determined phase velocity and attenuation coefficient data for the four encapsulated microbubble suspensions. These are the data used in the Kramers-Kronig integrals.

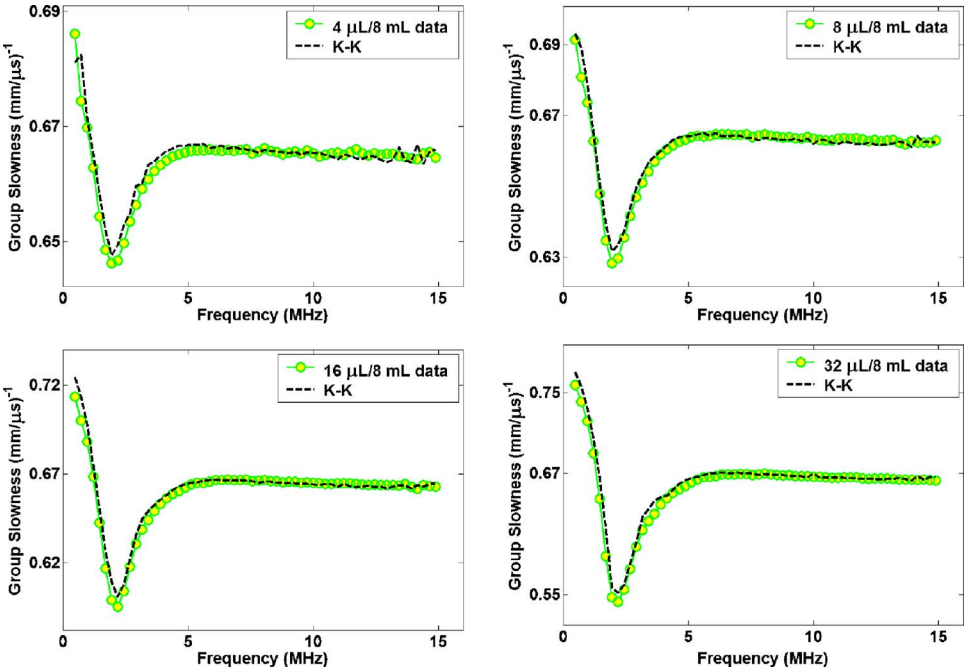


FIG. 4. (Color online) Comparisons of the Kramers-Kronig predictions and the experimental data for the group slowness (i.e., $1/c_g$) for the four suspensions.

understand the relationship between the resolution of resonant structures and the bandwidth limits of the data, a Lorenztian Hilbert transform pair has been utilized to provide some analytical understanding of the problem.[5] Details regarding this model and its applicability to the current problem are provided in the Appendix. The model has four inputs, the frequency of peak response ω_r, the damping constant Γ, and the high and low frequency limits of the data (Ω and σ). This model is not meant to describe the microbubble behavior, only to serve as a simple resonant type analog to aid in judging the nature and impact of finite bandwidth effects on KK calculations. A reasonable correspondence between the widths of the resonant structures is achieved when taking $\omega_r/2\pi=2.2$ MHz, and $\Gamma/2\pi=1.4$ MHz. To calculate the impact of the finite-bandwidth artifacts on the KK results using the model, we will consider the function $g'(\omega)$, and its finite bandwidth approximation $g'(\omega,\sigma,\Omega)$, whose behavior is analogous to that of the group velocity. To make a quantitative comparison, we will first

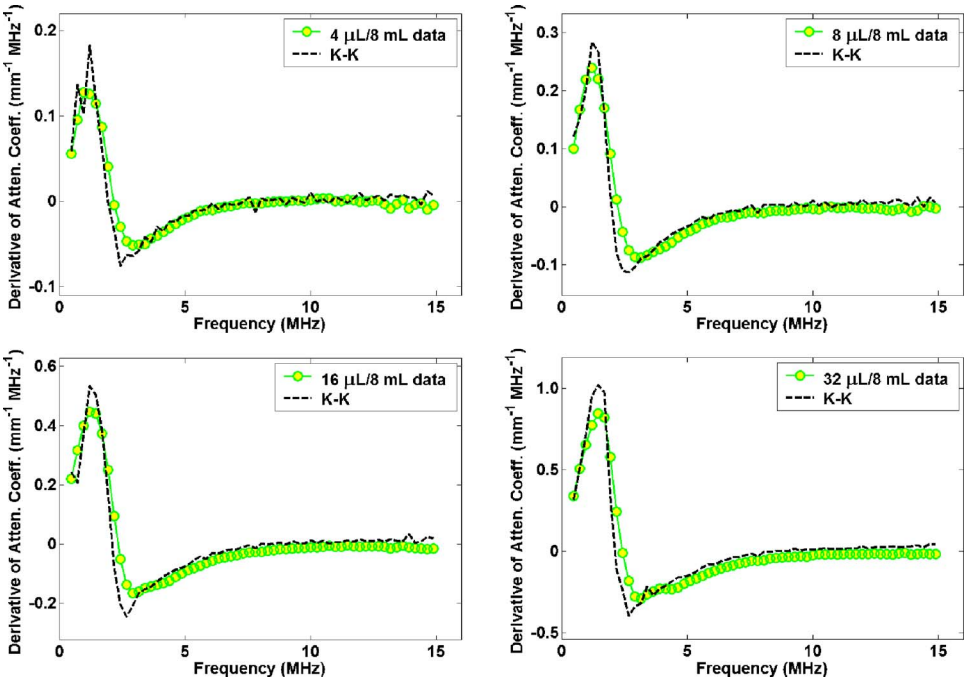


FIG. 5. (Color online) Comparisons of the Kramers-Kronig predictions and the experimental data for the frequency derivative of the attenuation coefficient for the four suspensions.

look at the difference in group slowness between the minimum value and at 8.06 MHz and look at the ratio of the KK prediction to the experimental data,

$$R_{\text{group}} = \frac{(1/c_g^{8\ \text{MHz}} - 1/c_g^{\max})_{\text{KK}}}{(1/c_g^{8\ \text{MHz}} - 1/c_g^{\max})_{\text{exp data}}}.$$

The analogous ratio from the model is given by

$$R_{\text{model}} = \frac{g'(\omega_2,\Omega,\sigma) - g'(\omega_1,\Omega,\sigma)}{g'(\omega_2) - g'(\omega_1)},$$

where $\omega_1/2\pi=2.2$ MHz, $\omega_2/2\pi=8$ MHz, $\Omega/2\pi=15$ MHz, and $\sigma/2\pi=0.5$ MHz. The model predicts that R_{model}, the ratio of the KK derived value to the actual value, will be 0.88. The ratio R_{group} for the microbubble data are 0.91, 0.88, 0.90, and 0.91 for the 4, 8, 16, and 32 μL cases, respectively. So the model suggests that the KK predictions are accurate to the levels one might expect based on our Lorentzian model. The model-based ratio improves to 0.932 if the bandwidth is extended to arbitrarily low frequencies. This would indicate that the ratio for the data might improve by about 6% for $\sigma\rightarrow 0$. Further improvements would require addition bandwidth at the high frequency end and the model indicates that the data would require at least an order of magnitude higher frequency limit to reach the 99% agreement level. For the attenuation derivative, the model versus data relationship is less quantitative, as the ratio of the KK predictions to the data for the ratio

$$\frac{(\alpha'_{\max} - \alpha'_{\min})_{\text{KK}}}{(\alpha'_{\max} - \alpha'_{\min})_{\text{exp data}}}$$

is greater than one and the corresponding model ratio is less than unity. This is likely due to the fact that the model function $g(\omega)$ is not as good a match to the phase velocity as $f(\omega)$ is to the attenuation coefficient, especially at the high frequency end of the spectrum.

B. KK relations for α and c_p, and the role of the subtraction frequency

In the earlier KK work with Albunex microbubble suspensions, the phase velocity and attenuation coefficient data have been shown to be consistent with finite-bandwidth KK relations [Eqs. (3) and (5)].[5] These twice-subtracted relations have also accurately transformed between $c_p(\nu)$ and $\alpha(\nu)$ for data from polymer microsphere suspensions.[11] However, for both the Albunex and microsphere suspensions the accuracies of the $c_p(\nu)$ and $\alpha(\nu)$ predictions depend critically on the choice of the subtraction frequency $\nu_0=\omega_0/2\pi$ used in the calculations, which controls the slope and offset of a linear factor in these relations. In the ideal case of $\sigma=0$ and $\Omega\rightarrow\infty$ (i.e., infinite bandwidth), the outcome should be independent of this choice. About one-fifth of the frequencies contained in the discrete data sets can produce reasonably good agreement when used as ν_0's, while a similar fraction produces strongly divergent results. To date, an objective physical justification for choosing ν_0 in the $c_p(\nu)$ and $\alpha(\nu)$ predictions has not been identified. In contrast, the high levels of agreement between KK predictions and data for both $c_g(\nu)$ and $\alpha'(\nu)$ clearly demonstrate the causal link in the Albunex microbubble data in a more satisfying manner, since the KK calculations produce shape-invariant predictions of the two quantities. The linear factors inherent to the original relations are essentially differentiated away in the process of deriving the relations for c_g and α'. As a result, the ν_0 choice only affects the offset of the c_g and α' predictions and not their shapes. Still, it should be noted that the twice subtracted relation predictions for $c_p(\nu)$ and $\alpha(\nu)$ [Eqs. (3)–(6)] use only parameters from within the measured data, and thus the fact that it works at all is suggestive of their correctness, in spite of the ω_0 issue. Also, the fact that the relations in this work are derived from the twice-subtracted relations also stands in their favor.

C. Power-law systems

An important class of nonresonant dispersive behavior in the megahertz ultrasound range is associated with power-law attenuation coefficients[15] of the form $\alpha(\omega)=\alpha_0\omega^y$, where $1\leqslant y<2$. Although the finite-bandwidth approach works for data exhibiting resonant structures, it is not as well suited for the power-law systems whose attenuation coefficient and dispersion behavior (usually globally varying as ω^{y-1}) are monotonic.[9] When attenuation/dispersion data do exhibit strong local variation (e.g., resonant structures), the KK integrals can rise above the finite bandwidth artifact and deliver locally accurate results due to the weighting of the locally singular kernels $1/(x-\omega)$ and $1/(x-\omega)^2$. However, in the power-law case, there is no local variation to effectively anchor the KK predictions, and wide bandwidths are the only defense against finite bandwidth artifacts. In earlier work,[9] it was shown that power-law systems can require extremely wide bandwidths to achieve accurate predictions of the power-law exponent of the attenuation coefficient. The situation is not greatly improved by moving to the relations for $1/c_g$, and α'. For the monotonically changing dispersion and attenuation behavior of power-law systems, the adherence of the data to the causally consistent power-law model remains the best indication that the results are causal. The converse is not necessarily true however; finite bandwidth data that only fit one component of the power-law complex wave number cannot be invalidated as acausal since the behavior of the system in the unknown spectral regions may deviate from the power-law type, and in the KK view, this out-of-band structure can influence the in-band result in an unknown manner.

VI. CONCLUSION

In this work, KK relations for determining the group velocity and frequency derivative of attenuation have been applied to data from encapsulated microbubble suspensions. These KK predictions yielded accurate predictions for the experimentally measured quantities without the shape-altering subtraction constants inherent in the twice-subtracted methods for determining phase velocity and attenuation. These KK relations for $c_g(\omega)$ and $\alpha'(\omega)$ are formulated in terms of the undifferentiated quantities $\alpha(\omega)$ and $c_p(\omega)$ so no derivative quantities are required as inputs.

This type of analysis should be useful for any data set exhibiting localized resonant-like structures, and also provides an alternative method for determining the group velocity curve. The success of these methods also lends validity to the twice subtracted KK approach for determining $\alpha(\omega)$ and $c_p(\omega)$, which is the basis for the derivation of the $c_g(\omega)$ and $\alpha'(\omega)$ relations.

ACKNOWLEDGMENTS

This research is an outgrowth of the body of work addressing Kramers-Kronig relations in acoustics by Professor James G. Miller of Washington University. I also wish to acknowledge former associates of Dr. Miller's group who were responsible for the Albunex data used in this work: Michael S. Hughes, Jon N. Marsh, Christopher S. Hall, and Gary H. Brandenburger.

APPENDIX

In previous work,[5] a Hilbert transform pair was introduced as a model for understanding the effects limited bandwidths can have on Kramers-Kronig predictions. The two model functions $f(\omega)$ and $g(\omega)$ are the real and imaginary components of a Lorentzian line-spectrum,

$$f(\omega)=\frac{1}{1+(\omega-\omega_r)^2/\Gamma^2}+\frac{1}{1+(\omega+\omega_r)^2/\Gamma^2}, \tag{A1}$$

$$g(\omega)=\frac{(\omega-\omega_r)/\Gamma}{1+(\omega-\omega_r)^2/\Gamma^2}+\frac{(\omega+\omega_r)/\Gamma}{1+(\omega+\omega_r)^2/\Gamma^2}. \tag{A2}$$

These are the real and imaginary parts of the Fourier transform of the real causal function

$$h(t)=\begin{cases}\frac{2}{\Gamma}\exp(-\Gamma t)\cos(\omega_r t), & t\geqslant 0\\ 0, & t<0,\end{cases} \tag{A3}$$

a damped harmonic oscillator whose natural resonance frequency ω_{res} is related to the given quantities by $\omega_r^2=\omega_{\text{res}}^2-\Gamma^2$. The only assumption we make is that $\omega_{\text{res}}>\Gamma$ (i.e., the oscillator is underdamped). Since the two functions (A1) and (A2) are the transforms of a causal function, they are Hilbert transforms of one another. One can then also formulate integral transforms that predict their derivatives

$$g'(\omega)=-\lim_{\substack{\Omega\to\infty\\ \sigma\to 0}}\left[\frac{1}{\pi}P\int_\sigma^\Omega\frac{f(x)-f(\omega)}{(x-\omega)^2}dx+\frac{1}{\pi}\int_\sigma^\Omega\frac{f(x)-f(\omega)}{(x+\omega)^2}dx\right] \tag{A4}$$

and

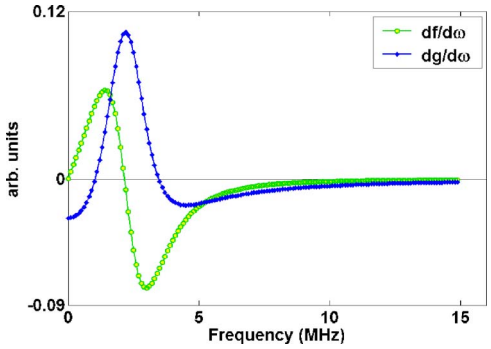


FIG. 6. (Color online) The two derivative components of the damped resonator model relevant in estimating finite-bandwidth artifacts of the bandlimited Kramers-Kronig integrals. The $g'(\omega)$ term behaves similarly to the group velocity while $f'(\omega)$ roughly corresponds with the attenuation derivative.

$$f'(\omega)=\lim_{\substack{\Omega\to\infty\\ \sigma\to 0}}\left[\frac{1}{\pi}P\int_\sigma^\Omega\frac{g(x)-g(\omega)}{(x-\omega)^2}dx-\frac{1}{\pi}\int_\sigma^\Omega\frac{g(x)+g(\omega)}{(x+\omega)^2}dx\right], \tag{A5}$$

where

$$g'(\omega)=\frac{1}{\Gamma}\left(\frac{1-(\omega-\omega_r)^2/\Gamma^2}{(1+(\omega-\omega_r)^2/\Gamma^2)^2}+\frac{1-(\omega+\omega_r)^2/\Gamma^2}{(1+(\omega+\omega_r)^2/\Gamma^2)^2}\right) \tag{A6}$$

and

$$f'(\omega)=-\frac{2}{\Gamma}\left(\frac{(\omega-\omega_r)/\Gamma}{(1+(\omega-\omega_r)^2/\Gamma^2)^2}+\frac{(\omega+\omega_r)/\Gamma}{(1+(\omega+\omega_r)^2/\Gamma^2)^2}\right). \tag{A7}$$

The $g'(\omega)$ term is analogous to the group velocity and the $f'(\omega)$ term is most comparable to the attenuation derivative. These two functions are shown in Fig. 6.

The KK integrals in Eqs. (A4) and (A5) have the following solutions:

$$g'(\omega,\Omega,\sigma)=g'(\omega)\theta(\Omega,\sigma,\omega_r,\Gamma)-f_s'(\omega)\varphi(\Omega,\sigma,\omega_r,\Gamma)+f'(\omega)\xi(\Omega,\sigma,\omega) \tag{A8}$$

and

$$f'(\omega,\Omega,\sigma)=f'(\omega)\theta(\Omega,\sigma,\omega_r,\Gamma)+g_s'(\omega)\varphi(\Omega,\sigma,\omega_r,\Gamma)+g'(\omega)\xi(\Omega,\sigma,\omega), \tag{A9}$$

where

$$\theta=\frac{1}{\pi}\left[\arctan\left(\frac{\Omega-\omega_r}{\Gamma}\right)+\arctan\left(\frac{\Omega+\omega_r}{\Gamma}\right)-\arctan\left(\frac{\sigma-\omega_r}{\Gamma}\right)-\arctan\left(\frac{\sigma+\omega_r}{\Gamma}\right)\right], \tag{A10}$$

$$\varphi=\frac{1}{2\pi}\ln\left(\frac{\Gamma^2+(\Omega+\omega_r)^2}{\Gamma^2+(\Omega-\omega_r)^2}\frac{\Gamma^2+(\sigma-\omega_r)^2}{\Gamma^2+(\sigma+\omega_r)^2}\right), \tag{A11}$$

$$\xi = \frac{1}{\pi} \ln\left(\frac{\Omega+\omega}{\Omega-\omega}\frac{\omega-\sigma}{\omega+\sigma}\right), \tag{A12}$$

$$f_s'(\omega) = -\frac{2}{\Gamma}\left(\frac{(\omega-\omega_r)/\Gamma}{(1+(\omega-\omega_r)^2/\Gamma^2)^2} - \frac{(\omega+\omega_r)/\Gamma}{(1+(\omega+\omega_r)^2/\Gamma^2)^2}\right), \tag{A13}$$

$$g_s'(\omega) = \frac{1}{\Gamma}\left(\frac{1-(\omega-\omega_r)^2/\Gamma^2}{(1+(\omega-\omega_r)^2/\Gamma^2)^2} - \frac{1-(\omega+\omega_r)^2/\Gamma^2}{(1+(\omega+\omega_r)^2/\Gamma^2)^2}\right). \tag{A14}$$

The exact functions and the finite bandwidth results are related by the following limits:

$$g'(\omega) = \lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} g'(\omega,\Omega,\sigma) \tag{A15}$$

and

$$f'(\omega) = \lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} f'(\omega,\Omega,\sigma). \tag{A16}$$

In terms of the functions defined in Eqs. (A10)–(A12), these limits are $\lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} \theta=1$, $\lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} \varphi=0$, and $\lim_{\substack{\Omega\to\infty \\ \sigma\to 0}} \xi=0$.

For the range of parameters relevant for this work ($\omega_r/2\pi=2.2$ MHz, $\Gamma/2\pi=1.4$ MHz, $\Omega/2\pi=15$ MHz, $\sigma/2\pi=1$ MHz) the finite bandwidth expressions are dominated by the terms inside the parentheses of Eq. (A6), (A7), (A13), and (A14). Thus, over a finite interval, the two predictions fall short of their full bandwidth prediction by the factor $\theta(\Omega,\sigma,\omega_r,\Gamma)$. The ξ term can be eliminated from the finite bandwidth result by the inclusion of the terms $-(x-\omega)f'(\omega)$ and $-(x-\omega)g'(\omega)$ in the numerators of the integrands in Eqs. (A4) and (A5), respectively. The terms that involve ξ are negligible everywhere except right on the boundary frequencies where they diverge as $\omega\to\Omega$, or $\omega\to\sigma$.

[1] J. S. Toll, "Causality and the dispersion relation: Logical foundations," Phys. Rev. **104**, 1760–1770 (1956).

[2] V. Lucarini, F. Bassani, K. E. Peiponen, and J. J. Saarinen, "Dispersion theory and sum rules in linear and nonlinear optics," Riv. Nuovo Cimento **26**, 1–120 (2003).

[3] K. F. Palmer, M. Z. Williams, and B. A. Budde, "Multiply subtractive Kramers-Kronig analysis of optical data," Appl. Opt. **37**, 2660–2673 (1998).

[4] G. W. Milton, D. J. Eyre, and J. V. Mantese, "Finite frequency range Kramers Kronig relations: Bounds on the dispersion," Phys. Rev. Lett. **79**, 3062–3065 (1997).

[5] J. Mobley, K. R. Waters, M. S. Hughes, C. S. Hall, J. N. Marsh, G. H. Brandenburger, and J. G. Miller, "Kramers-Kronig relations applied to finite bandwidth data from suspensions of encapsulated microbubbles," J. Acoust. Soc. Am. **108**, 2091–2106 (2000); **112**, 760–761 (2002).

[6] J. Mobley, K. R. Waters, and J. G. Miller, "Causal determination of acoustic group velocity and frequency derivative of attenuation with finite-bandwidth Kramers-Kronig relations," Phys. Rev. E **72**, 016604 (2005).

[7] R. L. Weaver and Y. H. Pao, "Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media," J. Math. Phys. **22**, 1909–1918 (1981).

[8] H. M. Nussenveig, *Causality and Dispersion Relations* (Academic, New York, 1972).

[9] J. Mobley, K. R. Waters, and J. G. Miller, "Finite-bandwidth effects on the causal prediction of ultrasonic attenuation of the power-law form," J. Acoust. Soc. Am. **114**, 2782–2790 (2003).

[10] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers-Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).

[11] K. R. Waters, J. Mobley, and J. G. Miller, "Causality-imposed (Kramers-Kronig) relationships between attenuation and dispersion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 822–833 (2005).

[12] J. N. Marsh, C. S. Hall, M. S. Hughes, J. Mobley, J. G. Miller, and G. H. Brandenburger, "Broadband through-transmission signal loss measurements of Albunex suspensions at concentrations approaching in vivo doses," J. Acoust. Soc. Am. **101**, 1155–1161 (1997).

[13] J. Mobley, J. N. Marsh, C. S. Hall, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "Broadband measurements of phase velocity in Albunex suspensions," J. Acoust. Soc. Am. **103**, 2145–2153 (1998).

[14] N. DeJong, L. Hoff, T. Skotland, N. Bom, "Absorption and scatter of encapsulated gas filled microspheres—Theoretical considerations and some measurements," Ultrasonics **30**, 95–103 (1992).

[15] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).

Ultrasonic transient bounded-beam propagation in a solid cylinder waveguide embedded in a solid medium

Laurent Laguerre[a] and Anne Grimault
Section Reconnaissance et Géophysique, Laboratoire Central des Ponts et Chaussées, Centre de Nantes, Route de Bouaye, BP 4129, 44341 Bouguenais Cedex, France

Marc Deschamps
Laboratoire de Mécanique Physique de Bordeaux, UMR CNRS 5469, Université de Bordeaux I, 351 Cours de la Libération, 33405 Talence Cedex, France

(Received 21 December 2005; revised 20 October 2006; accepted 11 January 2007)

A semianalytical solution alternative and complementary to modal technique is presented to predict and interpret the ultrasonic pulsed-bounded-beam propagation in a solid cylinder embedded in a solid matrix. The spectral response to an inside axisymmetric velocity source of longitudinal and transversal cylindrical waves is derived from Debye series expansion of the embedded cylinder generalized cylindrical reflection/transmission coefficients. So, the transient guided wave response, synthesized by inverse double Fourier-Bessel transform, is expressed as a combination of the infinite medium contribution, longitudinal, transversal, and coupled longitudinal and transversal waveguide sidewall interactions. Simulated $(f, 1/\lambda_z)$ diagrams show the influence of the number of waveguide sidewall interactions to progressively recover dispersion curves. Besides, they show the embedding material filters specific signal portions by concentrating the propagating signal in regions where phase velocity is closer to phase velocity in steel. Then, simulated time waveforms using broadband high-frequency excitation show that signal leading portions exhibit a similar periodical pattern, for both free and embedded waveguides. Debye series-based interpretation shows that double longitudinal/transversal and transversal/longitudinal conversions govern the time waveform leading portion as well as the radiation attenuation in the surrounding cement grout. Finally, a methodology is deduced to minimize the radiation attenuation for the long-range inspection of embedded cylinders. [DOI: 10.1121/1.2536759]

PACS number(s): 43.35.Cg, 43.20.Gp, 43.20.Mv, 43.20.Bi [PEB] Pages: 1924–1934

I. INTRODUCTION

Wave propagation in free elastic cylinders has been studied extensively since the pioneering works of Pochhammer (1876) and Chree (1889). Some of the major results for free and clad cylinders are reviewed and discussed in Thurston (1978) and Pavlakovic *et al.* (2001), for instance. The classical way of solving the waveguide propagation problem is first by identifying the dispersion curves of transverse modes according to an imposed longitudinal guiding direction of propagation. To do so, the roots of the characteristic dispersion equation (the determinant of the global matrix of the whole stress/strain boundary conditions at layer interfaces, Lowe, 1995) have to be found in terms of frequency and wave number. Once these eigenvalues are known, the eigenvectors (modes) can then be found. Finally, decomposition of the source on the eigenvector basis leads to modes amplitudes (Puckett and Peterson, 2005).

However, finding roots of the dispersion equation is usually a difficult, time-consuming task, which necessitates robust search algorithms, especially because of large magnitude variations in the absolute value of the dispersion between roots, or existence of very close modes, as well as difficulties in deriving complex Bessel functions, for leaky system cases (Lowe, 1995; Pavlakovic *et al.*, 2001; Barshinger *et al.*, 2002).

Thus, an alternative semianalytical formulation is considered herein which does not rely on the use of mode propagation and must be thought of as a complementary approach to the above-mentioned one. Indeed, the purpose of this paper is to first derive the axisymmetric elastic isotropic spectral response of a solid cylinder embedded in an infinite solid matrix to an incident spectrum of cylindrical waves coming from the inner cylinder. Particularly, the generalized cylindrical reflection and transmission coefficients of the embedded cylinder are expanded, for longitudinal and transversal potentials, as Debye series of local reflection and/or transmission coefficients at the solid/solid interface, respectively. Debye series expansion which allow for the decomposition of a global physical process into a series of local interactions was used for fluid loaded plate problems (Deschamps and Chengwei, 1991; Deschamps and Hosten, 1992) for either isotropic elastic or viscoelastic anisotropic materials. Danila *et al.* (1995) extend the formalism to fluid-fluid concentric (or not) interfaces. Fluid-loaded layered-plate or -cylinder have been adressed by Zeroug and Felsen (1994, 1995) with special emphasis on nonspecular beam reflection. The case of the guided wave propagating in a cylinder embedded in solid matrix arising from an inside bounded beam and devel-

[a]Electronic mail: laurent.laguerre@lcpc.fr

 0001-4966/2007/121(4)/1924/11/$23.00

oped in this paper was not, to the author's knowledge, yet addressed. It is an extension to the previous works of Danthez and Deschamps (1989) dealing with the spatiotemporal response of a free cylinder to a single longitudinal excitation.

Then, from the spectral response, the spatiotemporal velocity field is derived at any location within the embedded cylinder from an inverse double Fourier-Bessel transform over time and space frequencies. The Bessel integration over all radial frequencies is thus performed for longitudinal (L) and transversal (T) cylindrical waves in the r direction whose amplitudes are weighted both by the spatial initial field spectra and cylindrical reflection or transmission coefficients under Debye series approximation, with plane-wave dependence in the z direction. Analog integral approaches have been used in geophysical domain especially in the acoustics of fluid-filled boreholes (Roever *et al.*, 1971; Chen and Toksöz, 1981) to retrieve surrounding material (possibly homogeneous or cylindrically stratified) characteristics from synthetic seismograms responses to a point source excitation at different receiver positions (all located in the inner fluid). For a similar configuration, Rao and Vandiver (1999) derived trapped modes from modal solutions.

In this problem, the "source" is circular with axial and radial velocity distributions initiated at $t=0$ and $z=0$, first within the infinite solid in Sec. II, then within the inner solid cylinder in Sec. III. Only the axisymmetric guided wave is considered for the inner waveguide and surrounding solid medium (i.e., Rayleigh and Stoneley waves are beyond the scope of the proposed model and thus not considered here).

The final synthesized spatiotemporal signal is then expressed as a combination infinite medium contribution and delayed longitudinal, transversal, coupled longitudinal/transversal, and transversal/longitudinal wave arrivals.

As an example of the application of the proposed "pulsed bounded-beam propagation" model for the development of nondestructive inspection methodologies for embedded cylinders, we focus on the axisymmetric guided propagation in a steel cylindrical bar embedded in an infinite grout matrix arising from initial field applied within the steel.

First, solutions are compared to dispersion curves to show how the modes can be progressively described by increasing the number of waveguide sidewall interactions. Then we show that the time response to a broadband high-frequency excitation gives similar results for the free and embedded waveguides in the leading portion of the signal. We observe as well that double longitudinal-transversal and transversal to longitudinal conversions govern this leading portion as well as the radiation attenuation in the surrounding cement grout. We show that the radiation attenuation for the embedded cylinder first filters out some specific signal portions of simulated time waveforms (or time to spatial frequencies diagrams) which is quite similar to the filtering due to Debye series truncation to the first terms for the free-cylinder case (even no leakage occurs in that case); second, we show that radiation attenuation levels are reduced for phase velocities closer to longitudinal phase velocity in steel. According to these simulations, a methodology is proposed to minimize the radiation attenuation.

II. PROBLEM STATEMENT AND SPECTRAL INTEGRAL REPRESENTATION

Before dealing with the source radiating within the embedded cylinder, this section presents the general formulation of the axisymmetric source radiating in an infinite solid medium and spectral transforms associated with the problem.

The equations governing the axisymmetric velocity potentials propagating in a linear elastic isotropic medium including a source s centered around the z axis and lying in the transversal plane at $z=0$ are

$$\begin{cases} \rho\dfrac{\partial^2\varphi(r,z,t)}{\partial t^2} = (\lambda+2\mu)\left(\dfrac{\partial^2}{\partial r^2}+\dfrac{1}{r}\dfrac{\partial}{\partial r}+\dfrac{\partial^2}{\partial z^2}\right)\varphi(r,z,t) \\ \rho\dfrac{\partial^2\psi(r,z,t)}{\partial t^2} = \mu\left(\dfrac{\partial^2}{\partial r^2}+\dfrac{\partial}{\partial r}\dfrac{1}{r}+\dfrac{\partial^2}{\partial z^2}\right)\psi(r,z,t) \\ \dfrac{\partial\varphi(r,z,t)}{\partial r}-\dfrac{\partial\psi(r,z,t)}{\partial z} = s_r(r,t)\big|_{z=0} \\ \dfrac{\partial\varphi(r,z,t)}{\partial z}+\left(\dfrac{\partial}{\partial r}+\dfrac{1}{r}\right)\psi(r,z,t) = s_z(r,t)\big|_{z=0} \end{cases}, \tag{1}$$

where ρ is the density and λ and μ are the Lamé constants.

The first two uncoupled equations are the classical scalar wave equations in the source-free medium, for the velocity scalar potential φ and the velocity vector potential ψ, respectively (with ψ the single orthoradial vector potential component due to cylindrical symmetry).

The last two coupled equations are the initial conditions on the velocity field.

Applying the respective double-integral Fourier-Bessel inverse transforms,

$$\Im_{0,p}^{-}[(\cdot)] = \int_{-\infty}^{+\infty}\int_{0}^{+\infty}(\cdot)J_0(2\pi pr)2\pi p\exp(j2\pi ft)dp\,df, \tag{2}$$

to scalar potential,

$$\varphi(r,z,t) = \Im_{0,p}^{-}[\Phi(p,z,f)], \tag{3}$$

and

$$\Im_{1,p}^{-}[(\cdot)] = \int_{-\infty}^{+\infty}\int_{0}^{+\infty}(\cdot)J_1(2\pi pr)2\pi p\exp(j2\pi ft)dp\,df, \tag{4}$$

to vector potential,

$$\psi(r,z,t) = \Im_{1,p}^{-}[\Psi(p,z,f)], \tag{5}$$

allows one to transform the space- and time-domain partial differential equations system (1) into an ordinary differential equations system in the (p,f) spectral domain (where p and f are the radial and temporal frequencies, respectively).

The general spectral solutions in the source-free medium are of the form

$$\Phi(p,z,f) = \Phi_0(p,f)\exp(-j2\pi w_l z), \tag{6}$$

for the scalar velocity potential, where the longitudinal axial frequency w_l is

$$w_l^2 = \left(\frac{f}{c_l}\right)^2 - p^2, \text{ with}$$

$$c_l = \sqrt{\frac{(\lambda + 2\mu)}{\rho}}, \quad \text{the longitudinal phase velocity}$$

and

$$\Psi(p,z,f) = \Psi_0(p,f)\exp(-j2\pi w_t z), \tag{7}$$

for the vector velocity potential, where the transversal axial frequency

w_t is

$$w_t^2 = \left(\frac{f}{c_t}\right)^2 - p^2, \text{ with}$$

$$c_t = \sqrt{\frac{\mu}{\rho}}, \quad \text{the transversal phase velocity.}$$

Finally, the expressions of the spatiotemporal velocity field for an axisymmetric bounded beam propagating in an infinite elastic isotropic medium are

$$v_r(r,z,t) = \Im_{1,p}^{-}\{2\pi E(f)[-p\Phi_0(p,f)\exp(-j2\pi w_l z) + jw_t\Psi_0(p,f)\exp(-j2\pi w_t z)]\},$$

$$v_z(r,z,t) = \Im_{0,p}^{-}\{2\pi E(f)[-jw_l\Phi_0(p,f)\exp(-j2\pi w_l z) + p\Psi_0(p,f)\exp(-j2\pi w_t z)]\}, \tag{8}$$

with $E(f)$ the time source spectrum.

Both radial and axial velocity components express as two contributions, a longitudinal one and a transversal one. Each contribution is thus expanded as an integral summation of cylindrical waves in the plane transversal to the z direction of propagation and plane and progressive waves in the z direction of propagation over all radial frequencies p. It is worth noting that the radial derivatives modify the transform kernel while the axial derivatives do not, leading hence to the first-order Bessel function dependence of the radial component and the zero-order Bessel function dependence of the axial component.

Finally, the Φ_0 and Ψ_0 amplitudes are determined from the knowledge of the velocity spatial spectra of the source $S_r(p,f)$ and $S_z(p,f)$ [these are derived from Bessel transforms of the radial and axial initial velocity profiles, $s_r(r)|_{z=0}$ and $s_z(r)|_{z=0}$, respectively].

Hence, the initial scalar and vector velocity potentials amplitudes are, respectively,

$$\Phi_0(p,f) = D^{-1}[pS_r(p,f) + jw_T S_z(p,f)]$$

$$\Psi_0(p,f) = D^{-1}[-jw_l S_r(p,f) + pS_z(p,f)], \tag{9}$$

where $D=2\pi(p^2+w_l w_t)$ is the determinant of the 2×2 system derived from the last two transformed equations in system (1).

Now, the aim of Sec. III is to derive the spatiotemporal velocity field in a cylindrical waveguide perfectly embedded in an infinite solid matrix and generated by an inside bounded beam.

III. SPATIOTEMPORAL VELOCITY FIELD IN AN EMBEDDED CYLINDRICAL WAVEGUIDE

A. Model assumptions

As mentioned in Sec. II, the velocity field at space and time origins expresses in the radial direction in terms of standing cylindrical waves, namely Bessel functions of zero-order J_0 and first-order J_1 (according to axial or radial component, respectively). To perform the study of the transient bounded beam propagation into the guide, i.e., the interaction of a short duration with limited spatial extension wave with the waveguide sidewalls, we need to express Bessel functions in terms of propagating waves. Thus, the Bessel function of order n is decomposed into a sum of two propagating cylindrical waves, one propagating radially outward from the z axis (namely the divergent wave), and one propagating radially inward from the z axis (namely the convergent wave), both being within the inner cylinder

$$J_n = \frac{1}{2}(H_n^+ + H_n^-), \quad \text{with } n = 0,1. \tag{10}$$

According to the time dependence $\exp(j2\pi ft)$ used throughout, the divergent wave H_n^+ is a Hankel function of the second kind $H_n^{(2)}$, and the convergent one H_n^- is a Hankel function of the first kind $H_n^{(1)}$, both of order n.

This decomposition can also be interpreted from the following physical considerations below. The axisymmetry of the problem imposes a null radial displacement on the revolution z axis. To satisfy this, we consider the z axis as a perfect reflecting virtual boundary (denoted now interface 1), upon which total reflection occurs with no mode conversion. As the boundary conditions concern potentials, the reflection coefficient of this virtual boundary is equal to unity. Hence, the reflection on the z revolution axis allows the recombination of the Bessel function. Now we will solve the problem of the axisymmetric elastic propagation in a cylindrical waveguide of radius b, infinite along the z axis of propagation, perfectly embedded in an infinite solid matrix, and arising from an initial velocity field at $z=0$ and $t=0$ within the inner waveguide ($r<b$). The initial velocity field is identical

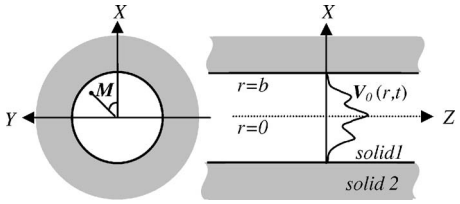


FIG. 1. Schematic representation of the initial velocity field for the for the embedded cylinder.

to the one described in Sec. II. It is centered on the z revolution axis and vanishes at $z=0$ and $t=0$ at the $r=b$ interface (Fig. 1).

At this stage, two additional assumptions related to the initial inner velocity field formulation have to be considered.

On the one hand, let us consider the divergent wave part of amplitude unity H_n^+ of the initial velocity field. It interacts with the embedded waveguide sidewalls to give rise first to four waves in the positive space and time regions, i.e., two reflected into the waveguide and two transmitted in the embedding medium, which in turn give rise to two reflected/transmitted waves, and so on (see Fig. 2). The reduced transmitted or reflected amplitudes are related to local reflection and transmission cylindrical coefficients R_{ij} and T_{ij} evaluated at interface $r=b$. Figure 2(a) shows the possible ray paths arising from the incident H_n^+ cylindrical divergent part, and involved in the construction of the velocity field in the guide for the z-positive region and $t>0$ and propagating in the z-positive direction.

On the other hand, the convergent wave part H_n^- of the velocity initial field can be seen as the final result of all multiple interactions with the embedded cylinder sidewalls coming from the z-negative region for $t<0$ propagating in the positive-z direction. In that case, exact recombination of the convergent part H_n^- of amplitude unity at the space and time origins is achieved. Figure 2(b) shows possible ray paths leading to the incident H_n^- cylindrical convergent part coming from the z-negative region and $t<0$ and propagating in the z-positive direction. The partition between incident, reflected, and transmitted waves at interface $r=b$ (Fig. 2) is such that stress-strain continuity conditions are satisfied at this interface.

The following section will be devoted to the derivation of the multiple reflected/transmitted waves amplitudes arising from H_n^+ only, since they constitute the causal part of the signal. On the basis of a Debye series expansion of generalized reflection and transmission coefficients, closed-form velocity field expressions will be derived.

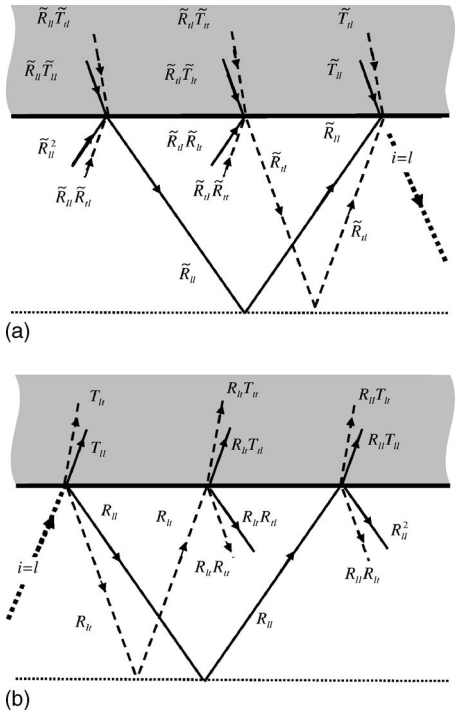


FIG. 2. (a) Example of the ray paths in the waveguide for the construction of the longitudinal H_n^- unity part. (b) Example of the ray paths in the waveguide arising from the longitudinal H_n^+ unity part, both for longitudinal incident polarization ($i=l$).

B. Theoretical derivation

Consider the general case where a longitudinal L_1^+ ($i=l$) and transversal T_1^+ ($i=t$) polarized divergent wave are incident on the waveguide sidewalls from the inside.

In medium 1, the spectral scalar Φ_1 potential (respectively, vector potential Ψ_1) solution is expressed as the sum of three potentials: one for the initial part in the infinite medium, one involving its reflection at $r=b$, and one involving the reflection of the vector potential (respectively, scalar potential) initial part at the same boundary.

In medium 2, the solution of the spectral scalar Φ_2 (respectively, vector Ψ_2 potential) is expressed as the sum of two potentials only, one involving the scalar initial part transmission at the boundary $r=b$ and one involving transmission of the vector potential (respectively, scalar potential) initial part at the same boundary.

Potentials in media 1 and 2 thus can be written

$$\begin{cases} \Phi_1 = \Phi_0[L_{1l}^- + L_{1l}^+ + X_{L_{1l}}^+ L_{1l}^+ + X_{L_{1l}}^- L_{1l}^-] + \Psi_0[X_{L_{1t}}^+ L_{1t}^+ + X_{L_{1t}}^- L_{1t}^-] & \text{for } r < b \\ \Psi_1 = \Psi_0[T_{1t}^- + T_{1t}^+ + X_{T_{1t}}^+ T_1^+ + X_{T_{1t}}^- T_1^-] + \Phi_0[X_{T_{1l}}^+ T_{1l}^+ + X_{T_{1l}}^- T_{1l}^-] & \\ \Phi_2 = \Phi_0[X_{L_{2l}}^+ L_{2l}^+] + \Psi_0[X_{L_{2t}}^+ L_{2t}^+] & \text{for } r > b \\ \Psi_2 = \Psi_0[X_{T_{2t}}^+ T_{2t}^+] + \Phi_0[X_{T_{2l}}^+ T_{2l}^+] & \end{cases}, \tag{11}$$

where $L_{1i}^{\pm}$ and $T_{1i}^{\pm}$ are the longitudinal and transversal divergent (+) and convergent (−) waves in medium 1; L_{2i}^+ and T_{2i}^+ are the longitudinal and transversal divergent (+) waves in medium 2, respectively, associated with the longitudinal part ($i=l$) and the

transversal part ($i=t$) of the initial field. Assuming the $\exp(j2\pi ft)$ time harmonic factor throughout, these waves are

$$L_{1i}^{\pm}(p_1,z,f)=\frac{1}{2}H_0^{\pm}(2\pi\alpha_{i1}r)\exp(-j2\pi w_{i1}z)$$

$$T_{1i}^{\pm}(p_1,z,f)=\frac{1}{2}H_1^{\pm}(2\pi\beta_{i1}r)\exp(-j2\pi w_{i1}z),$$

$$L_{2i}^{+}(p_1,z,f)=\frac{1}{2}H_0^{+}(2\pi\alpha_{i2}r)\exp(-j2\pi w_{i1}z),$$

$$T_{2i}^{+}(p_1,z,f)=\frac{1}{2}H_1^{+}(2\pi\beta_{i2}r)\exp(-j2\pi w_{i1}z). \tag{12}$$

The index i denotes the incident polarization of the initial field $i=l$ or $i=t$ and needs the following expressions:

$$\alpha_m=p_m,\quad w_1=w_{l_1},\quad \text{and } \beta_m=q_m(w_{l_1}), \tag{13}$$

for the incident longitudinal wave case ($i=l$),

$$\beta_m=p_m,\quad w_1=w_{t_1},\quad \text{and } \alpha_m=q_m(w_{t_1}), \tag{14}$$

for the incident transversal wave case ($i=t$).

The index m is for the medium 1 or 2.

The α_m and β_m are the generic forms of the radial frequencies associated with the reflected (or transmitted) waves, given an incident polarization i. Here, attention must be paid to the derivations of the radial frequencies q_1, p_2, and q_2 for the reflected cross-polarized wave and transmitted co- and cross-polarized waves, respectively. Let us consider the longitudinal incident case, for instance. Since the guided wave propagates as a whole down the z direction, q_1, p_2, and q_2 are deduced from the conservation of the axial longitudinal frequency w_{l_1} (longitudinal incident case) in the following dispersion relations:

$$p_m^2+w_{l_1}^2=\frac{f^2}{c_{l_m}^2}, \tag{15}$$

$$q_m^2+w_{l_1}^2=\frac{f^2}{c_{t_m}^2}. \tag{16}$$

Hence, substituting the w_{l_1} expression from (15) into (16) gives for medium 1

$$q_1(w_{l_1})=\sqrt{\frac{f^2}{c_{t_1}^2}-\frac{f^2}{c_{l_1}^2}+p_1^2}. \tag{17}$$

Similarly, for medium 2, $p_2(w_{l_1})$ and $q_2(w_{l_1})$ are, respectively,

$$p_2(w_{l_1})=\sqrt{\frac{f^2}{c_{l_2}^2}-\frac{f^2}{c_{l_1}^2}+p_1^2},\quad \text{and}\quad q_2(w_{l_1})=\sqrt{\frac{f^2}{c_{t_2}^2}-\frac{f^2}{c_{l_1}^2}+p_1^2}. \tag{18}$$

Similar derivations can be performed for the transversal incident case to obtain, respectively, $q_1(w_{t_1})$, $p_2(w_{t_1})$, $q_2(w_{t_1})$ by considering the axial transversal frequency w_{t_1} conservation.

The six unknown amplitudes of potentials in (11), relative to the incident polarization i ($i=l,t$), have to be determined. They are denoted by the vector $\boldsymbol{X}_i$,

$$\boldsymbol{X}_i=[X_{L_{2i}}^{+},X_{T_{2i}}^{+},X_{L_{1i}}^{-},X_{T_{1i}}^{-},0,0,X_{L_{1i}}^{+},X_{T_{1i}}^{+}]^{\text{trans}}, \tag{19}$$

(where the superscript trans is for vector transposition).

Here, we proceed the same way as was done in Deschamps and Chengwei (1991) for a solid plate immersed in two fluids half-spaces, and extend this formulation to the solid cylinder embedded in an infinite solid matrix. According to the z-axis symmetry, the upper half-space is the embedding solid matrix, the lower half-space is the vacuum, and the layer is the solid cylinder. As mentioned before, the source is inside the embedded layer.

The solutions are written under Debye series form, provided $|[R]|<1$

$$\boldsymbol{X}_i=\sum_{N=0}^{\infty}[R]^N\boldsymbol{S}_{i_1}, \tag{20}$$

where N is the number of interactions with interface 2.

This formulation retains explicitly the multiple reflection/refraction interactions with the boundaries.

$$\boldsymbol{S}_{i_1}=[T_{il},T_{it},R_{il},R_{it},0,0,0,0]^{\text{trans}}, \tag{21}$$

is the reflection/transmission column vector of the first incident polarization interaction ($i=l$ or $i=t$) with interface 2, and

$$[R]=\begin{bmatrix}[0] & [R_2]\\ [R_1] & [0]\end{bmatrix}, \tag{22}$$

the reflection/transmission matrix $[R]$.

The 4×4 refraction/reflection submatrices at interface 2 and 1, respectively $[R_2]$ and $[R_1]$, are of the form

$$[R_2]=\begin{bmatrix}0 & 0 & T_{ll} & T_{tl}\\ 0 & 0 & T_{lt} & T_{tt}\\ 0 & 0 & R_{ll} & R_{tl}\\ 0 & 0 & R_{lt} & R_{tt}\end{bmatrix}, \tag{23}$$

and

$$[R_1]=\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}. \tag{24}$$

R_{ll}, R_{lt}, R_{tl}, R_{tt} are the reflection coefficients (in terms of potential) of a cylindrical wave at the solid/solid interface 2 (at $r=b$) for longitudinal/longitudinal, longitudinal/transversal, transversal/longitudinal, transversal/transversal conversions, respectively. Similarly T_{ll}, T_{lt}, T_{tl}, T_{tt} are the transmission coefficients.

Re-expressing (20) in terms of double reflection/transmission interactions allows us to split the 8×8 system into two independent 4×4 systems such as,

$$\boldsymbol{X}_{i_A} = \sum_{N=0}^{\infty} [R_2]^N \boldsymbol{R}_i, \tag{25}$$

$$\boldsymbol{X}_{i_B} = [R_1] \sum_{N=0}^{\infty} [R_2]^N \boldsymbol{R}_i, \tag{26}$$

with

$$\boldsymbol{R}_i = [T_{il}, T_{it}, R_{il}, R_{it}]^{\text{trans}}. \tag{27}$$

Then, according to the structure of the $[R_1]$ matrix, the following trivial result is obtained:

$$X^-_{L_1} = X^+_{L_1} \text{ and } X^-_{T_1} = X^+_{T_1} \tag{28}$$

[it is worth noting (28) would no longer be valid for hollow cylinder configuration for instance]. Thus, as mentioned before, the perfect reflection on the z-revolution axis allows recombination of the Bessel function.

Finally, rearranging the 4×4 subsystems (25) and (26) into two independent 2×2 subsystems, for a polarized incident wave ($i=l$ or $i=t$), yields the final explicit expressions of the amplitudes of reflected convergent waves,

$$\begin{bmatrix} X^-_{L_{1i}} \\ X^-_{T_{1i}} \end{bmatrix} = \sum_{N=0}^{\infty} \begin{bmatrix} R_{ll} & R_{tl} \\ R_{lt} & R_{tt} \end{bmatrix}^N \begin{bmatrix} R_{il} \\ R_{it} \end{bmatrix}, \tag{29}$$

and these of the transmitted divergent waves,

$$\begin{bmatrix} X^+_{L_{2i}} \\ X^+_{T_{2i}} \end{bmatrix} = \sum_{N=0}^{\infty} \begin{bmatrix} T_{ll} & T_{tl} \\ T_{lt} & T_{tt} \end{bmatrix} \begin{bmatrix} R_{ll} & R_{tl} \\ R_{lt} & R_{tt} \end{bmatrix}^{N-1} \begin{bmatrix} R_{il} \\ R_{it} \end{bmatrix}. \tag{30}$$

Then, the space and time velocity field at a distance z from the source (with $i=l$ or $i=t$) is the double Bessel-Fourier transform of the spectral scalar and vector potentials (13) spatial derivatives, where unknown amplitudes $X^+_{L_{1i}}$, $X^+_{T_{1i}}$, $X^+_{L_{2i}}$, $X^+_{T_{2i}}$ are determined from their respective Debye series expansions (29) and (30).

For the sake of clarity, we only show below the velocity field for $r<b$ using the following synthesized form:

$$v(r,z,t) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} (2\pi)^2 p_1 E(f) [V_{ll}(p_1,f) + V_{lt}(p_1,f) + V_{tl}(p_1,f) + V_{tt}(p_1,f)] \exp(-j2\pi ft) dp_1\, df. \tag{31}$$

So, in comparison with (13), the inner waveguide field (31) must now include the four new involved contributions arising from the reflections of the scalar potential and vector potential.

For $v_{r_1}(r,z,t)$, the radial velocity component in medium 1, the V_{ll}, V_{lt}, V_{tl}, V_{tt} are, respectively,

$$V_{ll}(p_1,f) = -p_1 [1 + X^+_{L_{1l}}] J_1(2\pi p_1 r) \Phi_0(p_1,f) \times \exp(-j2\pi w_{l_1} z),$$

$$V_{lt}(p_1,f) = j w_{l_1} X^+_{T_{1l}} J_1(2\pi q_1(w_{l_1}) r) \Phi_0(p_1,f) \times \exp(-j2\pi w_{l_1} z),$$

$$V_{tl}(p_1,f) = -q_1(w_{t_1}) X^+_{L_{1t}} J_1(2\pi q_1(w_{t_1}) r) \Psi_0(p_1,f) \times \exp(-j2\pi w_{t_1} z),$$

$$V_{tt}(p_1,f) = j w_{t_1} [1 + X^+_{T_{1t}}] J_1(2\pi p_1 r) \Psi_0(p_1,f) \times \exp(-j2\pi w_{t_1} z), \tag{32}$$

and for $v_{z_1}(r,z,t)$, the axial velocity component in medium 1, the V_{ll}, V_{lt}, V_{tl}, V_{tt} are, respectively,

$$V_{ll}(p_1,f) = -j w_{l_1} [1 + X^+_{L_{1l}}] J_0(2\pi p_1 r) \Phi_0(p_1,f) \times \exp(-j2\pi w_{l_1} z),$$

$$V_{lt}(p_1,f) = q_1(w_{l_1}) X^+_{T_{1l}} J_0(2\pi q_1(w_{l_1}) r) \Phi_0(p_1,f) \times \exp(-j2\pi w_{l_1} z),$$

$$V_{tl}(p_1,f) = -j w_{t_1} X^+_{L_{1t}} J_0(2\pi q_1(w_{t_1}) r) \Psi_0(p_1,f) \times \exp(-j2\pi w_{t_1} z),$$

$$V_{tt}(p_1,f) = p_1 [1 + X^+_{T_{1t}}] J_0(2\pi p_1 r) \Psi_0(p_1,f) \times \exp(-j2\pi w_{t_1} z). \tag{33}$$

TABLE I. Material parameters.

Material	c_l (ms^{-1})	c_t (ms^{-1})	ρ (kg m^{-3})	radius (m)
Steel	5960	3260	7,932	8×10^{-3}
Cement grout	2810	1700	1600	∞

IV. NUMERICAL EXAMPLES AND DISCUSSION

An example of the application of the proposed solution technique is the guided wave propagation in a steel cylindrical bar embedded in an infinite grout matrix (material properties are in Table I). Even if the use of elastic guided wave propagation appears to be promising in the field of civil engineering for nondestructive evaluation of long length of rod, strands, free or embedded in cement grout (Pavlakovic *et al.*, 2001; Na *et al.*, 2002; Laguerre *et al.*, 2002; Finno and Chao, 2005), predictive modeling is still needed for designing new inspection methodologies. At this stage, we chose to highlight some specific capabilities of our model in addition to modal theory results.

A. Model excitation

The waveguide spatiotemporal excitation is applied through the cross section by axial v_z and/or radial v_r initial velocity components. Here, initial velocity field $\mathbf{v}(r,z=0,t) = \mathbf{v}(r)e(t)$ derives from a scalar potential Φ_0 only [so Ψ_0-dependent terms cancel out in (32) and (33)]. We also assume a Gaussian radial distribution for $v_z(r)$, even if the model imposes no restrictions on the initial spatial velocity distribution,

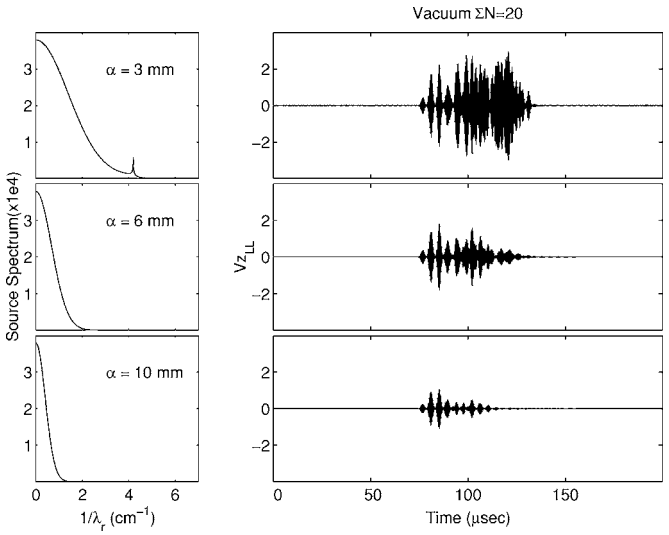


FIG. 3. Time history signal for three radii Gaussian beams (3, 6, and 10 mm) at $z=426$ mm for the free-steel cylinder waveguide.

$$v_z(r) = \begin{cases} v_0 \exp\left(-\dfrac{\pi r^2}{\gamma^2}\right) & r < b \\ 0 & r > b \end{cases}. \tag{34}$$

$\gamma/\sqrt{\pi}$ is the radius of the Gaussian beam, for which the velocity is decreased by a $1/e$ factor of its axis value and γ an arbitrary constant.

The Gaussian spatial distribution allows the analytical derivation of spectral quantities, and

$$V_z(p, z=0) = \int_0^{+\infty} v_0 \exp\left(-\frac{\pi r^2}{\gamma^2}\right) J_0(2\pi p r) r\, dr, \tag{35}$$

gives

$$V_z(p, z=0) = v_0 \exp(-\pi \gamma^2 p^2). \tag{36}$$

The time dependence $e(t)$ is a Gaussian windowed sinusoidal burst, whose center frequency and bandwidth can be varied.

B. Model simulations

Simulations were conducted, both in time and spectral domains, for the free-steel cylinder (in vacuum) case and the steel cylinder embedded in a cement grout matrix, at different propagation distances z, given a radial position in the guide.

An arbitrary frequency domain was chosen from 1.5 to 3.5 MHz.

To show the influence of the source spectrum on the time history signals, simulations were performed for the free waveguide at three different radii Gaussian beams $\gamma/\sqrt{\pi}$ (with $\gamma=3\times10^{-3}$, 6×10^{-3} and 10^{-2}) given a fixed number of interactions N (each source amplitude spectrum is normalized by its radius Gaussian beam). We observe (Fig. 3) that the source directivity tends to decrease the overall amplitude of the time signal, and make the leading portion of the signal predominant over the trailing one. Moreover, contrary to the trailing portion, the characteristic oscillating behavior of the leading portion is not really affected by the source directivity.

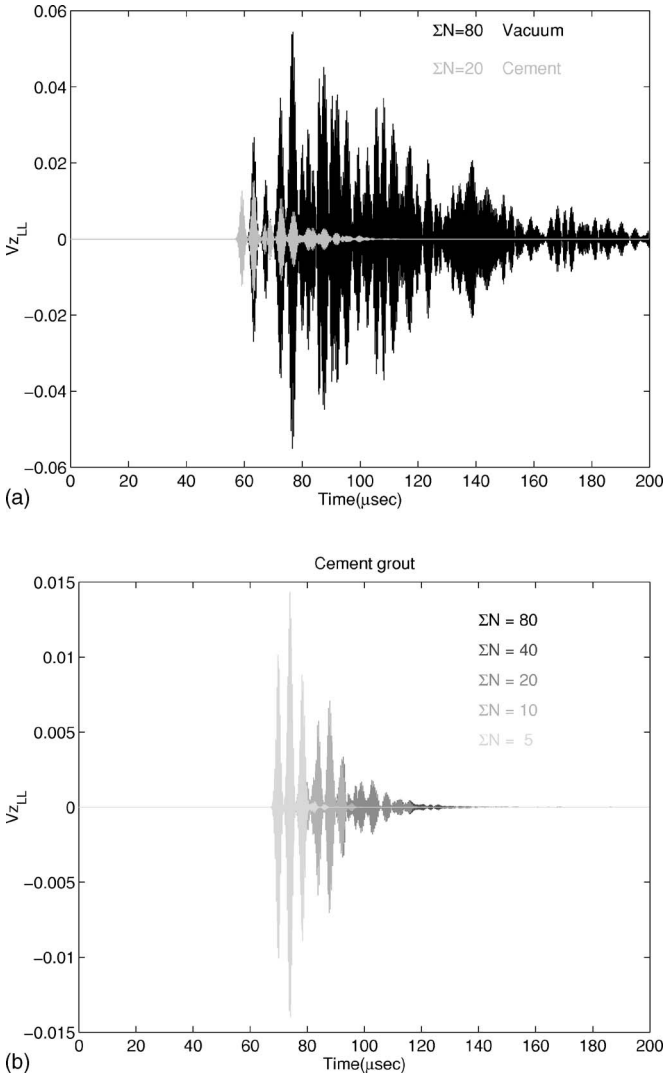


FIG. 4. (a) Time history signals for the free-steel waveguide (for $N=80$) and steel waveguide embedded in cement grout case (for $N=20$). (b) Convergence of the time solution for different N at $z=321$ mm.

In the following, to maximize sidewall interactions in the simulation to reveal the waveguide propagation effect, the radius Gaussian beam was chosen small ($\gamma=2\times10^{-3}$).

1. Time-domain signals

Figure 4(a) compares time waveforms in a free and embedded waveguide, respectively. The simulations were performed for a number of interactions ensuring the convergence of the results (as we will see below) for the given time window, that is $N=20$ for the free waveguide and $N=80$ for the embedded waveguide (as the solution is the superimposition of time-delayed echoes, one benefit of Debye series expansion is to avoid temporal aliasing for time-gated signals by adjusting N since the global response cannot be directly numerical as it theoretically spreads over an infinite time window).

The multiple sidewall reflections which give rise to constructive and destructive interferences in the trailing signal portion for the free-cylinder time waveform are strongly filtered out for the embedded cylinder signal. Indeed, the necessary condition for leakage $v_t^{\text{steel}} > v_t^{\text{grout}}$ is fulfilled for the

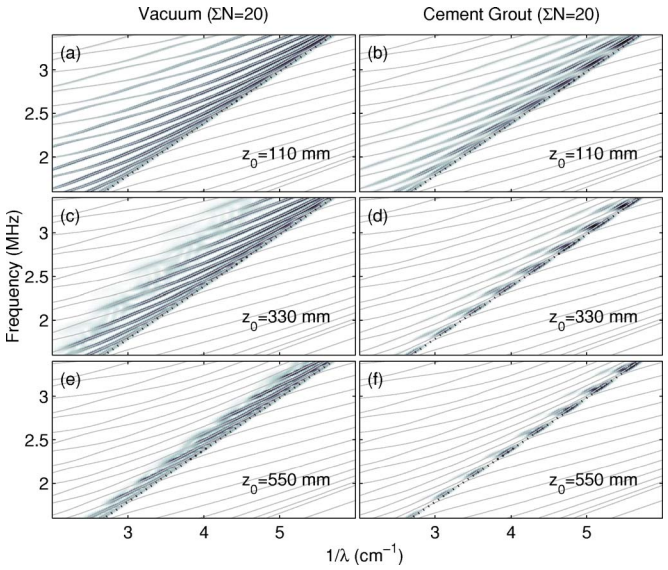


FIG. 5. (Color online) Simulated (f,w) diagrams (gray-scaled image) and longitudinal dispersion curves (gray solid lines) for $N=20$ and $z_0=110$, 330, and 550 mm—free steel cylinder case (a), (c), (e) and steel cylinder embedded in cement grout (b), (d), (f).

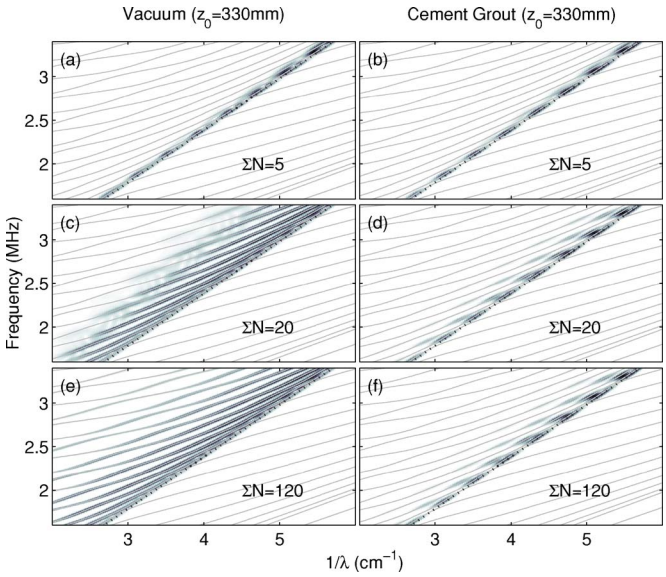


FIG. 6. (Color online) Simulated (f,w) diagrams (gray-scaled image) and longitudinal dispersion curves (gray solid lines) for $N=5$, 20, and 120 at $z_0=330$ mm—same cases as Fig. 5.

materials involved and so wave propagating in steel radiates into the surrounding cement grout. Besides, for any number of interactions, although the signal amplitude in the embedded cylinder signal is as expected lower than that in the free cylinder, we observe the arrival times in the leading portion are not really modified by the presence of the embedding material.

Finally, in order to test the solution convergence, simulations are shown in Fig. 4(b) for the embedded waveguide at $N=5,10,20,40,80$. The trailing part of the signal is progressively built as N increases. Here, the convergence is reached for $N=40$. (Calculation of the normalized root-mean-square (rms) error between predicted time waveforms at two following N—as mentioned above—gives 52, 17.4, 2.3, 2×10^{-2}%, respectively.)

2. Time-to-space frequencies (f,w) diagrams

The ability of the model was also considered by comparing the simulation results with dispersion curves in cylindrical waveguides. We processed the time-history signals simulated at 512 z positions using two-dimensional Fourier transform (Alleyne and Cawley, 1991). The dispersion curves were derived from modal solutions using DISPERSE software (*Disperse*, 2001).

To have a global view of how the model parameters affect the construction of dispersion curves, cross simulations were performed for both free and embedded waveguide by varying first propagation distances z_0 (z_0 is the mean distance over which the 2D transform is performed) at constant number of interactions N, then by varying the N parameter at z_0 constant.

Figures 5 and 6 show the results for $z_0=110$, 330, and 550 mm at $N=20$ and for $N=5$, 20, and 120 at $z_0=330$ mm, respectively. In the resulting $(f,w=1/\lambda_z)$ representations, we observe for the free-waveguide case [Figs. 5(a), 5(c), and 5(e)] a very good agreement between the simulated diagrams (image of the amplitude spectra normalized by their respective maximum) and $L(0,n)$ longitudinal mode dispersion curves (solid gray lines). Since the excitation is only longitudinal, note that simulated signal phase velocities describe only the portion of the dispersion curves greater than the longitudinal phase velocity in steel c_l (dotted black line).

We observe as well that increasing the propagation distance while keeping the number of interactions with the cylinder sidewalls constant leads progressively, first, to a poorer resolution of the modes and second, to their incomplete description (only the portion close to the c_l is described). For the embedded waveguide case [Figs. 5(b), 5(d), and 5(f)], the agreement of the simulated diagrams is very good for all propagation distances. The major difference with the free-waveguide case is that the solid embedding material causes some specific (f,w) portions to vanish. These modes cannot be fully described for these specific (f,w) portions because they leak energy from steel into the surrounding cement grout material as they propagate. Moreover, we observe that regions with minimum leakage attenuation have phase velocities closer to longitudinal phase velocity in steel and exhibit a periodical pattern. This phenomenon has already been reported in Pavlakovic *et al.* (2001) from modal approach simulation results. They observed that leakage attenuation minima coincide with the phase velocity which is just above the bulk longitudinal wave speed of the fastest material.

The simulations for different N at constant $z_0=330$ mm (Fig. 6) confirm the above major trends for increasing number of sidewall interactions. For example, in the case of the free waveguide [Figs. 6(a), 6(c), and 6(e)], the number of interactions must be increased with the propagation distance for a full description of the signal. Indeed, the diagram for $N=120$ at $z_0=330$ mm is quite similar to the one at $N=20$ at $z_0=110$ mm, while for the embedded waveguide case the convergence is faster due to leakage attenuation [Figs. 6(b), 6(d), and 6(f)] which causes multiple sidewalls interactions to cancel out.

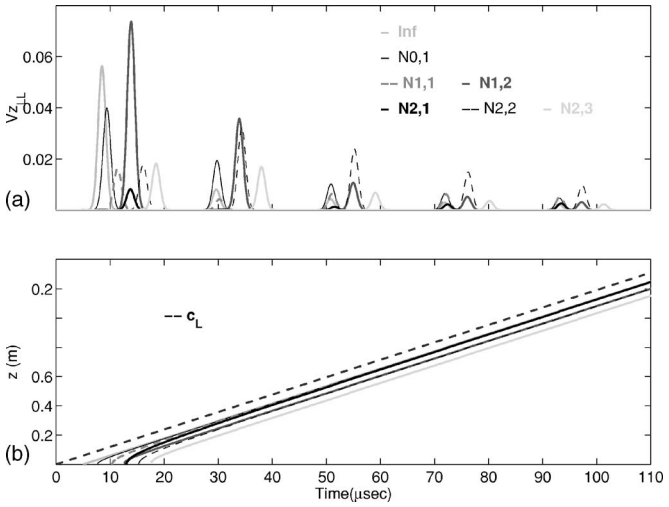


FIG. 7. (a) Time waveform envelopes of infinite medium and first terms of the longitudinal-to-longitudinal Debye series for the free-waveguide case at $z=21$, 141, 273, 399, and 525 mm. (b) Above-individual contributions time of flight versus propagation distance.

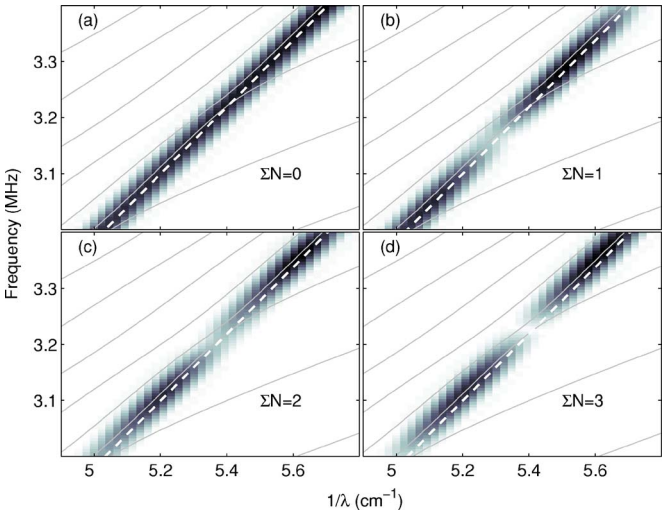


FIG. 8. (Color online) Evolution of the wave packet phase velocity for N varying from 0 to 3 with regard to longitudinal phase velocity in steel (white dashed line).

3. Discussion

Time waveform construction involving Debye series consists of the superimposition of time-delayed echoes. To highlight some capabilities of time waveform interpretation using Debye series, particularly for the leading part of the signal, we expand the longitudinal to longitudinal series (29) to the very first contributions,

$$X_{L_{1l}} = \overbrace{R_{ll}}^{N=0} + \overbrace{R_{ll}^2 + R_{lt}R_{tl}}^{N=1} + \overbrace{R_{ll}^3 + R_{lt}R_{tl}[2R_{ll} + R_{tt}]}^{N=2} + \cdots, \tag{37}$$

with $N_{1,1}=R_{ll}^2$, $N_{1,2}=R_{lt}R_{tl}$, and $N_{2,1}=R_{ll}^3$, $N_{2,2}=2R_{ll}R_{lt}R_{tl}$, and $N_{1,3}=R_{lt}R_{tt}R_{tl}$. Figure 7(a) shows the envelope of the first arrivals (up to $N=2$) at $z=21$, 141, 273, 399, and 525 mm), for the free waveguide, after source spectrum weighting and double Fourier-Bessel inverse transform. The infinite medium contribution (denoted by Inf) is also included. As the propagation distance increases, since the infinite medium contribution is weaker and weaker, we observe that the time waveform is rapidly governed by the time delay between the longitudinal/longitudinal contributions (with no polarization conversion) $(LL, LLL, \ldots)$ and the longitudinal/transversal and transversal/longitudinal conversions (LTL), and between the (LTL) and the $(LTLTL)$ and so on. This separation is effective because of the high-frequency broadband excitation which gives a pulse duration shorter than the time needed for a transversal wave to cross the guide. However, if we recall Fig. 4, only the four first wave packets are time delayed because, beyond $N=5$, the number of contributions between two successive double-mode conversions (LTL) grows faster to give rise to interference as attested to by the following reformulation of (29) in terms of double (LTL) mode conversion, using combinatorial analysis:

$$X_{L_{1l}} = \sum_{c=1}^{\infty} R_{ll}^c + \sum_{n=1}^{\infty}\sum_{c=n}^{\infty}\sum_{s=n}^{\infty} C_{c+1}^n C_s^{n-1} R_{ll}^{c-n+1} R_{tt}^{s-n+1} (R_{lt}R_{tl})^n,$$

where c and s are the number of longitudinal and transversal interactions, respectively, n the number of double-mode conversions, and C_j^i expresses the combination.

Figure 7(b) shows that the pulse's propagation is dispersive at short distances and reaches the longitudinal velocity in steel for long distances (that is, for distances greater than the cylinder radius). Since the contributions propagate at c_L, this can be used to simplify the calculation of the time duration between LL and LTL contributions. Indeed, even if a reflected pulse is the result of plane waves impinging on the waveguide sidewalls at various angles weighted by the source angular spectrum and reflection coefficients, one can consider as a first approximation that time delay can be calculated at a grazing angle. Hence, the time delay between (LL) and (LTL) or (LTL) and $(LTTL)$, for instance, is

$$\Delta t = \frac{2b}{\cos\theta_t}\left[\frac{1}{c_T} - \frac{\sin\theta_T \cos\theta}{c_L}\right], \tag{38}$$

where θ_L, θ_T, θ are the longitudinal, transversal, and emission angles $(\theta_L=\pi/2-\theta)$, respectively.

Hence, for $\theta=0$, this delay is only a function of the cylinder diameter and the longitudinal and transversal velocities

$$\Delta t = 2b\frac{\sqrt{c_L^2 - c_T^2}}{c_L c_T}. \tag{39}$$

The measured delay, 4.09 μs, is very close to the estimated theoretical one [4.109 μs from relation (39)].

It is also remarkable to observe in Fig. 8 that these are the interferences between the first arrivals propagating at c_L which progressively lead the wave packet to travel at a velocity slightly higher than c_L, for both free or embedded waveguide.

Moreover, it is worth noting that these are the interferences between the (LL) and (LTL) and between the higher-

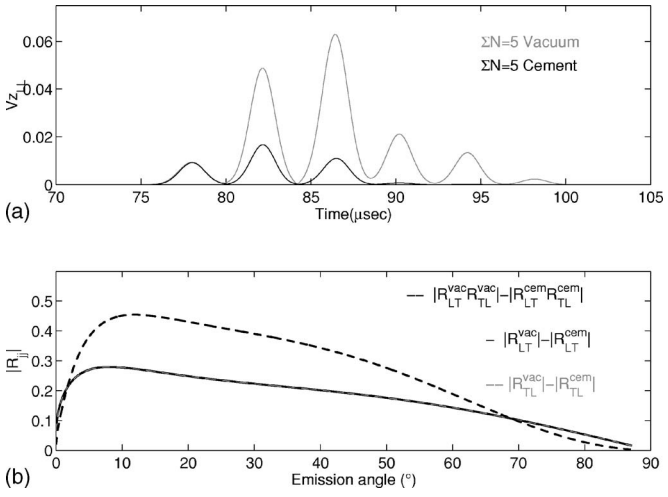


FIG. 9. (a) Influence of the surrounding medium on the time amplitude for $N=5$ at $z=433$ mm. (b) Comparison of steel/vacuum and steel/cement reflection coefficients (R_{lt}, R_{lt}, and $R_{lt}R_{tl}R_{lt}$, respectively) magnitudes versus emission angle at the 2.5-MHz center frequency.

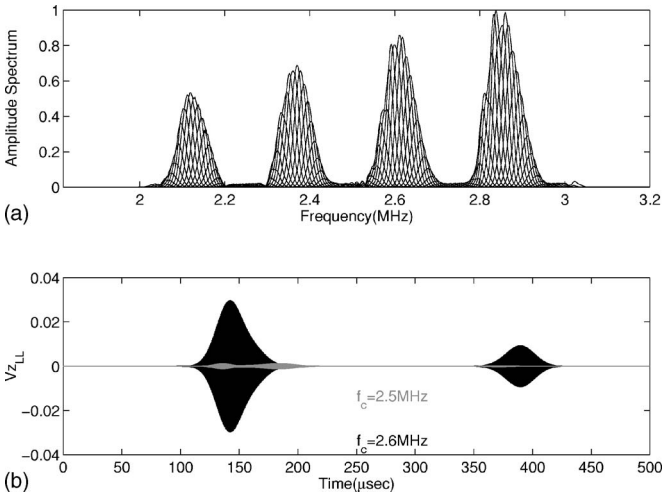


FIG. 10. (a) Normalized amplitude spectra of the 50-kHz narrow-band time history signals with center frequencies varying from 2 to 3 MHz after 1.2-m propagation through the embedded cylinder. (b) Time history signals propagated after 1.2- and 2-m propagation in the embedded cylinder for center frequencies of 2.5 and 2.61 MHz, respectively.

order (LTL) successive conversions which lead to the periodic behavior in the $f, w=1/\lambda_z$ diagrams along a line close to c_L.

From Fig. 9(a), for the embedded cylinder, we clearly observe the predominance of (LTL) double-mode conversion contribution in the process of leakage attenuation. The decrease of the double-conversion (LTL) contribution with the presence of surrounding material confirms this. For $N=5$ the multiple (LL) contributions do not modify the first arrival amplitude, while the successive arrivals depending upon double (LTL) conversion are strongly attenuated. This is the consequence of the decrease of the reflection coefficient R_{lt} and R_{tl} due to the surrounding medium as shown in Fig. 9(b) for the center frequency of the excitation.

Finally, with regard to results of the response of an embedded cylinder to a broadband high-frequency excitation, it appears clear that, to proceed to long-range inspection of such configuration as a grouted steel bar, it is necessary to use a narrow-band frequency pulse centered around these bright points.

Simulations were performed for a sinusoidal burst in a Gaussian window, by varying its center frequency by from 2 to 3 MHz by 10-kHz steps and using a 50-kHz constant bandwidth. Figure 10(a) shows the amplitude spectra of time history signals propagated through a 1.2-m, 8-mm-radius cement grouted steel. We retrieve the periodic pattern mentioned before and this confirms the very frequency selective behavior of the embedded cylinder. Time history signals at 1.2 and 2- m from the source with very close center frequencies (2.5 and 2.61 MHz, respectively) emphasize this point.

V. CONCLUSIONS

In order to predict the effect of the solid embedding material on the propagation of axisymmetric guided waves in a solid cylinder, we developed a semianalytical pulsed bounded-beam propagation model. Materials are assumed isotropic and elastic, the embedding material is infinite. An initial axial and radial axisymmetric velocity field was applied within the inner cylinder. Because the embedding material shear wave velocity is lower than the inner material shear wave velocity, it is partially guided in the inner cylinder and radiated in the embedding material. The radial solution in the waveguide in space- and time-frequencies domain is thus expressed as the contributions of an incident circular wave, a reflected wave (the guided wave) arising, and a transmitted wave (the radiating wave). The full reflected wave is expressed as a combination of the longitudinal-to-longitudinal, longitudinal-to-transversal, transversal-to-longitudinal, and transversal-to-longitudinal contributions. Each contribution is in turn expanded in terms of local reflection and transmission coefficients at the solid/solid interface (i.e., as Debye series). The full transmitted wave is derived in a similar way. The spatiotemporal field is obtained anywhere in the waveguide from inverse Fourier-Bessel transform over radial and time frequencies after source spectra amplitude weighting.

First simulations for a steel cylinder embedded in an infinite cement grout matrix are performed in the ultrasonic range and compare very favorably with classical modal curves, providing a well-suited number of interactions. Time-domain waveforms derived at different positions along the propagation direction emphasize the influence of the inner sidewall interaction and constitute benchmark waveforms for experimental time signals.

Time response to a broadband high-frequency excitation gives similar results for the free and embedded waveguides in the leading portion of the signal. We observe that double longitudinal-transversal and transversal to longitudinal conversions govern this leading portion as well as the radiation attenuation in the surrounding cement grout. We show that radiation attenuation levels are reduced for phase velocities closer to longitudinal phase velocity in steel. According to these simulations, a methodology is proposed to minimize the radiation attenuation.

Alleyne, D., and Cawley, P. (**1991**). "A two-dimensional Fourier transform method for the measurement of propagating multimode signals," J. Acoust. Soc. Am. **89**(3), 1159–1168.

Barshinger, J., Rose, L., and Avioli, M. (**2002**). "Guided wave resonance tuning for pipe inspection," Drying Technol. **124**, 303–310.

Chen, C., and Toksöz, M. (**1981**). "Elastic wave propagation in a fluid-filled borehole and synthetic acoustic logs," Geophysics **46**(7), 1042–1053.

Chree, C. (**1889**). "The equations on an isotropic elastic solid in polar and cylindrical coordinates, their solutions, and applications," Trans. Cambridge Philos. Soc. **14**, 250–369.

Danila, E., Conoir, J.-M., and Izbicki, J.-L. (**1995**). "The generalized Debye series expansion: Treatment of the concentric and nonconcentric cylindrical fluid-fluid interfaces," J. Acoust. Soc. Am. **98**(6), 3326–3342.

Danthez, J.-M., Deschamps, M., and Gérard, A. (**1989**). "Réponse spatiotemporelle d'un guide cylindrique à un faisceau borné. I. Partie théorique ("Spatiotemporal response of a cylinder to a bounded beam: Theoretical part")," J. Acoust. **2**, 119–125.

Deschamps, M., and Chengwei, C. (**1991**). "Reflection/refraction of a solid layer by Debye's series expansion," Ultrasonics **29**, 288–293.

Deschamps, M., and Hosten, B. (**1992**). "The effects of viscoelasticy on the reflection and transmission of ultrasonic waves by orthotropic plate," J. Acoust. Soc. Am. **91**(4), 2007–2015.

DISPERSE (2001). "A system for generating dispersion curves," User's manual version 2.0.16d.

Finno, R., and Chao, H. (**2005**). "Guided waves in embedded concrete piles," J. Geotech. Geoenviron. Eng. **124**(10), 965–975.

Laguerre, L., Aime, J.-C., and Brissaud, M. (**2002**). "Magnetostrictive pulse-echo device for nondestructive evaluation of cylindrical steel materials using longitudinal guided waves," Ultrasonics **39**(7), 503–514.

Lowe, M. (**1995**). "Matrix techniques for modelling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(4), 525–542.

Na, W., Kundu, T., and Ehsani, M. (**2002**). "Ultrasonic guided waves for steel-bar concrete interface testing," Mater. Eval. **60**, 437–444.

Pavlakovic, B., Lowe, M., and Cawley, P. (**2001**). "High-frequency low-loss ultrasonic modes in imbedded bars," ASME J. Appl. Mech. **68**, 67–75.

Pochhammer, J. (**1876**). "Uber die fortpflanzungsgeschwindigkeiten kleiner schwingungen in einem unbergrenszten isotropen kreiscylinder ("On the propagation velocities of small vibrations in an infinite isotropic cylinder")," J. Reine Angew. Math. **81**, 324–336.

Puckett, A., and Peterson, M. (**2005**). "A semi-analytical model for predicting multiple waveguide propagating axially symmetric modes in cylindrical waveguides," Ultrasonics **43**, 197–207.

Rao, V., and Vandiver, J. (**1999**). "Acoustics fluid-filled boreholes with pipe: Guided propagation and radiation," J. Acoust. Soc. Am. **105**(6), 3057–3066.

Roever, W., Rosenbaum, J., and Vining, T. (**1971**). "Acoustic waves from an impulsive source in a fluid-filled borehole," J. Acoust. Soc. Am. **55**(6), 1144–1157.

Thurston, R. (**1978**). "Elastic waves in rods and clad rod," J. Acoust. Soc. Am. **64**(1), 1–37.

Zeroug, S., and Felsen, L. (**1994**). "Nonspecular reflection of two and three-dimensional acoustic beams from fluid-immersed plane-layered elastic structures," J. Acoust. Soc. Am. **95**(6), 3075–3089.

Zeroug, S., and Felsen, L. (**1995**). "Nonspecular reflection of two and three-dimensional acoustic beams from fluid-immersed cylindrically layered elastic structures," J. Acoust. Soc. Am. **98**(1), 584–598.

Wave propagation along transversely periodic structures

Mihai V. Predoi[a)]
Department of Mechanics, University Politehnica Bucharest, Splaiul Independentei, 313, Sect. 6, Bucharest 77201 Romania

Michel Castaings, Bernard Hosten, and Christophe Bacon
Laboratoire de Mécanique Physique, Université Bordeaux 1, UMR CNRS 5469, 351 cours de la Libération, 33405 Talence Cedex, France

(Received 18 October 2006; revised 12 December 2006; accepted 5 January 2007)

The dispersion curves for guided waves have been of constant interest in the last decades, because they constitute the starting point for NDE ultrasonic applications. This paper presents an evolution of the semianalytical finite element method, and gives examples that illustrate new improvements and their importance for studying the propagation of waves along periodic structures of infinite width. Periodic boundary conditions are in fact used to model the infinite periodicity of the geometry in the direction normal to the direction of propagation. This method allows a complete investigation of the dispersion curves and of displacement / stress fields for guided modes in anisotropic and absorbing periodic structures. Among other examples, that of a grooved aluminum plate is theoretically and experimentally investigated, indicating the presence of specific and original guided modes. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2534256]

PACS number(s): 43.35.Cg, 43.20.Mv [RLW] Pages: 1935–1944

I. INTRODUCTION

Guided waves have a significant potential for structural health monitoring (SHM) and nondestructive evaluation (NDE) due to their relatively long distance of propagation and sensitivity to discontinuities along the propagation path, anywhere through the material thickness. These advantages are partially diminished by the existence of several propagating modes at any frequency, wave dispersion and attenuation. The first objective to be attained in a SHM or NDE technique is to determine the dispersion curves and the modal characteristics such as waveguide stress-displacement pattern and wave velocities. Several methods have been developed for the dispersion equation solution. For simple geometries of the waveguide such as plates or cylinders made of homogeneous or multilayered, isotropic materials, numerical methods are used to determine solutions of analytical transcendental equations.[1] More general solutions can be obtained using methods based on the superposition of bulk waves, including, for example, the transfer matrix method[2,3] or the surface impedance matrix method.[4] These methods are oriented towards multilayered structures of viscoelastic, anisotropic materials, but the root searching method in the complex plane of wave numbers can sometimes miss some solutions. This shortcoming and the inability to investigate waveguides of arbitrary cross section, has led to the development of a new technique called semianalytical finite element method (SAFE) also referred to in the literature as spectral or waveguide finite element methods.[5–28] This method uses standard eigenvalue finding routines, which have been implemented into commercial finite element method (FEM) codes several decades ago and considerably improved since then. The main advantage of the SAFE method is that only the cross section of the waveguide, i.e., the section normal to the direction of wave propagation, has to be meshed by finite elements. The waves are assumed to be harmonic. Two directions of investigation have been followed in the last three decades:

(a) For plane structures of constant thickness, the SAFE method is applied on a line segment running normally through the thickness [one-dimensional (1D)-SAFE]. The finite elements are one-dimensional elements. The accuracy for high-order wave number values and displacements/stresses field shapes increases with the number of elements. The use of a one-dimensional finite element mesh to describe the cross-sectional deformation of a laminate was first used three decades ago.[5,6] The characteristic equation for free-wave propagation was expressed as a linear eigenvalue problem in frequency. Later works demonstrated that the complex roots of the dispersion curves could be obtained by writing the characteristic equation as an eigenvalue problem in wave numbers.[7] A similar approach was employed to investigate the dispersion curves for anisotropic laminates, based also on the state of plane strains.[8] Fixing the direction of propagation in elastic laminated composite cylinders, the linear mesh has been used to investigate the dispersion curves.[9] The authors used a full displacement field for the cylinder and the characteristic equation was formulated as a linear eigenvalue problem in frequency.

In a study of plate edge reflection of Lamb waves, the dispersion curves for elastic plates were determined using the same approach, fixing the wave number and formulating the eigenvalue problem in frequencies.[10,11] Convergence criteria for mesh size are determined. Wave propagation and damping in linear viscoelastic laminates have also been investigated using the 1D-SAFE method.[12] More recently, the

[a)]Author to whom correspondence should be addressed. Electronic mail: predoi@cat.mec.pub.ro

method has been applied to two classes of viscoelastic plates and a first promising tentative was made to model wave interaction with a spar reinforced wing.[13]

(b) For waveguides having an arbitrary closed-contour cross section, the [two-dimensional (2D)-SAFE] method was applied onto the two-dimensional cross section of the waveguide normal to the direction of propagation. The method had been used for the first time more than 30 years ago for triangular waveguides.[14] The use of a finite element mesh to describe the cross-sectional deformation of a waveguide has been discussed extensively in studies concerned with wave propagation in linear elastic thin walled shells,[15] railway tracks,[16–18] rib stiffened plates,[19] rods[20] and fluid filled pipes.[21] In each of these studies, the cross-sectional area of the waveguide is of finite extent and the finite elements are used to model the three-dimensional displacement field across the normal section of the waveguide. Another direction of study using the FEM method consisted in assuming the waveguide as a periodic chain of the given guide section over one element depth, repeating this cell along the guide using periodicity conditions.[22–24] This approach can be applied to waveguides of arbitrary cross section and any curve taken as waveguide mean fiber. A composite plane structure made of a homogeneous substrate and periodic particle reinforcements was also investigated via specially designed finite elements in order to obtain the effective mechanical properties.[25] Waveguides with periodic geometry along the direction of wave propagation and separating into several branches have also been addressed.[26] Dispersion curves for layered composites have been determined using spectrally formulated finite elements for beams[27] and for elastic composite plates.[28] This method requires a fast Fourier transform of the structure response to a point excitation.

The present work proposes an extension of the 2D-SAFE method for modeling wave propagation along guides of finite thicknesses, but of infinite widths. This allows, for instance, a complete investigation of the dispersion curves for waveguides having infinitely periodic geometry and/or material properties along the width of the waveguide. Specific periodic boundary conditions are implemented and applied to the sides of a finite meshed domain in order to render this domain infinitely repeated. The method is first validated by predicting the dispersion curves (complex wave numbers) for Lamb-like and SH-like modes guided along an anisotropic and viscoelastic material plate, and by comparing the results to those obtained with a semianalytical model. Then, the case of a grooved aluminum plate is theoretically and experimentally investigated, indicating the presence of new guided modes.

II. FUNDAMENTALS OF THE SAFE METHOD

The usual approach for structures of constant thickness is to consider plane strain state for the displacements fields.[5–13] In these conditions, the 1D-SAFE method can be applied across the plate thickness, which is the unique geometric parameter. This approach has been applied to Lamb waves and can be extended to include shear horizontal (SH) waves.

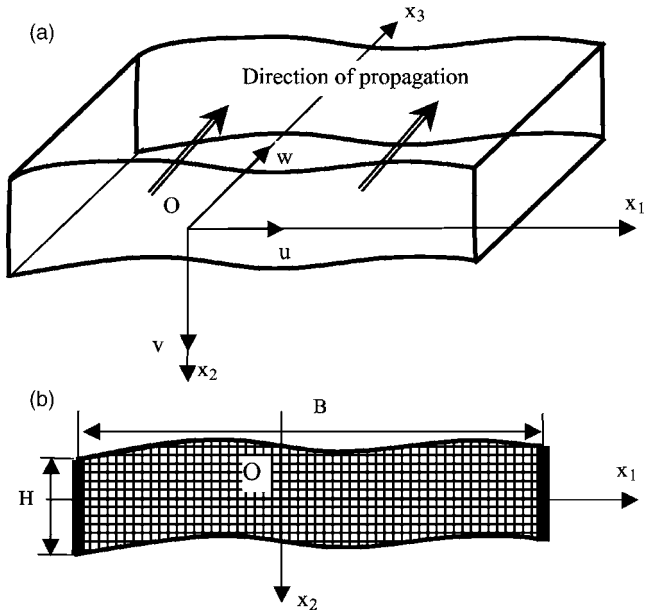


FIG. 1. Schematic of the problem, (a) in three dimensions (3D) and (b) in 2D; thick lines in (b) point out the two sides of the cross section where periodic boundary conditions are applied.

In the following, we consider the three-dimensional elasticity approach, so that no simplifications are made to the elastic tensor, nor to the displacement field. The harmonic guided waves are assumed to propagate along the Ox_3 axis (Fig. 1). Consequently, the displacements vector in the waveguide can be written

$$u_i(x_1,x_2,x_3,t) = U_i(x_1,x_2)e^{I(kx_3-\omega t)} \quad I=\sqrt{-1}, \tag{1}$$

in which k is the wave number and $\omega=2\pi f$ the angular frequency, f being the frequency, the subscript $i=1,2,3$ here and in the following. The problem has two geometric variables and will require the analysis onto the normal cross section of the waveguide (section width B and plate thickness H), corresponding to the Ox_1x_2 plane in Fig. 1. The displacements gradients deduced:

$$\frac{\partial u_i}{\partial x_1} = \frac{\partial U_i}{\partial x_1}e^{I(kx_3-\omega t)},$$

$$\frac{\partial u_i}{\partial x_2} = \frac{\partial U_i}{\partial x_2}e^{I(kx_3-\omega t)},$$

$$\frac{\partial u_i}{\partial x_3} = ikU_i e^{I(kx_3-\omega t)}. \tag{2}$$

The differential equations of motion in an elastic domain (D) of mass density ρ and elastic constants C_{ijkl} are

$$\sum_{j,k,l=1}^{3}\left[C_{ikjl}\frac{\partial^2 U_j}{\partial x_k \partial x_l}\right] + \rho\omega^2 U_i = 0 \ in \ D; \quad i=1,2,3. \tag{3}$$

Using Eq. (2), after some intermediary transformations, one gets

$$C_{ikjl}\frac{\partial^2 U_j}{\partial x_k \partial x_l} + I(C_{i3jk} + C_{ikj3})\frac{\partial (kU_j)}{\partial x_k} - kC_{i3j3}(kU_j) + \rho\omega^2 \delta_{ij} U_j = 0 \tag{4}$$

with summation over the indices $j=1,2,3$ and k, $l=1,2$. The stresses on the boundaries δD of (D) are written

$$T_i = \sum_{j,k,l=1}^{3} C_{ikjl}\frac{\partial U_j}{\partial x_l} n_k; \quad on\ \delta D \tag{5}$$

in which n_k are the components of $\mathbf{n}$, the outward unit vector normal to the boundary δD. Applying again the derivation expressions (2), the stresses become

$$T_i = C_{ikjl}\frac{\partial U_j}{\partial x_l} n_k + IC_{ikj3}(kU_j)n_k; \quad on\ \delta D \tag{6}$$

with summation over the indices $j=1,2,3$ and k, $l=1,2$.

In a commercial FEM code,[29] the input formalism for eigenvalues problems has the general expression

$$\nabla \cdot (c\,\nabla \tilde{U} + \alpha \tilde{U} - \gamma) - \beta \nabla \tilde{U} - a\tilde{U} + \lambda d_a \tilde{U} = 0 \tag{7}$$

in which all matrix coefficients admit complex values, which is essential for viscoelastic materials, and $\tilde{U}$ represents the set of variables to be determined. If $\gamma=0$, Eq. (7) can be expressed in the form

$$C_{ijkl}\frac{\partial^2 \tilde{U}_j}{\partial x_k \partial x_l} + (\alpha_{ijk} - \beta_{ijk})\frac{\partial \tilde{U}_j}{\partial x_k} - a_j \tilde{U}_j + \lambda d_{ij}\tilde{U}_j = 0. \tag{8}$$

The generalized Neumann boundary conditions on δD are expressed in the same FEM code as

$$\mathbf{n} \cdot (c\,\nabla \tilde{U} + \alpha \tilde{U}) + q\tilde{U} = 0, \tag{9}$$

which can be expressed using the above expressions as

$$C_{ijkl}\frac{\partial \tilde{U}_j}{\partial x_l} n_k + \alpha_{ijk}\tilde{U}_j n_k + q_{ij}\tilde{U}_j = 0. \tag{10}$$

The quadratic eigenvalue problem (4) in wave numbers can be cast into linear form by introducing a new vector variable $\mathbf{V}$ as

$$\mathbf{M} \cdot \mathbf{V} = k\mathbf{M} \cdot \mathbf{U}, \tag{11}$$

in which $\mathbf{M}$ is an arbitrary diagonal matrix. In order to write Eqs. (4) and (6) in the form of Eqs. (8) and (10), respectively, a new set of variables $\tilde{\mathbf{U}}=[U_1 U_2 U_3 V_1 V_2 V_2]^T$ must be introduced. With this new set of variables, the FEM coefficients must be

$$d_a = \begin{bmatrix} \mathbf{0} & \mathbf{D} \\ \mathbf{M} & \mathbf{0} \end{bmatrix}; \quad c = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \quad \alpha = \begin{bmatrix} \mathbf{0} & I\mathbf{A} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \tag{12}$$

$$\beta = \begin{bmatrix} \mathbf{0} & -I\mathbf{B} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}; \quad a = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix}.$$

It is assumed that $\mathbf{0}$ represents a zero matrix of appropriate dimension. The submatrices are

$$\mathbf{D} = \begin{bmatrix} -C_{55} & -C_{54} & -C_{53} \\ -C_{45} & -C_{44} & -C_{43} \\ -C_{35} & -C_{34} & -C_{33} \end{bmatrix}; \tag{13}$$

$$\mathbf{M} = \begin{bmatrix} -\rho\omega^2 & 0 & 0 \\ 0 & -\rho\omega^2 & 0 \\ 0 & 0 & -\rho\omega^2 \end{bmatrix}.$$

$$\mathbf{C} = \begin{bmatrix} \begin{bmatrix} C_{11} & C_{16} \\ C_{61} & C_{66} \end{bmatrix} & \begin{bmatrix} C_{16} & C_{12} \\ C_{66} & C_{62} \end{bmatrix} & \begin{bmatrix} C_{15} & C_{14} \\ C_{65} & C_{64} \end{bmatrix} \\ \begin{bmatrix} C_{61} & C_{66} \\ C_{21} & C_{26} \end{bmatrix} & \begin{bmatrix} C_{66} & C_{62} \\ C_{26} & C_{22} \end{bmatrix} & \begin{bmatrix} C_{65} & C_{64} \\ C_{25} & C_{24} \end{bmatrix} \\ \begin{bmatrix} C_{51} & C_{56} \\ C_{41} & C_{46} \end{bmatrix} & \begin{bmatrix} C_{56} & C_{52} \\ C_{46} & C_{42} \end{bmatrix} & \begin{bmatrix} C_{55} & C_{54} \\ C_{45} & C_{44} \end{bmatrix} \end{bmatrix}. \tag{14}$$

$$\mathbf{A} = \begin{bmatrix} \begin{bmatrix} C_{15} \\ C_{65} \end{bmatrix} & \begin{bmatrix} C_{14} \\ C_{64} \end{bmatrix} & \begin{bmatrix} C_{13} \\ C_{63} \end{bmatrix} \\ \begin{bmatrix} C_{65} \\ C_{25} \end{bmatrix} & \begin{bmatrix} C_{64} \\ C_{24} \end{bmatrix} & \begin{bmatrix} C_{63} \\ C_{23} \end{bmatrix} \\ \begin{bmatrix} C_{55} \\ C_{45} \end{bmatrix} & \begin{bmatrix} C_{54} \\ C_{44} \end{bmatrix} & \begin{bmatrix} C_{53} \\ C_{43} \end{bmatrix} \end{bmatrix}; \tag{15}$$

$$\mathbf{B} = \begin{bmatrix} \begin{bmatrix} C_{51} \\ C_{56} \end{bmatrix} & \begin{bmatrix} C_{56} \\ C_{52} \end{bmatrix} & \begin{bmatrix} C_{55} \\ C_{54} \end{bmatrix} \\ \begin{bmatrix} C_{41} \\ C_{46} \end{bmatrix} & \begin{bmatrix} C_{46} \\ C_{42} \end{bmatrix} & \begin{bmatrix} C_{45} \\ C_{44} \end{bmatrix} \\ \begin{bmatrix} C_{31} \\ C_{36} \end{bmatrix} & \begin{bmatrix} C_{36} \\ C_{32} \end{bmatrix} & \begin{bmatrix} C_{35} \\ C_{34} \end{bmatrix} \end{bmatrix},$$

where $C_{IJ}(I,J=1,\ldots,6)$, is the contracted notation for the complex stiffness moduli C_{ikjl}.

It can be easily verified that stress free boundary conditions correspond to Neumann conditions

$$\mathbf{n}[c\,\nabla \hat{\mathbf{U}} + \alpha \hat{\mathbf{U}}] = \mathbf{0}. \tag{16}$$

Rigidly fixed boundaries can also be implemented by use of the Dirichlet boundary conditions $\hat{\mathbf{U}}=\mathbf{0}$ on the corresponding boundaries.

Another special case of interest is that of periodic boundary conditions. It is a particular case of Neumann boundary conditions. The variables and their derivatives up to the element order are forced to take identical values on a pair of boundaries of the structure. For the case of a rectangular domain $[-B/2B/2]\times[H/2H/2]$, connecting all the nodal displacements along the $x_1=-B/2$ edge to those along the $x_1=B/2$ edge [Fig. 1(b)], for quadratic finite elements means

$$\hat{\mathbf{U}}|_{-B/2} = \hat{\mathbf{U}}|_{B/2}$$

$$\nabla \hat{\mathbf{U}}|_{-B/2} = \nabla \hat{\mathbf{U}}|_{B/2}. \tag{17}$$

These conditions represent continuity of displacements and stresses between the two edges, without imposing any particular values.

TABLE I. Complex viscoelastic coefficients C_{IJ} (GPa) for the anisotropic material.

C_{IJ}	$J=1$	2	3	4	5	6
$I=1$	$21.25(1+0.02i)$	$7.5(1+0.02i)$	$13.25(1+0.02i)$	0	$-6.25(1+0.02i)$	0
2		$15(1+0.02i)$	$7.5(1+0.02i)$	0	$0.5(1+0.02i)$	0
3			$21.25(1+0.02i)$	0	$-6.25(1+0.02i)$	0
4	Symm.			$3.7(1+0.02i)$	0	$-0.25(1+0.03i)$
5					$10.25(1+0.02i)$	0
6						$3..7(1+0.02i)$

III. ANISOTROPIC VISCOELASTIC COMPOSITE PLATE

The objective of our first use of the 2D-SAFE method is to verify the correct implementation of the general anisotropic formalism, and to check the efficiency and the interest of the periodic boundary conditions. The waveguide is then a plate, so that the semianalytical method presented in Ref. [4] can be used for supplying data used for validating the SAFE results. This plate is made of an anisotropic, viscoelastic material the properties of which are given in Table I (complex C_{IJ} in GPa). These properties are those of a unidirectional composite made of glass fibers and epoxy matrix. The C_{IJ} are expressed in the coordinate axis $(0,x_1,x_2,x_3)$ of Fig. 1, and the direction of the propagation makes a 45° angle with the fibers, which are contained in the plane of the plate. In these conditions, the plane of propagation does not coincide with a plane of symmetry of the material. Therefore, all guided modes are supposed to produce three nonzero displacement components, and plane strain conditions are not considered here. The mass density is $\rho=1.8\ \mathrm{g/mm^3}$.

The first approach is to study a rectangular cross-section waveguide of thickness $H=1$ mm and to vary the width B of the section in order to quantify the effect of the ratio B/H on the dispersion curves of the guided modes. Figures 2(a) and 2(b) present the phase velocities predicted with the SAFE method (dots) for $B=2$ mm and $B=4$ mm, respectively and compare these results with those obtained by using the surface impedance method,[4] that considers the plate as infinitely wide (lines). Stress free boundary conditions are applied on all four sides of the meshed rectangle that represents the cross section of the guide, according to the 2D-SAFE method. The SAFE analysis is performed at 18 frequencies between 100 and 1800 kHz. The phase velocities shown in Fig. 2(a) indicate that $B/H=2$ corresponds to a relatively good approximation for simulating the A_0-type mode along the infinitely wide plate, and to a bad approximation for all other modes. Among the solutions supplied by the SAFE method, torsion modes are obtained for this rod of rectangular cross section. These, of course, are not found by the semianalytical model developed for infinitely wide plates. Higher order modes are even more difficult to recognize among the obtained SAFE solutions due to the complexity of their modal shape. As shown in Fig. 2(b), better agreement is obtained for the A_0 mode in the case $B/H=4$, but many other modes which are typical for a rod have no equivalent in plate dispersion curves. It has been checked that increasing the B/H ratio leads to a better approximation of the infinitely wide plate, but this, of course, implies the use of larger numbers of elements and is therefore time consuming.

This comparison between wave numbers for plates of infinite width B with those for structures of finite width, which are in general considered to be rods, can be useful in applications with relatively narrow planar specimens. In fact, an infinitely wide planar structure can be seen as juxtaposition of an infinite number of finite domains of width B and height H alongside their width. This leads to the idea of modeling an infinitely wide structure by a periodic domain of finite width B. Periodic boundary conditions (PBCs) defined by Eqs. (17) are set on boundaries $x_1=-B/2$ and $x_1=B/2$. Stress-free conditions remain applied along the $x_2=\pm H/2$ boundaries, considering the plate in vacuum. Since (PBCs) impose continuity and no other restrictions on the displacements and stress fields along the $\pm B/2$ side edges, the SH modes are included in the solution. Another important remark is that the value for the width B in the SAFE method is not important for the solution, since the resulting structure is an infinitely wide plate made of identical, adjacent blocks with continuity of both displacements and stresses at their junctions. The effective B value can then be chosen very small in order to reduce the number of mesh elements. If the planar structure is layered, there is no restriction on the number or on the mechanical properties of the superposed layers. In addition, the structure can include periodic inhomogeneities along the width, i.e. along the Ox_1 direction, and this important property will be discussed in the following sections.

The previous composite plate is thus investigated again, but this time by meshing a strip domain of width $B=0.25$ mm only, and imposing periodic boundary conditions (PBCs) at both sides located at $x_1=\pm B/2$. The obtained phase velocities [Fig. 2(c)] and modal attenuation curves [Fig. 2(d)] are presented as dots and compared to the semianalytical simulations (lines). The results are in excellent agreement for all the guided modes in the whole frequency range of investigation, including the SH modes. The number of degrees of freedom for this SAFE model was 4110, and the computation for finding 50 complex wave number roots for each of the 18 frequencies took around 4 min using a 1.8 GHz, bi-processor MACINTOSH G5 machine.

The number and accuracy of eigenvalues to be determined by the FEM code can be selected, imposing, however, a minimum number of degrees of freedom of the FEM mesh. Most of these roots are complex numbers even for elastic materials. Selecting the propagating or quasi-propagating

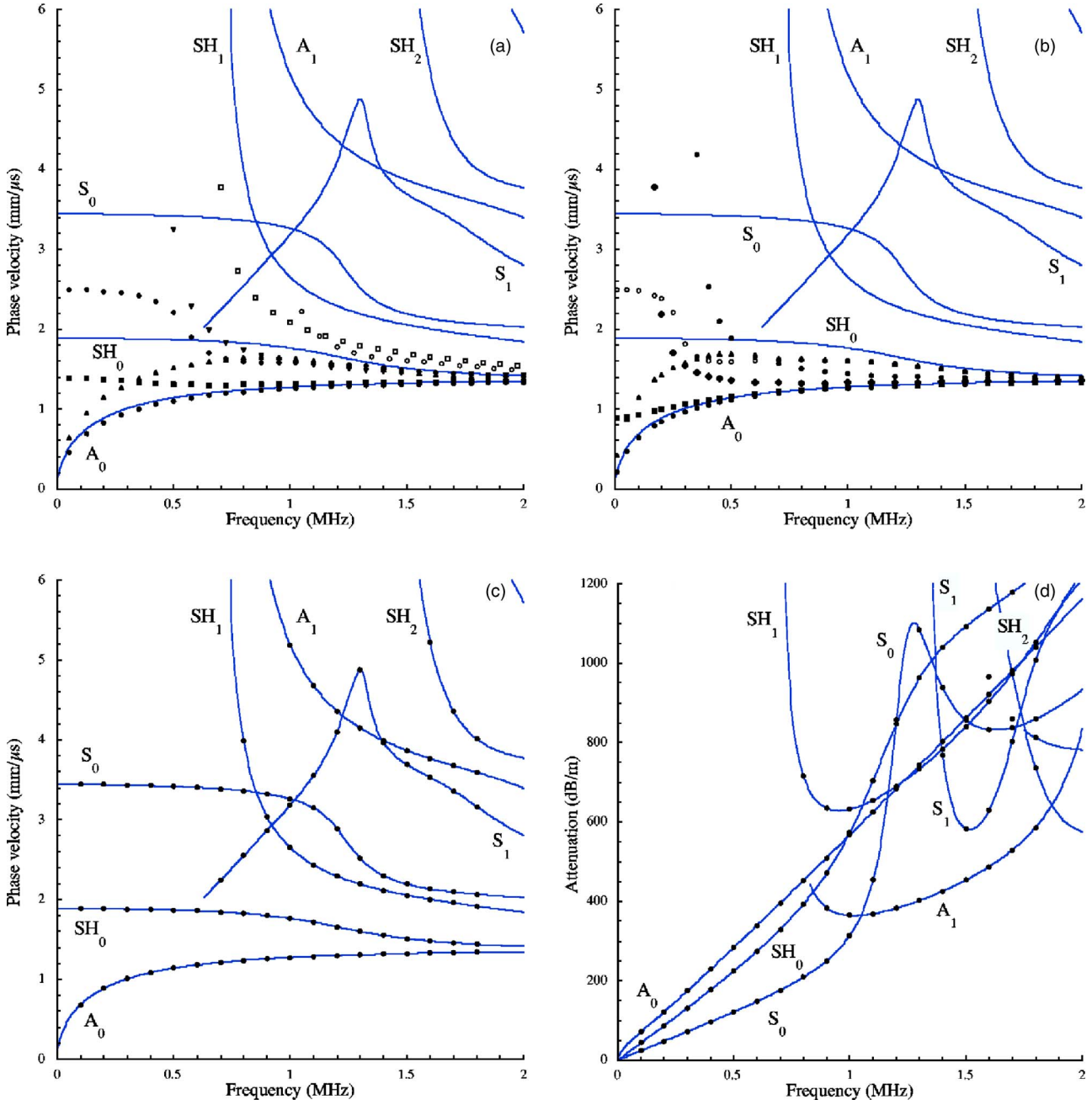


FIG. 2. (Color online) Dispersion curves (phase velocities or attenuations) for Lamb and SH-like modes propagating along a nonprincipal plane of an anisotropic and viscoelastic material plate; (●): data obtained by 2D-SAFE for (a) $B/H=2$ and no PBC, (b) $B/H=4$ and no PBC, (c), (d) any value of B/H and using PBC; (—): data predicted using the semianalytical method described in Ref. 4 and valid for $B/H\rightarrow\infty$ (B: section width, H: plate thickness, PBC: periodic boundary conditions).

modes for elastic and respectively viscoelastic structures and modal sorting according to their nature is the next step, considered as postprocessing. This step is done by following a mode along the frequency range, as long as eigenvalues are not very close or otherwise by modal displacements inspection. In the case of layered structures some eigenvalues can represent particular projections (labeled k_3) of existing wave numbers k, onto the Ox_3 axis. In fact, it can be checked that these projections verify this formula: $k_3^2=k^2-k_1^2=k^2-(2n\pi/B)^2$; $n=1,2,\ldots$. These projections are always obtained as double eigenvalues, since oblique propagating waves can be symmetrically inclined with respect to the Ox_3 axis and are thus easily eliminated. Moreover, these projections depend on B, contrary to the modes propagating along the Ox_3 axis, which is another way to eliminate them.

IV. PERIODICALLY GROOVED PLATE

A. Results obtained with the periodic SAFE method

Periodic planar structures are widely used in technical applications. Two periodicity directions are commonly encountered in planar structures: through-thickness periodic structures such as multilayered composites and along span periodic structures. The second case is addressed in the fol-

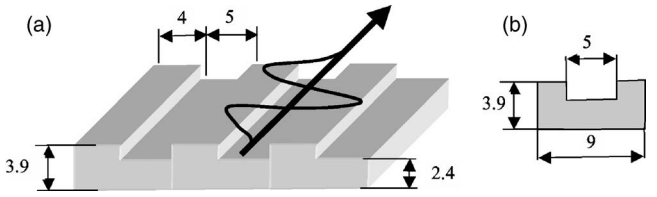


FIG. 3. Geometry of the periodic structure and direction of wave propagation; (a) 3D view of part of the structure and (b) 2D cross section of 1 period (PBCs are applied at each side of this elementary cell). Dimensions are in mm.

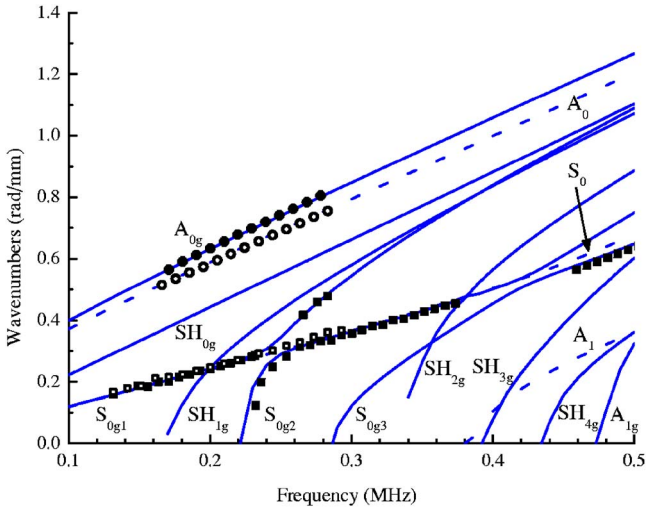


FIG. 4. (Color online) Dispersion curves for modes guided along the aluminum plates; (lines): 2D SAFE predictions using the PBC for the pristine plate (---) and for the grooved plate (—), (dots): experimental data with (○) for A_0 and (□) for S_0 along the pristine plate, and (●) for A_{0g} and (■) for S_{0g1}, S_{0g2} or S_{0g3} along the grooved plate.

lowing, with the particularity of having the wave propagation direction perpendicular to the material and/or geometrical periodicity [Fig. 3(a)]. The dispersion curves are determined for a periodically grooved aluminum plate having geometric data indicated in Fig. 3(a). The periodicity conditions are implemented in the SAFE method and applied at each lateral boundary of the elementary cell of the waveguide as shown in Fig. 3(b). The elasticity coefficients for the aluminum plate of uniform thickness 3.9 mm, called pristine plate, have been determined by best fit of the predicted phase velocities of the Lamb wave modes A_0 and S_0 with experimental data, according to the technique described in Ref. 30. The values thus determined are $C_{11}=105$ GPa, $C_{66}=25$ GPa, and mass density is $\rho=2720$ kg/m^3.

The frequency range 0.1–0.5 MHz has been selected for the interesting feature of the guided waves. In this frequency range there are only three Lamb waves for the pristine plate (A_0, S_0, and A_1). As shown in Fig. 4, the grooved plate reveals interesting changes in the dispersion curves. The A_{0g} mode of the grooved plate has higher wave numbers than those of the A_0 mode of the pristine plate, which can be easily explained by the reduction of the mean thickness. In fact, the wave number values for the grooved plate lie between those of the pristine plate and those of a 2.4-mm-thick plate, which is the thickness of the structure in the grooved regions. More interesting is the splitting phenomenon of the S_0 mode of the pristine plate, into three branches called S_{0g1}, S_{0g2} and S_{0g3}, respectively. In the low frequency regime, the dispersion curve of the S_{0g1} mode of the grooved plate follows that of the S_0 mode of the pristine plate, up to

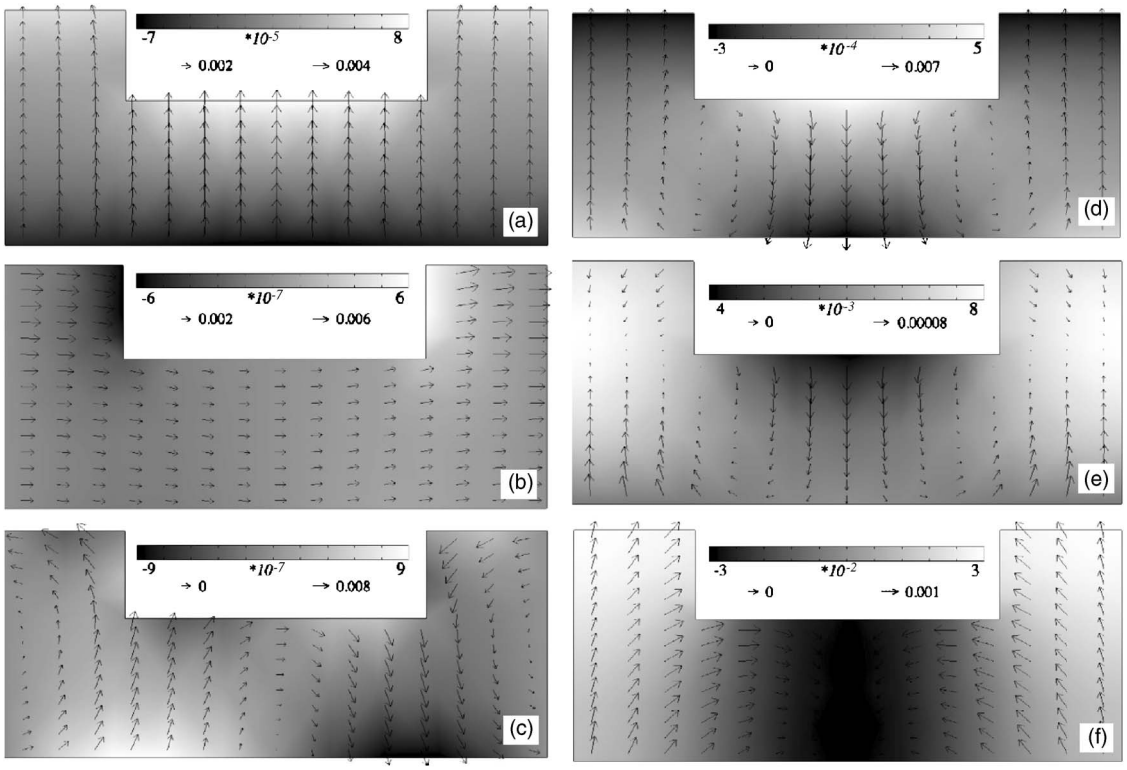


FIG. 5. Displacement mode shapes for guided waves at 300 kHz: (a) A_{0g} with $k_{A0g}=847.8$ m^{-1}, (b) SH_{0g} with $k_{SH0g}=662.6$ m^{-1}, (c) SH_{1g} with $k_{SH1g}=580.5$ m^{-1}, (d) S_{0g1} with $k_{S0g1}=545.5$ m^{-1}, (e) S_{0g2} with $k_{S0g2}=363.8$ m^{-1} and (f) S_{0g3} with $k_{S0g3}=122.9$ m^{-1}. Gray scale indicates the amplitude of U_3 displacements and arrows show the resulting U_1 and U_2 displacements.

0.24 MHz. Then it becomes strongly dispersive beyond this frequency, having wave number values increasing with frequency, whereas a new mode called S_{0g2} takes the relay from its cutoff frequency and is following up the pristine plate mode S_0 up to 0.42 MHz. Another new mode with a cutoff close to 0.285 MHz, called here S_{0g3}, reaches the curve of the pristine plate S_0 mode, after 0.42 MHz.

The displacement patterns for the guided modes at 0.3 MHz are presented in Fig. 5. The highest wave number corresponds to an antisymmetric mode with quasi identical U_2 displacements, along the Ox_2 direction, and for any given x_1 position [Fig. 5(a)]. This mode exhibits a flexural motion and was therefore named A_{0g}, similarly to the fundamental antisymmetric mode of the pristine plate. Using a similar remark for the U_1 displacements along the Ox_1 direction, the next mode can be called SH_{0g} [Fig. 5(b)]. Then Fig. 5(c) displays x_1 periodically localized torsion motions which represent considerable modifications of a shear horizontal mode, which is called here SH_{1g} mode. To our knowledge, this displacement field is reported for the first time in this study, and remains to be confirmed by future studies.

The other new feature predicted by using this method, which is the splitting of the classical S_0 mode of the pristine plate into several branches, is due to the geometrical periodicity (see Fig. 4). The mode called S_{0g1} has important longitudinal (U_3) displacements along the Ox_3 axis [see gray scale level in Fig. 5(d)] and comparable U_2 displacements in the Ox_2 direction (see arrow scale on same figure). These vertical displacements U_2 do not keep the same sign along the width of the U-shaped cell (at a given position x_2), as it is usually the case for symmetric-like modes in plates. The modes called S_{0g2} and S_{0g3} also produce significant displacements in the axial direction Ox_3. The S_{0g2} mode produces more uniform longitudinal (U_3) displacements over the cross section, than the two other branches, i.e., than S_{0g1} and S_{0g3}, corresponding thus more closely to the S_0 mode of the pristine plate. The S_{0g3} mode produces quasi-uniform longitudinal displacements across the thicker regions of the U-shaped cell, which are of opposite phase in comparison to the quasi-uniform displacements produced in the thinner region. It has also noticeable U_1 displacements in the cross-section plane, especially in the thinner domain. These properties make this mode a potential candidate for structural health monitoring (SHM) of periodically reinforced structures. The important longitudinal displacement fields of the three modes S_{0gi} ($i=1,2,3$) suggest the use of a contact transducer coupled to one edge of the plate for launching them in experiments as it is presented in the next section.

B. Experimental data on the grooved plate

The experimental setup shown in Fig. 6 consists of two capacitive, circular, air-coupled transducers of 45 mm in diameter, one piezoelectric, contact, rectangular transducer 150 mm long and 40 mm wide, plus a specific positioning device and electronics for setting up the system, displacing the air-coupled transducers, exciting the transmitters, filtering, amplifying and storing signals for postprocessing. Guided waves are generated either by one properly inclined air-coupled transducer for the antisymmetric-like modes, or by the contact piezoelectric transducer for the symmetric-like modes. The receiver is systematically a properly oriented air-coupled transducer.

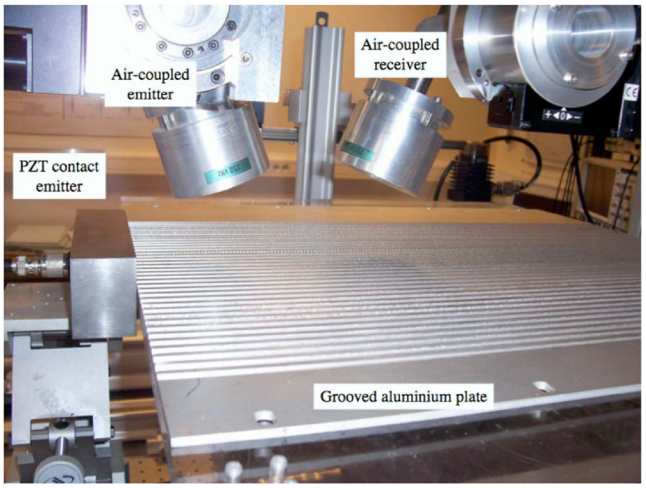


FIG. 6. (Color online) Experimental setup used for generating-detecting guided modes along the pristine or the grooved plates.

A 10-cycle toneburst, with different central frequencies set equal to 180, 220 or 260 kHz depending on the aimed mode(s), was used for exciting the transmitters. The air-coupled receiver was moved along the path of propagation and a series of signals were captured for various positions of this element, in order to process a double Fourier transform of these signals, thus providing a frequency-wave number diagram, as described in previous papers.[30] Two series of data have been acquired: one for the pristine aluminum plate and the second for the grooved plate both on the plane side and on the grooved side of the sample. The first series has been used for characterizing the aluminum sample properties (C_{11} and C_{66}). The measured wave numbers are plotted in Fig. 4 as empty circles for A_0 and empty squares for S_0, respectively, and compared to the wave numbers calculated with the optimized C_{IJ}. The wave numbers of the A_{0g} mode produced along the grooved plate have also been measured using the two air-coupled transducers. These experimental data perfectly confirm the numerical predictions made with the 2D SAFE method using the PBC. Wave numbers have also been measured for the S_{0g1}, S_{0g2} and S_{0g3} modes. These experimental data are in very good agreement with the numerical predictions too, and confirm the splitting of the dispersion curve of the S_0 mode into three branches. These results can be considered as sufficiently relevant for the validation of the SAFE-PBC method.

V. REINFORCED CONCRETE SLABS

Many applications of the SAFE-PBC method can be foreseen. One of them can be the NDE of reinforced concrete slabs with steel bars. The example investigated in this paragraph uses a geometric configuration chosen for the noticeable influence of the steel bars on the concrete slab dispersion curves. The 50-mm-thick and 100-mm-wide periodic cell being presented in Fig. 7 represents this infinite wide structure. The steel bars are 20 mm in diameter and their

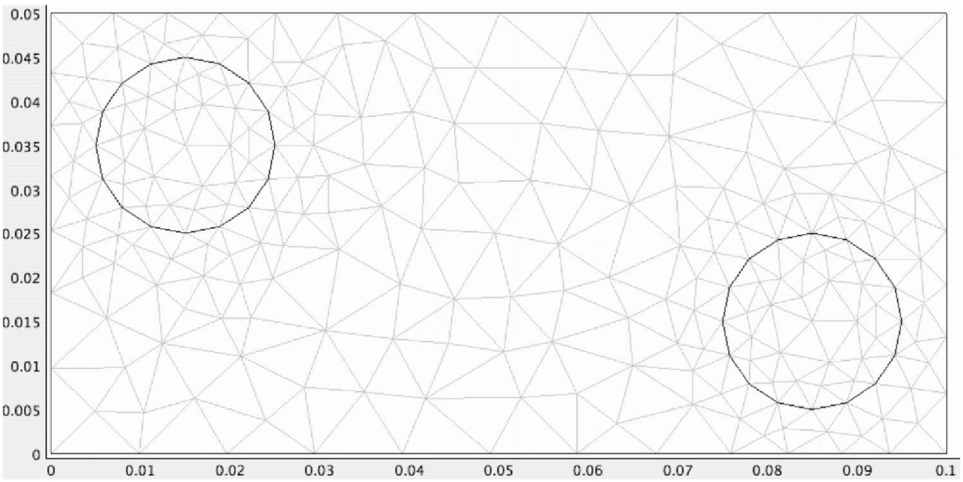


FIG. 7. Geometric parameters of the reinforced concrete slab (values along the axis are in meters).

axes are separated by 70 mm in the direction Ox_1 and by 20 mm in the direction Ox_2. This corresponds to a cross-sectional area ratio of 12.5% between steel domain and concrete domain, which allows emphasizing the influence of the steel bars on the dispersion curves of modes guided along the structure. Lower ratios have indicated negligible changes in the dispersion curves of the modes guided along the nonreinforced concrete.

The material data used in these numerical simulations are those of concrete,[31] which is assumed to be homogeneous and dissipating material: $C_{11}=41+i0.82$ GPa, $C_{66}=16+i0.32$ GPa, mass density 2300 kg/m^3. The elastic properties of the steel bars are $C_{11}=280$ GPa, $C_{66}=80$ GPa and mass density 7900 kg/m^3. Two numerical simulations have been run, one for the concrete slab without reinforcing, and the second one for the reinforced slab defined in Fig. 7. The results of the first simulation are plotted as continuous lines in Fig. 8 and for the second one by dots on the same figure.

Figure 8 shows the influence of the steel bars on the dispersion curves of modes that would exist in the not-reinforced slab. Concerning the real parts of wave numbers, the smallest influence is on the A_0 mode, which is practically not affected by the presence of the bars, in the whole frequency range of investigation (0–50 kHz). The SH_0 mode gets higher wave number values and the S_0 mode gets lower wave number values under the influence of the reinforcement. The SH_1 mode gets lower wave number values in the vicinity of its cutoff frequency, and keeps practically the same values as the frequency increases without reinforcement. Only solutions with ratios between the imaginary and real parts of the wave numbers lower than 10% have been selected. It is also interesting to note that the attenuation remains quite small (lower than 15 dB/,m) for three of the considered modes below 35 kHz, thus indicating that a very low frequency regime is suitable for inspecting large reinforced concrete structures using ultrasonic guided waves. In such NDE purposes, the S_0-like mode seems to be a good candidate since its attenuation doubles, at 30 kHz, when the material properties of the bars are changed from those of steel to those of concrete. Therefore, it may be expected that the attenuation of this mode be sensitive to the quality of the material constituting the reinforcing bars, e.g., if corroded or not.

The cross-section displacement fields for the first three modes at 30 kHz are presented in Fig. 9. The arrows indicate displacements in the cross-section plane (U_1 and U_2), and the gray scale indicates displacements U_3 along the direction of propagation. The amplitudes of these U_3 motions are not comparable from one mode to the other since there is no power normalization, for example. However, they indicate the behavior of the modes. In Fig. 9(c), it is visible that the

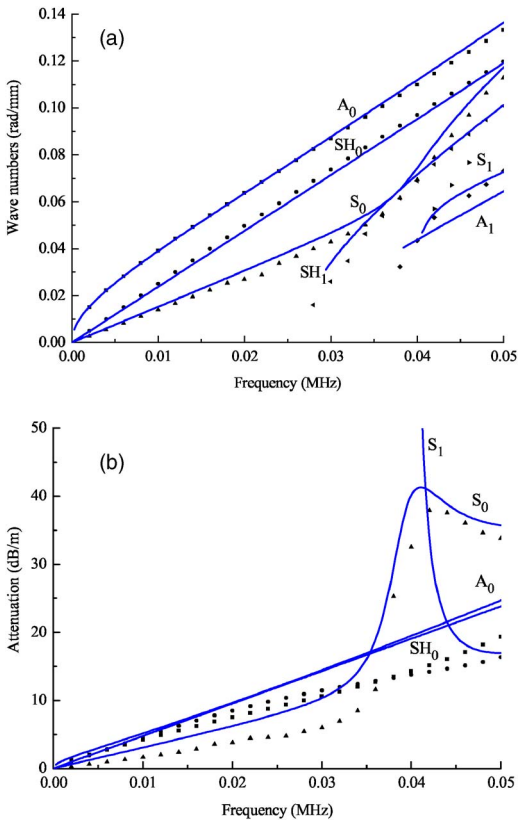


FIG. 8. (Color online) Dispersion curves of modes guided along the concrete slab (continuous lines) and along the reinforced concrete slab (dotted lines); (a) wave numbers and (b) attenuation.

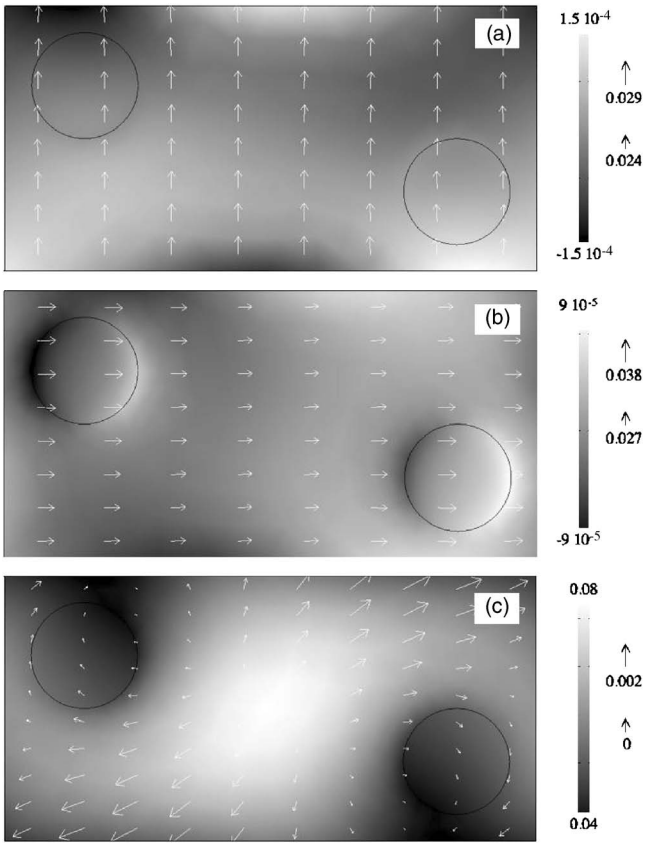


FIG. 9. Modal shapes for the reinforced concrete slab at 30 kHz for (a) A_0-like mode with $k_{A0}=87.1-i0.53$ m^{-1}, (b) SH_0-like mode with $k_{SH0}=73.8-i0.57$ m^{-1} and (c) S_0-like mode with $k_{S0}=43-i0.35$ m^{-1}.

U_3 motion is almost uniform across the concrete, and smaller in the steel bars with no change in sign in the whole section of the structure. This is typical of a longitudinal mode. Therefore, we can say that this mode is a S_0-like mode. The cross-section U_1 and U_2 displacements displayed in Figs. 9(a) and 9(b) show that these two modes are similar to A_0 and SH_0, respectively. The A_0-like mode is practically not affected by the presence of the steel bars, since its three displacement components present no significant variations across the concrete-steel interfaces. However, the SH_0-like mode is slightly sensitive to the presence of the bars, since its U_3 displacement component changes at these interfaces. These wave modes could not be predicted by using plane-strain based models.

VI. CONCLUSIONS

The propagation of guided wave modes along structures having periodic distributions of geometrical and/or of material properties along their width, which is normal to the direction of propagation and to the thickness, was investigated. The proposed method is an extension of the standard 2D-SAFE method, which is restricted to structures of finite widths, since the periodicity imposed along the width of the structures makes them infinitely wide. The periodicity conditions implemented in the SAFE method offer new possibilities to obtain the dispersion curves for a general class of laminar waveguides. This is particularly interesting for studying guided wave modes that produce energy across the whole thickness of the structure (as Lamb or SH-like modes, for example), but crossways part of the width only, so without interacting with the sides of the structure. In addition, this method also allows changes in the geometry and/or in the material properties through the thickness to be considered, as done by many authors.

The propagation of Lamb and SH wave modes along an anisotropic, viscoelastic, composite plate of infinite width has been investigated, as a validation case. The phase velocities and attenuations of the modes have been predicted and successfully compared to those calculated by a plane-strain, semianalytical model. Periodically reinforced planar structures have also been investigated. First, the case of a grooved, aluminum plate has been considered. The dispersion curves predicted with the extended 2D SAFE method have been validated by experimental data obtained with contact and air-coupled transducers set on a specifically machined sample. The results showed the original effect of the grooves on the dispersion curves, e.g., the splitting of the fundamental symmetric mode of the aluminum plate of uniform thickness into several branches. The second case consisted of a concrete slab reinforced by cylindrical steel bars. In these two cases of geometric periodicity along the width, the mode shapes across the elementary cell have been presented for some modes at given frequencies, for better understanding the mechanical behavior of the structure, and therefore the nature of the modes. The results obtained in this paper emphasize the potentialities of the method for a wide variety of SHM or NDE applications based on the propagation of ultrasonic guided waves.

ACKNOWLEDGMENTS

This work was partially supported by the Romanian Ministry of Education and Research under the Research of Excellence Program-Contract No. 6110, CEEX-SINERMAT.

[1]M. J. S. Lowe, "Matrix technique for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control vol. **42**, 525–542 (1995).
[2]W. T. Thomson, "Transmission of elastic waves through a stratified medium," J. Appl. Phys. **21**, 89 (1950).
[3]N. A. Haskell, "The dispersion of surface waves on multilayered media," Bull. Seismol. Soc. Am. **43**, 17–34 (1953).
[4]B. Hosten and M. Castaings, "Surface impedance matrices to model the propagation in multilayered media," Ultrasonics **41**, 501–507 (2003).
[5]S. Dong and R. Nelson, "On natural vibrations and waves in laminated orthotropic plates," J. Appl. Mech. **39**, 739–745 (1972).
[6]R. Nelson and S. Dong, "High frequency vibrations and waves in laminated orthotropic plates," J. Sound Vib. **30**, 33–44 (1973).
[7]S. Dong and K. Huang, "Edge vibrations in laminated composite plates," J. Appl. Mech. **52**, 433–438 (1985).
[8]S. Datta, A. Shah, R. Bratton, and T. Chakraborty, "Wave propagation in laminated composite plates," J. Acoust. Soc. Am. **83**, 2020–2026 (1988).
[9]Z. Xi, G. Liu, K. Lam, and H. Shang, "Dispersion and characteristic surfaces of waves in laminated composite circular cylindrical shells," J. Acoust. Soc. Am. **108**, 2179–2186 (2000).
[10]J. M. Galan and R. Abascal, "Numerical simulation of Lamb wave scattering in semi-infinite plates," Int. J. Numer. Methods Eng. **53**, 1145–1173 (2002).
[11]J. M. Galan and R. Abascal, "Lamb mode conversion at edges. A hybrid boundary element-finite-element solution," J. Acoust. Soc. Am. **117**, 1777–1784 (2005).
[12]P. J. Shorter, "Wave propagation and damping in linear viscoelastic laminates," J. Acoust. Soc. Am. **115**, 1917–1925 (2004).

[13]I. Bartoli, A. Marzania, F. L. di Scalea, and E. Viola, "Modeling wave propagation in damped waveguides of arbitrary cross-section," J. Sound Vib. **295**, 685–707 (2006).
[14]P. E. Lagasse, "Higher-order finite-element topographic guides supporting elastic surface waves," J. Acoust. Soc. Am. **53**, 1116–1122 (1973).
[15]L. Gavric, "Finite element computation of dispersion properties of thin walled waveguides," J. Sound Vib. **173**, 113–124 (1994).
[16]L. Gavric, "Computation of propagative waves in free rail using a finite element technique," J. Sound Vib. **185**, 531–543 (1995).
[17]T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," Ultrasonics **41**, 175–183 (2003).
[18]T. Hayashi, C. Tamayama, and M. Murase, "Wave structure analysis of guided waves in a bar with an arbitrary cross section," Ultrasonics **44**, 17–24 (2006).
[19]U. Orrenius and S. Finnveden, "Calculation of wave propagation in rib stiffened plate structures," J. Sound Vib. **198**, 203–224 (1996).
[20]T. Mazuch, "Wave dispersion modelling anisotropic shells and rods by the finite element method," J. Sound Vib. **198**, 429–438 (1996).
[21]S. Finnveden, "Spectral finite element analysis of the vibration of straight fluid-filled pipes with flanges," J. Sound Vib. **199**, 125–154 (1997).
[22]W. X. Zhong and F. W. Williams, "On the direct solution of wave propagation for repetitive structures," J. Sound Vib. **181**, 485–501 (1995).
[23]L. Gry and C. Gontier, "Dynamic modeling of railway track: A periodic model based on a generalized beam formulation," J. Sound Vib. **199**, 531–558 (1997).
[24]B. R. Mace, D. Duhamel, M. J. Brennan, and L. Hinke, "Finite element prediction of wave motion in structural waveguides," J. Acoust. Soc. Am. **117**, 2835–2843 (2005).
[25]C. H. Yang, H. Huh, and H. T. Hahn, "Investigation of effective material properties in composites with internal defect or reinforcement particles," Int. J. Solids Struct. **42**, 6141–6165 (2005).
[26]J.-M. Mencik and M. N. Ichchou, "Multi-mode propagation and diffusion in structures through finite elements," Eur. J. Mech. A/Solids **24**, 877–898 (2005).
[27]D. R. Mahapatra and S. Gopalakrishnan, "A spectral finite element model for analysis of axial-flexural-shear coupled wave propagation in laminated composite beams," Compos. Struct. **59**, 67–88 (2003).
[28]A. Chakraborty and S. Gopalakrishnan, "A spectrally formulated finite element for wave propagation analysis in layered composite media," Int. J. Solids Struct. **41**, 5155–5183 (2004).
[29]COMSOL, User's Guide and Introduction. Version 3.2 by—COMSOL AB 2005 http://www.comsol.com/ Accessed on 3/5/2007.
[30]B. Hosten, M. Castaings, H. Trétout, and H. Voillaume, "Identification of composite materials elastic moduli from Lamb wave velocities measured with single-sided, contact-less, ultrasonic method," *Review of Progress in Quantitative Non Destructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, AIP Conf. Proc., New York, vol. **20.B**, p. 1023–1030 2001.
[31]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).

Phased array focusing with guided waves in a viscoelastic coated hollow cylinder

Wei Luo[a] and Joseph L. Rose
Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 26 August 2006; revised 25 January 2007; accepted 1 February 2007)

Guided wave phased array focusing has shown many advantages in long-range pipeline inspection, such as, longer inspection distance, greater wave penetration power and higher detection resolution. Viscoelastic coatings applied to a large percentage of pipes for protection purposes created some challenges in terms of focusing feasibility and inspection ability. Previous studies were all based on bare pipe models. In this work, guided wave phased array focusing in viscoelastic coated pipes is studied for the first time. Work was carried out with both numerical and experimental methods. A three-dimensional finite element model was developed for quantitatively and systematically modeling guided waves in pipes with different viscoelastic materials. A method of transforming measured coating properties to finite element method inputs was created in order to create a physically based model of guided waves in coated pipes. Guided wave focusing possibilities in viscoelastic coated pipes and the effects from coatings were comprehensively studied afterwards. A comparison of focusing and nonfocusing inspections was also studied quantitatively in coated pipe showing that focusing increased the wave energy and consequently the inspection ability tremendously. This study provides an important base line and guidance for guided wave propagation and focusing in a real field pipeline under various coating and environmental conditions. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2711145]

PACS number(s): 43.35.Zc, 43.20.Mv, 43.35.Mr [SFW] Pages: 1945–1955

I. INTRODUCTION

Pipelines are used in almost every industry to provide large-scale distribution of products, such as gas, oil, and water. Defective pipelines can cause fatal failures, property damage, and high litigation and replacement costs. When pipelines are aging, inspection and monitoring becomes indispensable. Ultrasonic guided waves, because of their long-range inspection ability, are now being used more and more as a very efficient and economical nondestructive evaluation method for pipeline inspection.[1,2] Current long-range guided wave techniques for pipeline inspection include axisymmetric waves, and nonaxisymmetric waves with partial loading and phased array focusing. Compared with these two techniques, the focusing technique can increase energy impingement, locate defects, and greatly enhance the inspection sensitivity and propagation distance of guided waves, thus consequently reducing inspection costs.

Viscoelastic coatings, such as bitumen and epoxy, are commonly used for protection against corrosion in the pipeline industry. The presence of viscoelastic coating results in changes of guided wave propagation characteristics. Coatings have been used in the pipeline industry since the 1930's and the coating type depends on the date of the coating application.[3] Because of a variation of coating materials and the complexity of the wave mechanics in a viscoelastic coated multilayered structure, many aspects and questions on guided wave inspection in coated pipes are still untouched and remain quite challenging. Such aspects include focusing feasibility, coating effect on focusing, quantitative comparisons of focusing and nonfocusing in terms of inspection ability.

Due to a limitation of experiments because of costs and the difficulty of finding theoretical solutions in viscoelastic coated pipes, numerical modeling studies are strongly called for exploring the potential of long-range guided wave pipeline inspection. In this work, a three-dimensional (3D) finite element method is studied for modeling guided waves in a bare pipe and also for a coated pipe. With help from the developed models, a series of studies are then carried out towards the above aspects.

II. GUIDED WAVES IN PIPES

A. Wave theory

Guided waves in the axial direction of a hollow cylinder include longitudinal and torsional waves with both axisymmetric and nonaxisymmetric modes. Typical dispersion curves in a pipe are plotted in Fig. 1, showing various wave modes in a pipe. Gazis first presented a full solution of all the axisymmetric and nonaxisymmetric wave modes in a hollow cylinder considering a three-dimensional wave field in the radial, angular and axial directions.[4] Later on, Ditri and Rose conducted the calculation of amplitude factors for all of the excited modes corresponding to a specific partial loading condition utilizing the normal mode expansion technique.[5] The study was followed by Li who investigated the field distribution of nonaxisymmetric longitudinal waves coming

[a]Author to whom correspondence should be addressed. Presently with GE Inspection Technologies, 50 Industrial Park Road, Lewistown, PA 17044. Electronic mail: wei.luo1@ge.com

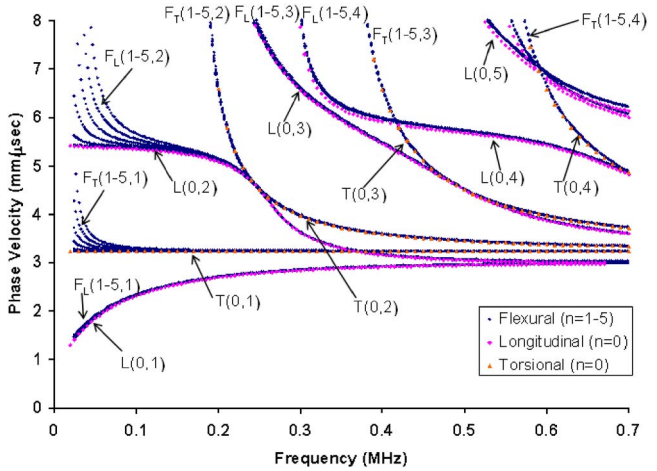


FIG. 1. (Color online) Phase velocity dispersion curves of axisymmetric and flexural modes in 10-in. schedule 40 steel pipe (outer radius = 136.5 mm, thickness =9.27 mm).

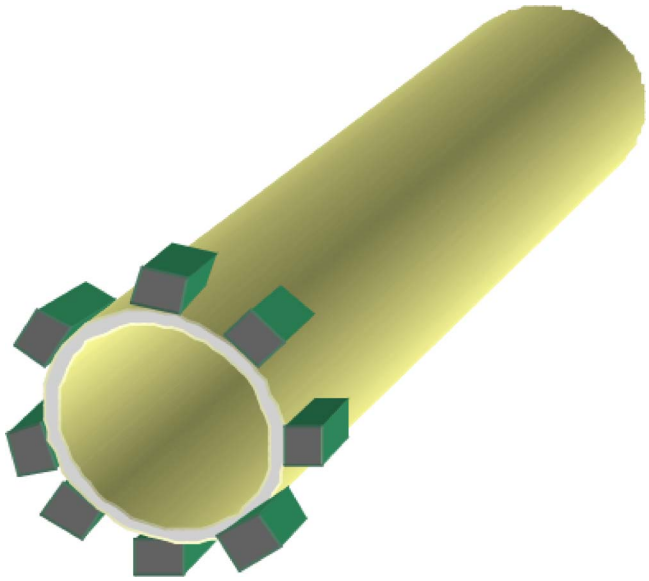


FIG. 2. (Color online) Illustration of an eight-channel guided wave phased array system.

up with an angular profile (overall field distribution for a certain partial source loading) taking into account the amplitude factors of every excited mode calculated with Ditri's method.[6] The angular profile is tunable and therefore can be focused at an expected distance and circumferential angle by using a phased array concept and a deconvolution algorithm for calculating the time delays and weights applied on each array channel.[7] Thereafter, the guided wave phased array mechanics was expanded further to the application of torsional waves.[8]

B. Focusing principle

Phased array focusing is realized by applying input time delays and amplitudes to an *N*-channel phased array. Different from bulk wave phased array ultrasonics, the time delay and amplitude applied on each channel for guided wave phased array focusing in a pipe is a nonlinear function of focal distance, pipe geometry, excitation source, and frequency. There are two key factors to understand the focusing principle of guided waves in a pipe: the angular profile of a single array channel, and the deconvolution algorithm.[6,7] Single channel angular profile can also be interpreted as the wave field distribution (associated with the displacement variation along the circumference of a pipe) under one channel loading condition.

Based on the angular profile of a single channel, focusing can be realized thereafter by using a deconvolution algorithm as shown in Eqs. (1)–(4). As an example, Fig. 2 shows the illustration of an eight-channel phased array system. The angular profiles of a single channel and the focused profile are plotted in Fig. 3.

$$H(\theta) \otimes A(\theta) = G(\theta), \tag{1}$$

$$A(\theta) = G(\theta) \otimes^{-1} H(\theta) = FFT^{-1}(1/\mathrm{H}(\omega)), \tag{2}$$

$$\text{Amplitude } A_i = |A(\theta)|_{\theta=\theta_i}, \tag{3}$$

$$\text{Time delay } \Delta t_i = \text{Phase}\{A(\theta)|_{\theta=\theta_i}\} = -\phi_i/2\pi f, \tag{4}$$

where G is the expected function of focal energy profile, H is the angular profile of a single partially loaded channel, A is the discrete weight function for excitation channels.

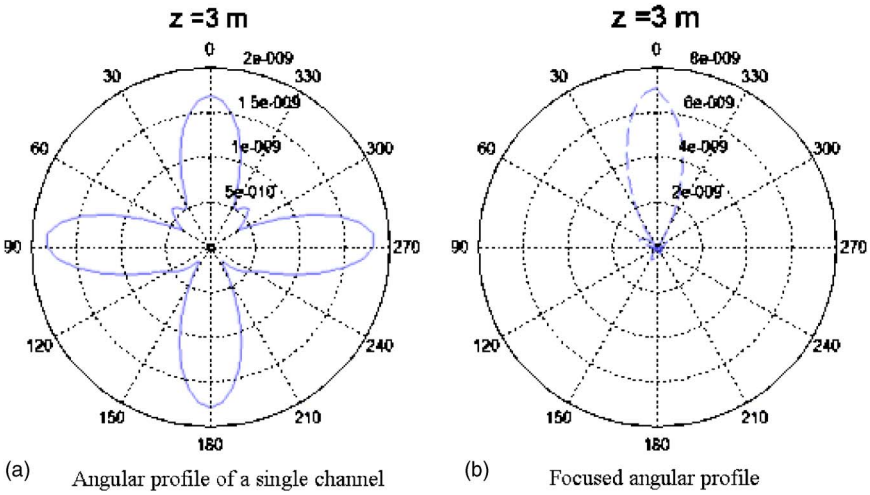


FIG. 3. (Color online) Angular profiles for a 45° single channel loading and eight-channel focusing with the 100 kHz second family longitudinal waves, in a 10 in.-schedule-40 steel pipe. (a) Angular profile of a single channel; (b) focused angular profile.

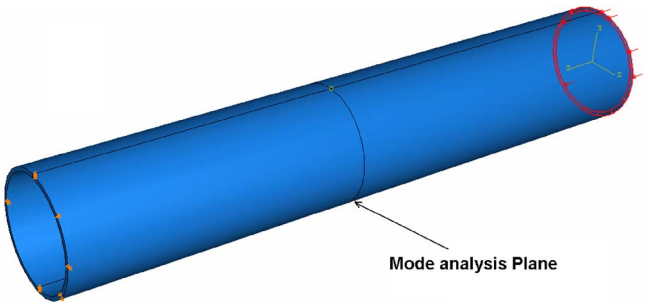


FIG. 4. (Color online) The finite element model for a 10 in.-schedule-40 pipe, with axial pressure loading at the right end. The middle plane is used for signal extraction and model analysis.

C. Viscoelastic multilayer

All the studies shown above on axial-direction guided waves in pipes are all based on an elastic single-layer bare pipe. Wave mechanics in a viscoelastic multilayer is much more complicated in terms of the multilayer problem and also the complex modal roots incurred by the viscoelasticity.[8] There has been some research work reported on guided waves in viscoelastic media,[9–14] such as plate waves in attenuative multilayers and measurement of liquid viscoelastic material property,[9] guided waves in sandwich composite structures,[10] modal wave analysis in composite,[11] and circumferential waves in viscoelastic multilayered pipe,[12] etc. Axial-direction guided waves in a viscoelastic multilayered pipeline were studied by Barshinger utilizing the global matrix method.[13,14] However, due to the complexity of finding full wave modes in a viscoelastic coated pipe, the study is limited to axisymmetric waves that are insufficient for studying phased array focusing in a viscoelastic-coated pipe.

Therefore, numerical modeling, i.e. finite element modeling (FEM), is conducted in this paper as a means of providing an insight into guided wave problems beyond the ability of analytical and experimental measures. Modeling can bring us closer to the optimal parameter choices for an inspection system design. Numerical modeling has been applied in examining guided waves by many researchers in the past two decades. Some typical modeling works on guided waves in pipelines are listed in Refs. 14–20. For example, some interesting studies with finite element modeling of plate waves by Castaings, Bacon and Hosten[15] and axisymmetric pipe torsional wave by Castaings and Bacon[16] in wave guides involving viscoelastic media have been reported. But finite element models of guided wave phased array focusing in a viscoelastic coated pipe have never been studied before. Three-dimensional finite element method based on ABAQUS/Explicit is used in this work to study guided wave propagation and focusing in a pipe. It turns out that guided wave mechanics is not only helpful but also necessary for running proper guided wave finite element models and interpreting the corresponding results.

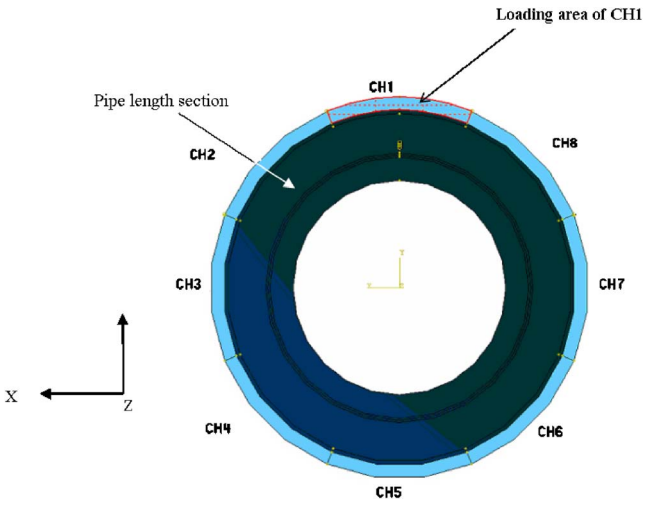


FIG. 5. (Color online) Finite element model for an eight-channel phased array focusing in a 10-in.-schedule-40 steel pipe. Loadings are applied with different time delays and amplitudes to the eight segments at one end of the pipe.

III. FE MODELING IN A HOLLOW CYLINDER

A. Finite element method in dynamics

Three-dimensional finite element methods of structural dynamics are used in this work to study guided wave in a pipe. Starting from Newton's second law, the global form of the finite element (FE) governing equation for dynamics can be acquired in Eq. (5) by using the virtual work principle.[21]

$$[M]\{\ddot{D}\} + [C]\{\dot{D}\} + [K]\{D\} = \{R^{\text{ext}}\}, \tag{5}$$

where $[M]$ is the mass matrix, $[C]$ is the damping matrix, $[K]$ is the stiffness matrix, and $\{R^{\text{ext}}\}$ is the external load, $\{D\}$ is the nodal degree of freedom as functions of time, $\{\dot{D}\}$ and $\{\ddot{D}\}$ are the first and second order derivatives of $\{D\}$, respectively.

Dynamics problems can usually be categorized into two types. One is the wave propagation problem that is related to fast loading and generated modes in a higher frequency range. The other is the structural dynamics problem that is related to much slower loading and lower modes, like the vibration or seismic problem. The explicit direct integration works best for wave propagation problems due to its lower computational cost.[11] In this work, a FE package ABAQUS/Explicit was used for the modeling of guided wave propagation and focusing in pipes.

B. Wave propagation models

Our work began with a simple 3D wave propagation model shown in Fig. 4 for testing the validity and accuracy of this method. The pipe model length is 1.6 m and the pipe wall thickness is 9.27 mm. A linear eight-node brick element (C3D8) is used here in order to reduce the total node number as well as the output file size. The element size is determined by the wavelength (108 mm for 50 kHz longitudinal wave) used in this model. Usually at least ten elements should be used in one wavelength in order to guarantee an effective representation of the wave field change in a wavelength. Given a frequency, the phase velocity dispersion curve could be used here to estimate the wavelength and then to determine the mesh size range. Another factor for mesh size determination is the model geometry. Like the pipe model, the wall is usually very thin, consequently requiring a smaller

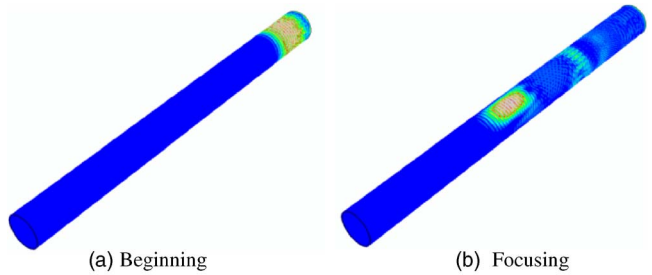


FIG. 6. (Color online) Eight-channel phased array focusing at 0° at a 1.5 meter axial distance in a 3-m-long 10-in.-schedule-40 pipe at a frequency of 100 kHz, showing resulting displacement profiles at the beginning and focusing steps. Focusing realizes a significantly higher Energy (5×) compared to the axisymmetric case, and 20× higher compared to partial loading.

element size. Considering two elements in the pipe wall thickness, therefore, the element size is chosen to be about 5 mm for meshing. For the pipe FE models run by other researchers, the element number from 1 to 3 in the pipe wall thickness was generally used with very good accuracy.[14,17] The total element number for this model is 99600. The computation time of 300 μs wave propagation time is about 10 min using a Dell Precision Workstation with a 3.0 GHz Xeon processor and 3 GB memory.

Basically, there are two ways to apply time-dependent loading in order to generate guided waves. The first one is to simulate the transducer loading behavior by defining a proper boundary condition pattern, called a boundary value problem. It is easy to implement but some unwanted modes may be generated because most of the time the applied loads cannot always match the wave structure of a certain wave mode. The second method is to prescribe the displacement of a cross section at the pipe end with the theoretical wave structure (displacement) of a certain mode. For axisymmetric modes, the wave structure is only a function of the wall thickness position. But for flexural modes, the wave structure is a function of both the radial and the circumferential positions. This method can generate a pure and specific mode by satisfying constraints of the mode. The trade-off is the complicity that may require displacement definition node by node.

The excitation frequency can be realized by using a windowed sinusoidal signal as the time-dependent amplitude of the pressure, also called a loading function. A narrow frequency band is helpful for a purer mode excitation and hence reducing the dispersion effects. However, too many cycles may result in a long time span. Usually 5–15 cycles are used.

The model accuracy was checked through comparing modeling results to analytical solutions, such as group velocity and wave structure. Wave propagation models under different frequencies and modes were carried out and the numerical results showed great consistence with the analytical group velocity and wave structures.

C. Wave focusing models

The finite element model of a multiple-channel focusing system can be realized by segmenting the pipe end and then applying the calculated time delays and amplitudes (Eqs. (1)–(4)) to each segment as shown in Fig. 5. The modeled focusing process was shown in Fig. 6 for zero degrees and a distance of 1.5 m away in a 10-in.-schedule-40 pipe with 100 kHz $L(0,2)$ and higher order flexural waves. The energy is increased tremendously (5×) at the focal point compared with axisymmetric (nonfocusing) loading. Modeling work also shows that focusing increases the energy by 20 times compared with only single channel loading as shown in Eq. (6).

$$\text{Energy ratio:}\quad E_{\text{Single Channel}}:E_{\text{Axisymmetic}}:E_{\text{Focusing}} = 1:4:20. \tag{6}$$

The angular profiles at the focusing distance for one channel (45° partial loading) and eight-channels are calculated by the finite element analysis and shown in Figs. 7(a)

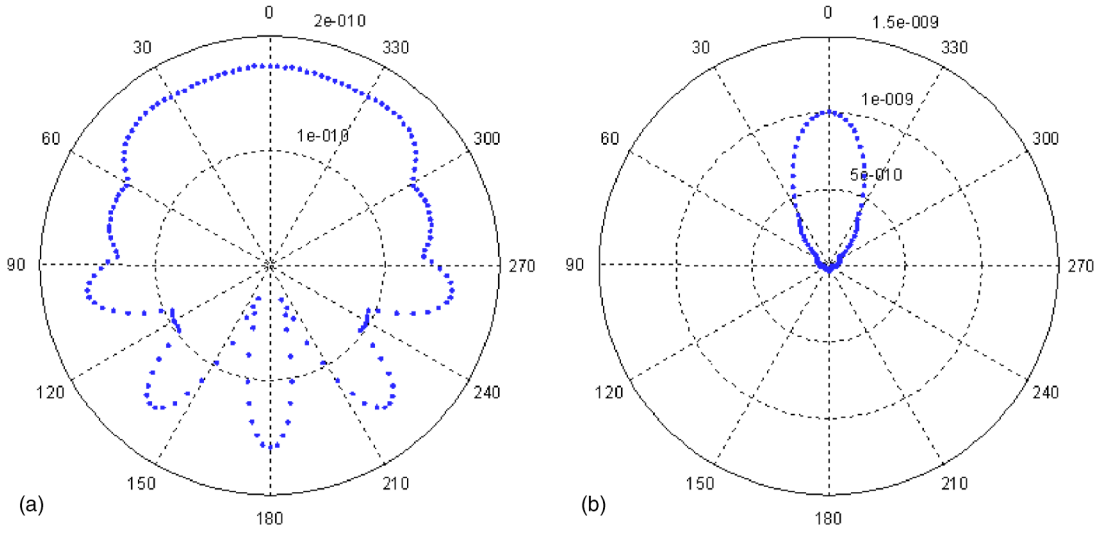


FIG. 7. (Color online) FEM numerical calculation results: angular profiles of the displacement field of the 100 kHz $L(0,2)$ mode at a distance of 1.5 m in a 10-in.-schedule-40 pipe (a) 45° partial loading and (b) eight-segment phased array loading. Note the maximum displacement amplitude in (b) is about 5.8 times larger than that in (a).

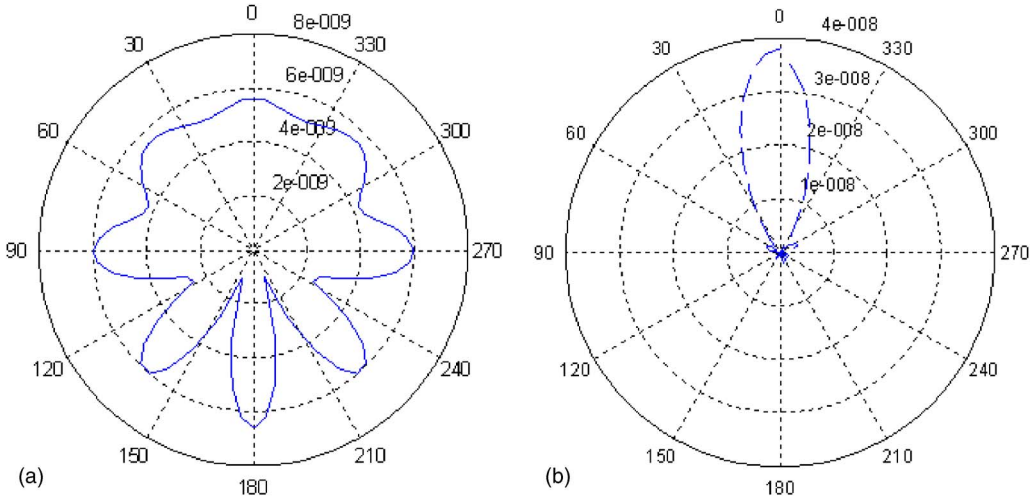


FIG. 8. (Color online) Analytical calculation results: angular profile of the displacement field of the 100 kHz $L(0,2)$ mode at a distance of 1.5 m in a 10-in.-schedule-40 pipe (a) 45° partial loading and (b) eight-segment phase array loading. Note the maximum displacement amplitude in (b) is also 5.8 times larger than that in (a).

and 7(b), respectively. The result is acquired through extracting the wave signals from the wave propagation model and then plotting the signal amplitudes. The angular profiles for the same cases but calculated analytically are shown in Fig. 8. Great consistency has been seen from the FEM numerical results and the analytical results. An amplitude increase (5.8 times) from partial loading to phased array loading can be seen from the numerically calculated angular profile. The result is extremely encouraging in that, at first, it once again demonstrates the accuracy of the finite element model, and secondly it indicates the value of a powerful finite element tool for further research work on wave scattering and wave attenuation using phased array loading.

IV. FE MODELING IN A VISCOELASTIC MULTILAYERED HOLLOW CYLINDER

It has been shown that guided wave focusing increases the inspection sensitivity tremendously. However, a guided wave focusing possibility and focusing profile changes in a coated pipe are still unknown. For the first time, finite element modeling of coated pipe is studied in terms of wave propagation and focusing. The challenge of setting up a physically effective FE model for a viscoelastic coated pipe is how to find appropriate representations of the coating viscoelastic properties as FE model inputs.

A. Coating viscoelasticity and damping

1. Rayleigh damping

Rayleigh damping is used in this paper to introduce the complex elastic modulus constants caused by viscoelasticity of the actual coating material. It is defined by a damping matrix formed as a linear combination of the mass and the stiffness matrices (also shown in Eq. (5)):

$$[C] = \alpha[M] + \beta[K], \tag{7}$$

where α and β are damping factors. With Rayleigh damping, the eigenvectors of the damped system are the same as the eigenvectors of the undamped system. Rayleigh damping can, therefore, be converted into critical damping fractions for each mode. For a mode, the fraction of critical damping can be expressed as[21]

$$\xi = \frac{\alpha}{2\omega} + \frac{\beta\omega}{2}, \tag{8}$$

where the mass proportional Rayleigh damping factor α damps the lower frequencies and the stiffness proportional damping factor β damps the higher frequencies. The α factor simulates the damping caused by the model movement through a viscous fluid and therefore it is related to the absolute model velocities. Since the frequency used in this study is in the ultrasonic wave range (>20 kHz), the contribution from the α_R factor is negligible. The β_R factor defines damping that is related to the material viscous property and proportional to the strain rate.

TABLE I. Elastic and viscoelastic material properties.

Material	c_1 (km/s)	α_1/ω	c_2 (km/s)	α_2/ω	ρ (gm/cm^3)
E & C 2057 / Cat9 Epoxy	2.96	0.0047	1.45	0.0069	1.60
Mereco 303 Epoxy	2.39	0.0070	.99	0.0201	1.08
Bitumastic 50 Coating	1.86	0.0230	0.75	0.2400	1.50

TABLE II. Calculated material damping properties.

Material	E^* (Pa)	G^* (Pa)	damping ratio $\eta=\beta_R\omega$	Damping factor β_R 30 kHz	50 kHz	100 kHz
E & C 2057/Cat9 Epoxy	9.03E9+1.92E8i	3.36E9+6.73E7i	2.12%	1.13E-7	6.76E-8	3.38E-8
Mereco 303 Epoxy	2.95E9+1.92E8i	1.06E9+4.27E7i	3.93%	2.08E-7	1.249E-7	6.25E-8
Bitumastic 50 Coating	2.18E9+7.6E8i	7.66E8+2.85E8i	34.9%	1.85E-6	1.11E-6	5.55E-7

The next question is how to connect the β_R factor with the complex elastic modulus or the acoustic parameters describing wave propagation and attenuation. According to a definition in vibration theory, the damping loss factor η is the ratio between dissipated energy and the input energy. It is expressed in equation (9)[22]

$$\eta=\frac{1}{Q}=2\xi, \tag{9}$$

where Q is the quality factor.

For a time harmonic case, according to the correspondence principle,[23] the stress-strain relationship for a viscoelastic material is changed by using the complex, viscoelastic modulus. Therefore, the complex elastic modulus E^* can be expressed as

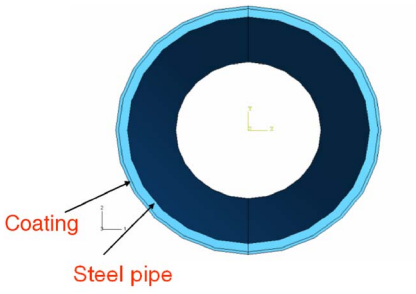
$$E^*=E'+iE'', \tag{10}$$

where E' is the storage modulus that defines the material stiffness, and E'' is the loss modulus that defines the energy dissipation of the material. Therefore, the damping loss factor η can be expressed as the ratio of the loss modulus and the storage modulus.[22]

$$\eta=\frac{E''}{E'}=2\xi=2\left(\frac{\alpha_R}{2\omega}+\frac{\beta_R\omega}{2}\right)\approx\beta_R\omega, \tag{11}$$

where the approximation sign works for the high frequency range of ultrasonic waves.

Then the relationship between Rayleigh damping factor and complex modulus can be expressed in Eq. (12). In next section, the estimation of complex modulus and then Rayleigh damping from measured acoustical properties will be introduced.



2D projection from one pipe end

FIG. 9. (Color online) ABAQUS model for guided wave propagation analysis in a 10-in.-schedule-40 pipe coated with 3 mm viscoelastic materials.

$$\beta_R\approx\frac{2\xi}{\omega}=\frac{E''}{\omega E'}. \tag{12}$$

2. Viscoelastic property estimation from acoustic measurement

For one-dimensional wave propagation in a viscoelastic material, the wave equation is expressed in Eq. (13), where u is the displacement and c is the complex wave velocity[23,25]

$$\frac{d^2u}{dx^2}=\frac{1}{c^*(i\omega)^2}\frac{d^2u}{dt^2}. \tag{13}$$

The solution to Eq. (13) in terms of the attenuation and phase velocity is

$$u(x,t)=Ae^{i(\omega t-k^*x)}=Ae^{i(\omega t-(k'+ik'')x)}=Ae^{k''x}e^{i(k'x-\omega t)}$$
$$=Ae^{-\alpha(\omega)x}e^{i\left(\frac{\omega}{c(\omega)}x-\omega t\right)}, \tag{14}$$

where k is the wave number, $'$ and $''$ indicates the real part and the imaginary part, respectively, and $\alpha(\omega)$ is the attenuation coefficient. Therefore, the imaginary part and the real part of the complex wave number can be expressed as in Eqs. (15) and (16), respectively.

$$k'=\left[\frac{\omega}{c(\omega)}\right]=\left[\frac{\omega}{c^*(\omega)}\right]' \tag{15}$$

$$k''=-\alpha(\omega)=\left[\frac{\omega}{c^*(\omega)}\right]''. \tag{16}$$

Consequently, the complex velocity $c^*(\omega)$ can be derived from Eqs. (15) and (16)

$$c^*(\omega)=\frac{1}{\frac{1}{c(\omega)}-i\frac{\alpha(\omega)}{\omega}}. \tag{17}$$

The velocity is specified to be complex and frequency dependent due to the viscoelastic material properties. Phase velocity $c(\omega)$ and the attenuation constant $\alpha(\omega)$ can be measured by experiments and then $c^*(\omega)$ can be acquired by Eq. (17). The complex shear modulus G^* (also the second Lamé constant μ^*) is calculated as in Eq. (18)[24]

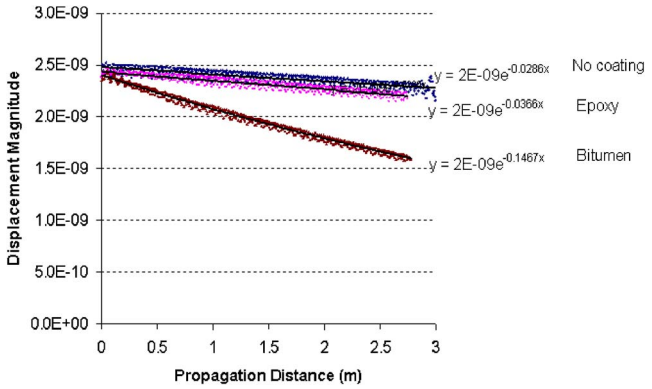


FIG. 10. (Color online) Displacement magnitude vs propagation distance for the $L(0, 2)$ mode at 30 kHz propagation in a 10-in.-schedule-40 pipe.

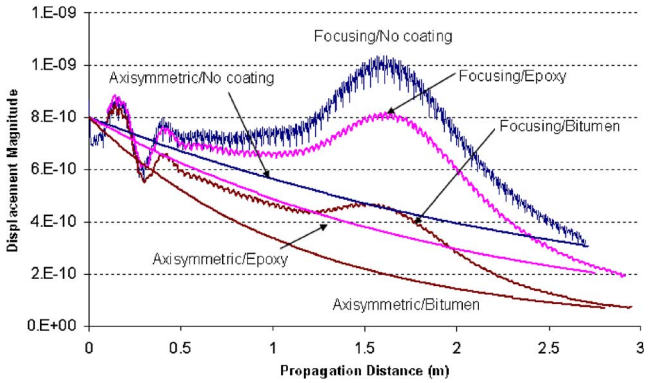


FIG. 11. (Color online) Axial profile for the 100 kHz $L(0, 2)$ wave with axisymmetric and phased array loading with a 1.5 meter focal distance, for no coating, 3 mm epoxy and 3 mm bitumen. Note that focusing increases the magnitude significantly at the focal point.

$$G^* = \mu^* = c_2^{*2} \cdot \rho = \left[\frac{1}{c_2(\omega)} - i\frac{\alpha_2(\omega)}{\omega} \right]^{-2} \cdot \rho = \left(\frac{c_2 \omega}{\omega - i c_2 \alpha_2} \right)^2 \cdot \rho, \quad (18)$$

where the subscript 2 indicates the variables for shear wave and ρ is the density of the material.

Young's modulus is expressed in Eq. (19)

$$E^* = \left[\frac{3 - 4\left(\frac{c_2^*}{c_1^*}\right)^2}{1 - \left(\frac{c_2^*}{c_1^*}\right)^2} \right] G^* = \left[\frac{3 - 4\left(\frac{c_2\omega - i c_1 c_2 \alpha_1}{c_1 \omega - i c_1 c_2 \alpha_2}\right)^2}{1 - \left(\frac{c_2\omega - i c_1 c_2 \alpha_1}{c_1 \omega - i c_1 c_2 \alpha_2}\right)^2} \right] \cdot \left(\frac{c_2 \omega}{\omega - i c_2 \alpha_2} \right)^2 \cdot \rho, \quad (19)$$

where the subscript 1 indicates the variables for longitudinal wave. Therefore, the complex modulus can be calculated from the measured velocities and attenuation for longitudinal and shear waves.

Coating materials used for the pipeline industry varies, depending on the pipeline buried time and the specific condition. Some typical viscoelastic coating materials are selected for this study and their measured material elastic and viscoelastic properties are shown in Table I.[13] The damping properties calculated using Eqs. (11) and (12) are shown in Table II.

B. Wave propagation and attenuation

With the calculated parameters in Table II as inputs to the ABAQUS models, some numerical experiments were carried out for pipes with coatings at several frequencies—30, 50, and 100 kHz. The finite element model of a 10 in. schedule 40 pipe with 3 mm coating is shown in Fig. 9. Linear eight-node brick element (C3D8) is used for this model. In the radial direction, two elements are used for the steel pipe wall and one for the coating layer. The total element number for this model is about 310 k. The computation time of 700 μs wave propagation time is about 50 min using a Dell Precision Workstation with a 3.0 GHz Xeon processor and 3 GB memory. Figure 10 shows the axial profile for the 30 kHz L(0,2) wave in a 10 in. schedule 40 pipe. The axial profile shows the displacement magnitude change along the axial direction at a certain circumferential angle (zero degrees here).

Computations were run on a bare pipe first in order to remove the attenuation from a bare pipe due to dispersion. An exponential curve was fit to the calculated axial profile, thus generating the term $e^{-\alpha(\omega)x}$ in Eq. (14). The attenuation can be calculated following Eqs. (20) and (21)

TABLE III. Attenuation introduced by 3 mm epoxy and bitumastic coating.

Material	Frequency	Attenuation constant α	Attenuation (dB/m)	Propagation distance with 50 dB attenuation (m)
Mereco 303 Epoxy	30 kHz	0.008	0.07	714
	50 kHz	0.0794	0.69	72.5
	100 kHz	0.144	1.25	40.0
Bitumastic 50 Coating	30 kHz	0.118	1.03	48.0
	50 kHz	0.739	6.42	7.9
	100 kHz	0.511	4.44	11.3

TABLE IV. Amplitude gain by focusing compared with axisymmetric waves under different frequencies, focal distances (FDs) and coating conditions.

Material	Gain (dB)		
	Frequency=100 kHz FDs=1.5 m	Frequency=50 kHz FDs=3 m	FD=1.5 m
No coating	8.37	4.51	5.37
Mereco 303 Epoxy	7.08	7.43	5.89
Bitumastic 50 Coating	7.38	5.20	5.67

$$\text{Attenuation (dB)} = 20\log_{10}(e^{-\alpha(\omega)x}) = -\alpha(\omega)x20\log_{10}(e) = -8.69\alpha(\omega)x, \tag{20}$$

$$\text{Attenuation (dB/m)} = -8.69\alpha(\omega). \tag{21}$$

The numerical experiments were also carried out for 50 and 100 kHz with all of the results summarized in Table III. The resulting wave propagation distance under a 50 dB attenuation law, for example, is also estimated based on the calculated attenuation (dB/m), showing that at a very low frequency, like 30 kHz, the wave can propagate much longer than at other frequencies. Results also show that attenuation is a function of material and frequency. For the less viscoelastic epoxy, the attenuation is much less than for the bitumastic coating. It is also noticeable that the attenuation increases monotonically with frequency for epoxy, but not for bitumastic coating whose attenuation at 50 kHz is higher than that at other frequencies. Therefore, in order to estimate the guided wave inspection distance of a coated pipe, it is necessary to analyze the specific situation based on the coating material and various guided wave parameters.

C. Phased array focusing

FEM modeling study of phased array focusing in coated pipes was started with several cases: no coating, a 3 mm epoxy coating and a 3 mm bitumen coating. Figure 11 shows the axial profiles for 100 kHz $L(0, 2)$ wave focusing with a 1.5 meter focal distance. The axial profiles for axisymmetric loading are also plotted for comparison purpose. It is very exciting to see that focusing is realized quite well at the expected distance and that the signal magnitudes are increased tremendously. As an example, at the focal point, the focused wave amplitude for the pipe coated with bitumen is even higher than the wave amplitude in a bare pipe. The quantitative information about the amplitude gain is tabulated in Table IV in which 8 dB gain can be acquired maximally from focusing. Angular profiles at the focal point are also plotted in Fig. 12 comparing with the ones in a bare pipe. Angular profiles are important because they help decide the inspection and circumferential sizing resolution potential. It is very nice to see from the profile after normalization the profile shape stays decently though there is a big amplitude loss.

Modeling work for a focal distance of 3 m was also carried out and the results are shown in Fig. 13, Fig. 14 and Table IV. The angular profiles in coated pipe become a little wider, leading to a little loss of circumferential resolution. The gains caused by focusing are generally smaller than those for the 1.5 meter focal distance and vary the three coating conditions. The gain for the epoxy is much larger than those for the bare pipe and bitumen, indicating focusing gain is also a function of coating materials besides the distance.

More generally, similar modeling work was also conducted for a different frequency, 50 kHz, in order to find the generality of the focusing effect in a coated pipe. Results are shown in Fig. 15, Fig. 16 and Table IV. Note that four-channel focusing was used rather than eight-channel focus-

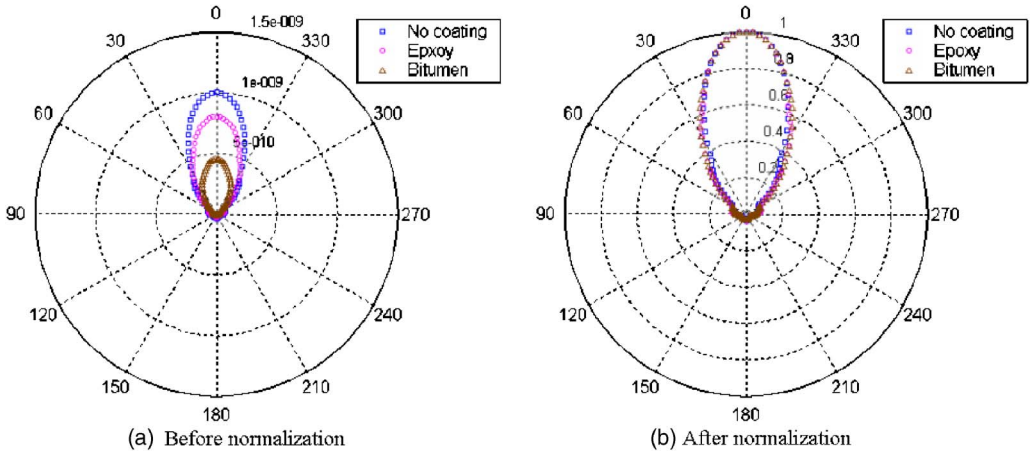


FIG. 12. (Color online) Angular profile at the focal point for a bare pipe and a pipe coated with 3 mm bitumen for 100 kHz longitudinal waves, showing that coating introduces some attenuation but that the profile shape stays the same.

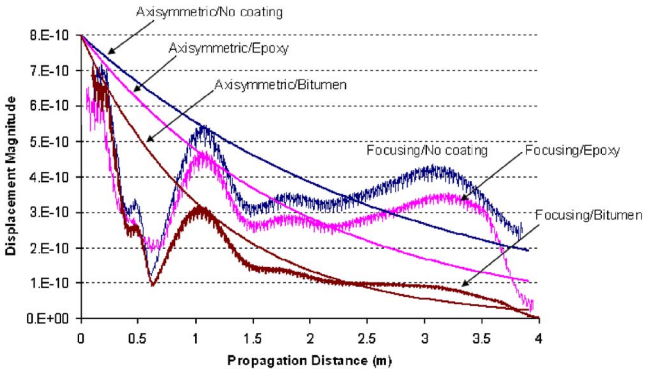


FIG. 13. (Color online) Axial profile for the 100 kHz $L(0, 2)$ wave with axisymmetric and phased array loading (focal distance equal to 3 m), for no coating, 3 mm epoxy and 3 mm bitumen. Note that focusing increases the magnitude significantly at the focal point.

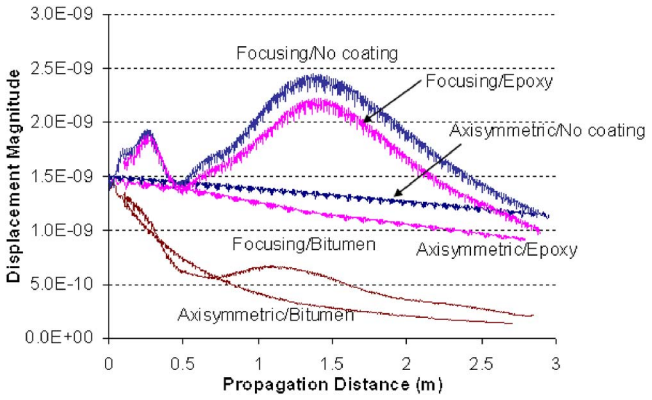


FIG. 15. (Color online) Angular profile for the 50 kHz $L(0, 2)$ wave with axisymmetric and phased array loading (focal distance equal to 1.5 m), for no coating, 3 mm epoxy and 3 mm bitumen. Note that focusing increases the magnitude significantly at the focal point.

ing for the frequency of 50 kHz. This is because modes are separated more in a low frequency range as shown in Fig. 1. In other words, fewer modes are available with slightly different phase velocities, thus resulting in fewer channel numbers. Results show that focusing was realized very well and gains were from 5 to 6 dB. Further study will be carried out to investigate any other possible effects from coating with other frequencies and/or materials and to confirm the conclusion that has been reached so far.

V. EXPERIMENTS

Guided wave experiments were also carried out on a coated pipe with a goal to demonstrate some of the observations and conclusions from the theoretical work. Due to experimental limitations, the experiments did not follow exactly the conditions of previous modeling work. An initial experiment is shown in Fig. 17. A phased transducer array was used as a transmitter on a 16 in. schedule 30 pipe coated with a viscoelastic wax coating with 2 mm thickness and 4 feet in length. A receiving transducer was placed 10 feet away from the transmitter to measure the wave field at various points along the circumference. Both $L(0,2)$ longitudinal waves and $T(0,1)$ torsional waves at 45 kHz were used for the focusing experiment. Figure 18(a) shows the experimental angular profiles for the 45 kHz longitudinal waves focused at a 10 foot distance and at 270° in the bare pipe and also in the coated pipe. An angle beam piezoelectric transducer was used for the longitudinal profile measurement. It can be seen that wave energy was focused at the expected angle for both the no coating and wax coating conditions, although there was some amplitude loss due to the wax coating. Torsional wave angular profiles were measured with a SH EMAT sensor and the results are shown in Fig. 18(b). The wave was once again focused quite well under the coating condition again with an amplitude loss. These two experiments agree quite well with the numerical result previously acquired. The impact of this work on the practical side of things is huge as it will provide us with a measure of the capability to inspect when certain coating conditions are in evidence and tedious trial and error approaches.

VI. CONCLUDING REMARKS

Long-range ultrasonic guided wave inspection of coated pipes has been studied through the use of numerical and experimental methods. First, a powerful 3D finite element tool for modeling guided wave propagation and focusing has been developed utilizing ABAQUS/Explicit. Procedures in-

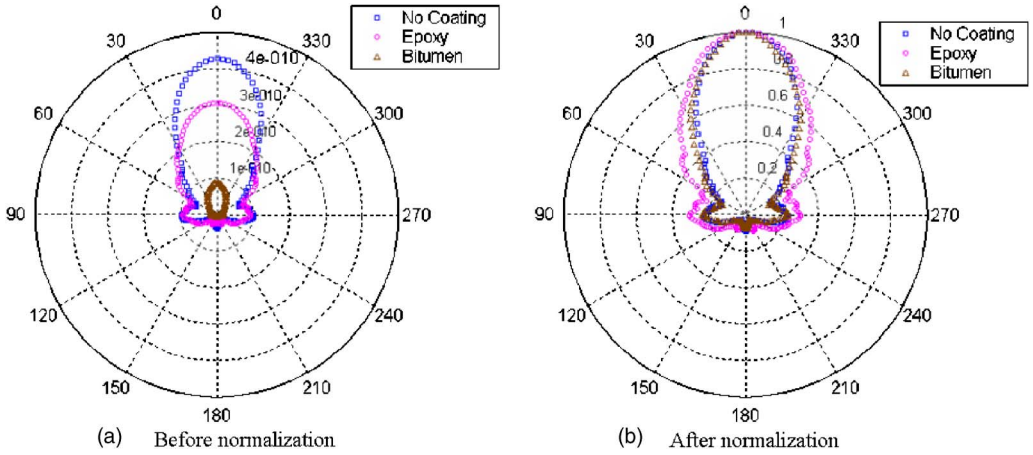


FIG. 14. (Color online) Axial profile for at the focal point for a bare pipe and a pipe coated with 3 mm bitumen for 100 kHz longitudinal waves, showing that coating introduces some attenuation but that the profile shape stays the same.

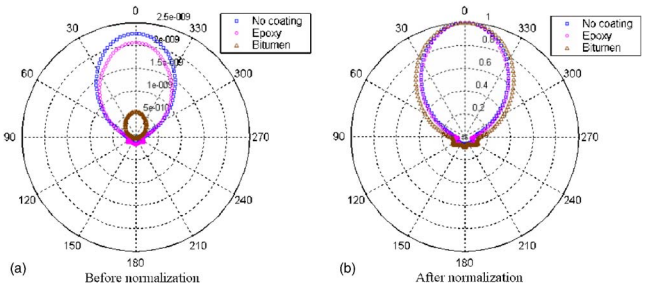


FIG. 16. (Color online) Angular profile at the focal point for a bare pipe and a pipe coated with 3 mm bitumen for 50 kHz longitudinal waves, showing that coating introduces some attenuation but that the profile shape stays the same.

cluding mode generation, data acquisition and result analysis have been established for the modeling and analysis of any wave type in a hollow cylinder. High consistency between modeling and theoretical results proves the validity and high accuracy of the guided wave FE models.

In order to study the coating effects on guided wave propagation and focusing, a two-layer 3D FE model was developed for the first time to study guided wave propagation and focusing in a viscoelastic coated pipe. A transformation algorithm from coating acoustic properties to complex viscoelastic coating properties has been developed to provide inputs for the 3D FE models. Rayleigh damping was utilized to introduce the damping caused by these viscoelastic properties calculated from the measurable acoustic properties. Wave propagation and focusing in pipes coated with any coating materials and any incident wave becomes possible provided the coating property is measurable. Wave propagation modeling for two coating materials shows that wave attenuation increases with frequency and coating viscoelasticity. Moreover, the modeling process also tells us that finite element modeling of guided waves is not just only running ABAQUS software, but also requires a deep understanding of wave mechanics as well as the necessary wave mechanics analytical calculations in providing inputs to the ABAQUS models.

With the FE models, it was the first time to study guided wave focusing in a coated pipe and the coating effect. It was shown by 3D FE modeling that focusing can be realized quite well in a pipe with viscoelastic materials for the frequencies studied. Coating introduces amplitude attenuation but without dramatically changing the focused angular profile, which was also confirmed by experiments. Modeling results show that phased array focusing often increases the wave energy by 9–16 dB for the studied frequencies and distances, indicating a much longer propagation distance and also improved defect detection sensitivity. This conclusion is extremely encouraging and useful for further coated pipe inspection although studies on more waves, frequencies and coating materials are still needed. This accomplishment has a high impact on the research and application of long-range pipeline inspection.

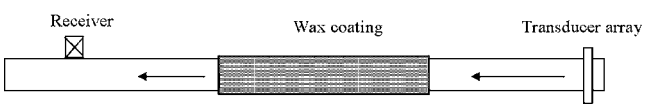


FIG. 17. Angular profile measurement experiment of a 16 in. schedule 30 pipe coated with one ply wax (2 mm in width, 4 feet covered length). A transducer was used as receiver 10 feet away from the transmitter.

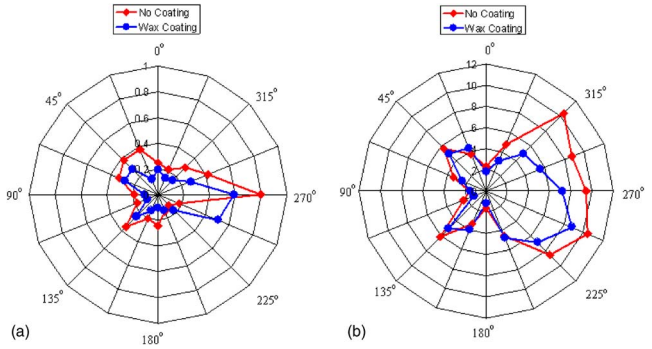


FIG. 18. (Color online) Comparison of angular profiles for uncoated and wax coated pipe when using 45 kHz guided wave focusing at 10 feet at 270°, in a 16-in.-schedule-30 pipe. The coating was one ply, 2 mm wax 4 feet in length. (a) Longitudinal (L [0, 2]), a piezoelectric angle beam transducer and (b) torsional (T[0, 1]), a SH EMAT was used as the receiver.

ACKNOWLEDGMENTS

Thanks are given to FBS, Inc., PA, Plant Integrity, Inc. UK, and the Department of Transportation, USA, for their technical and financial support of the thesis work. We also thank Penn State Colleagues J. Mu and L. Zhang for setting up the experimental system and J. K. Van Velsor for miscellaneous help.

[1]J. L. Rose, "Standing on the shoulders of giants: An example of guided wave inspection," Mater. Eval. **60**, 53–59 (2002).
[2]D. N. Alleyne and P. Cawley, "Long range propagation of lamb wave in chemical plant pipework," Mater. Eval. **45**, 504–508 (1997).
[3]S. Papavinasam and R. W. Revie, "Standards for pipeline coatings," *Workshop on Advanced Coatings for R&D for Pipelines and Related Facilities*, National Institute of Standards and Technology, Gaithersburg, June 9–10 (2005).
[4]D. Gazis, "Three dimensional investigation of the propagation of waves in hollow circular cylinders," J. Acoust. Soc. Am. **31**, 568–578 (1959).
[5]J. J. Ditri and J. L. Rose, "Excitation of guided wave modes in hollow cylinders by applied surface tractions," J. Appl. Phys. **72**, 2589–2597 (1992).
[6]J. Li and J. L. Rose, "Excitation and propagation of non-axisymmetric guided waves in a hollow cylinder," J. Acoust. Soc. Am. **109**, 457–464 (2001).
[7]J. Li and J. L. Rose, "Implementing guided wave mode control by use of a phased transducer array," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 761–768 (2002).
[8]Z. Sun, L. Zhang, B. Gavigan, T. Hayashi, and J. L. Rose, "Ultrasonic flexural torsional guided wave pipe inspection potential," *Proceedings, ASME Pressure Vessel and Piping Division Conference*, **456**, 29–34 (2003).
[9]F. Simonetti, "Sound propagation in lossless waveguides coated with attenuative materials," Ph.D thesis, Imperial College London (2003).
[10]M. Castaings and B. Hosten, "Guided waves propagating in sandwich structures made of anisotropic, viscoelastic, composite materials," J. Acoust. Soc. Am. **113**, 2622–2634 (2003).
[11]W. J. Xu, F. Jenot, and M. Ourak, "Modal waves solved in complex wave number," *Review of Progress in Quantitative Nondestructive Evaluation* (AIP, New York, 2004), vol. **24**, pp. 156–163.
[12]W. Luo, J. L. Rose, J. K. Van Velsor, M. Avioli, and J. Spanner, Circumferential guided waves for defect detection in coated pipe," *Review of Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, (2006), Vol. **25**, pp. 165–172.
[13]J. N. Barshinger and J. L. Rose, "Guided wave propagation in an elastic

hollow cylinder coated with a viscoelastic material," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1547–1556 (2004).
[14]P. Cawley, M. J. S. Lowe, and F. Simonetti, "The variation of reflection coefficient of extensional guided waves in pipes from defects as a function of defect depth, axial extent, circumferential extent and frequency," J. Mech. Eng. Sci. **216**, pp. 1131–1141 (2002).
[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]M. Castaings and C. Bacon, "Finite element modeling of torsional wave modes along pipes with absorbing materials," J. Acoust. Soc. Am. **119**, 3741–3751 (2006).
[17]W. Zhu, "A simulation for guided elastic wave generation and reflection in hollow cylinders with corrosion defects," J. Pressure Vessel Technol. **124**, 108–117 (2002).
[18]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).
[19]T. Hayashi, Z. Sun, J. L. Rose, and K. Kwawshima, Analysis of flexural mode focusing by a semianalytical finite element method," J. Acoust. Soc. Am. **113**, 1241–1248 (2003).
[20]W. Luo, X. Zhao, and J. L. Rose, "A guided wave plate experiment for a pipe," ASME J. Pressure Vessel Technol. **127**, 345–350 (2005).
[21]R. D. Cook, D. S. Malkus, M. E. Plesha, and R. J. Witt, *Concepts and Applications of Finite Element Analysis* (Wiley, New York, 2001).
[22]J. Sun and C. Wang, *Theory of Mechanical Noise Control (in Chinese)* (Northwestern Polytechnical University Press, Xian, China, 1993).
[23]R. M. Christensen, *Theory of Viscoelasticity: An Introduction* (Academic, New York, 1981).
[24]J. Krautkramer and H. Krautkramer, *Ultrasonic Testing of Materials*, 4th ed. (Springer-Verlag, Berlin, 1990).
[25]R. H. Blanc, "Transient wave propagation methods for determining the viscoelastic properties of solids," J. Appl. Mech. **60**, pp. 763–768 (1993).

Determination of a response function of a thermocouple using a short acoustic pulse

Yusuke Tashiro[a)]
Department of Crystalline Materials Science, Nagoya University, Nagoya 464-8603, Japan

Tetsushi Biwa
Department of Mechanical Systems and Design, Tohoku University, Sendai 980-8579, Japan

Taichi Yazaki
Aichi University of Education, Kariya 448-8542, Japan

(Received 19 September 2006; revised 14 December 2006; accepted 8 January 2007)

This paper reports on an experimental technique to determine a response function of a thermocouple using a short acoustic pulse wave. A pulse of 10 ms is generated in a tube filled with 1 bar helium gas. The temperature is measured using the thermocouple. The reference temperature is deduced from the measured pressure on the basis of a laminar oscillating flow theory. The response function of the thermocouple is obtained as a function of frequency below 50 Hz through a comparison between the measured and reference temperatures. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2535534]

PACS number(s): 43.35.Zc, 43.20.Ye, 43.20.Mv, 43.58.Gn [RR] Pages: 1956–1959

I. INTRODUCTION

An acoustic wave in a gas-filled tube can convert acoustic *heat flow* $Q=\rho_m T_m\langle SU\rangle$ into *work flow* $I=\langle PU\rangle$ or vice versa through heat exchange with the tube wall,[1–3] where S, U, P respectively denote the oscillating entropy, velocity, and pressure of the gas. In addition, ρ_m and T_m are the mean density and temperature; angular brackets represent the time average over a period. A method for the accurate determination of I in the tube has been established experimentally through simultaneous measurements of P and U,[4] which has contributed to elucidating thermoacoustic phenomena.[5–7] However, the measurement method of Q in the tube has not been well established because of the difficulty of measuring the oscillating entropy S. Therefore, we specifically address the total energy flow $H=Q+I$, rather than Q itself. The total energy flow H is written as $H=\rho_m C_P\langle TU\rangle$ for an ideal gas, where C_P is an isobaric specific heat and T the oscillating temperature of the gas. Once T is measured, H can be determined and Q is then deduced by subtracting the measured I from H.

The dynamic calibration of a thermometer is necessary for accurate measurements of the time-dependent temperature T of the gas because the amplitude damping and the phase delay are unavoidable in any thermometer. The comparison between an existing technique for a cold wire anemometer and our method is important. However, as long as we know, only a few paper compensate the response of thermometer not only for amplitude but also for phase. Therefore, it is difficult to compare our method to the existing technique. Huelsz and Ramos[8] used a cold wire anemometer for oscillating temperature measurements of an acoustic wave in a tube. They stressed the need to employ a scale corrected for the phase delay of the anemometer. Tagawa *et al.*[9,10] described an experimental method to determine, quantitatively, the time constant τ for the thermocouple to achieve thermal equilibrium with a surrounding gas. However, that study presumes, without experimental verification, that the response function of the thermocouple is given as $Z=(1+i\omega\tau)^{-1}$ used in a first-order lag system, where ω is an angular frequency.

We have reported the dynamic calibration of a thermocouple using temperature oscillations caused by a monofrequency acoustic wave in a tube as the reference temperature.[11] This method determines the response function Z_{ex} as defined in Ref. 11 only for a given frequency f, i.e., $f=\omega/2\pi$. For that reason, many experiments are necessary to determine Z_{ex} for a wide frequency range. In this paper, we report the experimental derivation of Z_{ex} using a single short acoustic pulse wave instead of a continuous wave. Because of the use of the pulse wave, the frequency dependence of Z_{ex} below 50 Hz was obtained very easily. Results show that Z_{ex} is consistent with that determined using a monofrequency acoustic wave,[11] which supports the validity of the pulse method.

II. EXPERIMENTAL SETUP AND PROCEDURE

The present experimental setup is illustrated schematically in Fig. 1. The tube length is 0.6 m, and the internal diameter $2r_0$ is 21 mm. One end of the tube was connected to a buffer tank that was filled with 1.3 bar helium gas via a solenoid valve. A rectangular voltage pulse, typically with 7 ms width, was fed to the solenoid valve to generate an acoustic pulse in the tube. The other end of the tube was connected to an orifice valve. The valve was used to change the width of the acoustic pulse, and to maintain the mean pressure P_m in the tube at atmospheric pressure (101 kPa)

[a)]Author to whom correspondence should be addressed. Electronic mail: tashiro@mizu.xtal.nagoya-u.ac.jp

 0001-4966/2007/121(4)/1956/4/$23.00

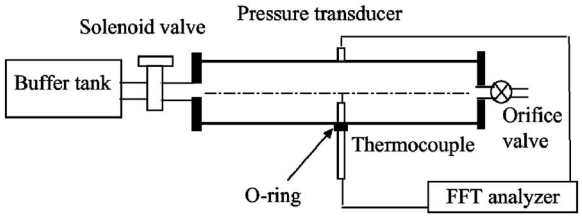


FIG. 1. Schematic illustration of the experimental apparatus.

throughout the experiments. Thermoacoustic effects associated with the continuous acoustic wave invariably cause a significant temperature gradient along the tube axis,[12] but the temperature change of the tube caused by the pulse wave was negligibly small in this experiment. A thermocouple of *K*-type (chromel-alumel) with wire diameter of 15 μm[11] was inserted into the tube through a narrow duct in such a way that the junction was positioned on the central axis of the tube. A small pressure transducer (DD102-1F; Toyoda Machine Works, Ltd.) was mounted on the tube wall at the same axial position where the thermocouple junction was located.

Electrical signals from the pressure transducer and thermocouple were recorded simultaneously with a multichannel 24 bit spectrum analyzer (sampling frequency 400 Hz). From Fourier transforms of their signals, the amplitude and phase spectra of the measured pressure and temperature were determined.

III. EXPERIMENTAL RESULTS

Figure 2(a) shows the measured pressure P associated with the acoustic pulse with 10 ms width. The corresponding temperature T_{ex} with height greater than 4 K was readily apparent, as shown in Fig. 2(b). The tails of their amplitude spectra shown in Figs. 2(c) and 2(d) extend up to 100 Hz, reflecting the width of P and T_{ex} in Figs. 2(a) and 2(b). These extended tails suggest that the response function Z_{ex} of the thermocouple is attainable up to this frequency. However, to maintain a reasonable S/N ratio, frequency components less than 50 Hz were used to determine Z_{ex}. We next consider gas temperature T resulting from the pressure P oscillating with the angular frequency ω before showing Z_{ex} of the thermocouple.

When the tube wall temperature is homogeneous, the temperature T on the central axis of the tube is theoretically given from a laminar oscillating flow theory[1,2] as

$$\frac{T}{T_S} = F, \tag{1}$$

where T_S represents the temperature oscillation in the adiabatic limit and is written for an ideal gas with a specific heat ratio γ as

$$T_S = \frac{\gamma - 1}{\gamma}\frac{T_m}{P_m}P. \tag{2}$$

The complex function F in Eq. (1) is written as

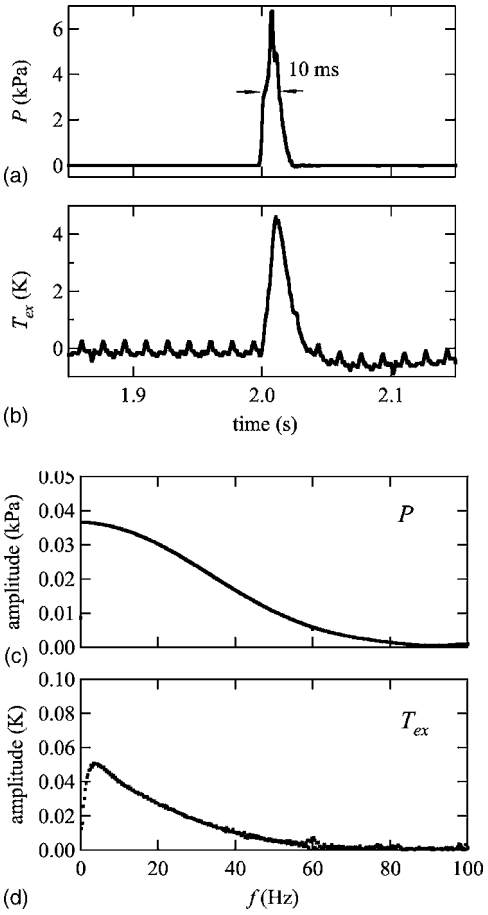


FIG. 2. Time dependence of (a) P and (b) T_{ex} and their amplitude spectra ((c) and (d)). The small periodic signals on T_{ex} in (b) are the 60 Hz ac power line noise.

$$F = 1 - \frac{1}{J_0(\sqrt{2}i^{3/2}r_0/\delta)}, \tag{3}$$

where δ is the thermal penetration depth given as $\delta = \sqrt{2\alpha/\omega}$ (α: thermal diffusivity of a gas), and J_0 is the 0th order complex Bessel function. The absolute value and argument of F, $|F|$ and Arg(F) were calculated from Eq. (3) using the present experimental condition ($\alpha = 1.80 \times 10^{-4}$ m^2/s) and plotted as solid lines in Fig. 3 as a function of f. With increasing f, $|F|$ and Arg(F) respectively approach unity and zero. In this region, δ is much thinner than r_0, and T can be regarded as T_S. On the other hand, $|F|$ and Arg(F) respectively approach zero and 90° when f approaches zero. In this limit, an isothermal acoustic wave is achieved.

We compare the theoretical F and the experimental F_{ex} instead of the true temperature T and the measured temperature T_{ex} to elucidate the response of the thermocouple. The function F_{ex} was determined experimentally from the transfer function given by the ratio of the Fourier transform of T_{ex} to that of T_S, the latter of which was obtained through Eq. (2) and the measured P. Here, the constants $T_m = 298$ K, $P_m = 101$ kPa, and $\gamma = 1.67$ were used. An ideal thermocouple with a perfect response to T is inferred to possess $|F_{ex}| = |F|$ and Arg(F_{ex}) = Arg(F). However, as shown in Fig. 3 $|F_{ex}|$ is smaller than $|F|$ and Arg(F_{ex}) lags behind Arg(F)

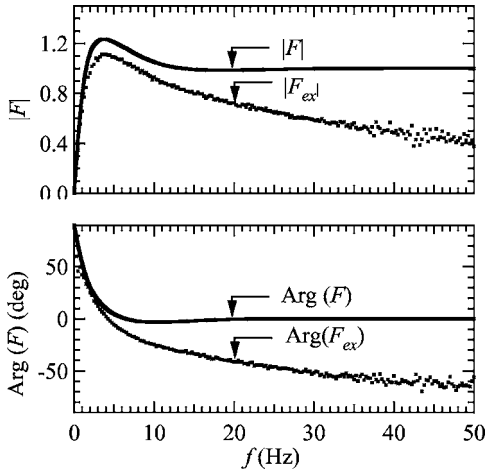


FIG. 3. Frequency dependence of absolute and argument values of F_{ex} and F. Solid lines and dots respectively represent F and F_{ex}.

regardless of f for the present thermocouple. The deviation of F_{ex} from F represents amplitude damping and phase delay of the thermocouple. Because their effects become large with increasing f, the thermocouple response worsens with increasing f.

We determined the response function Z_{ex} of the thermocouple from the relation $Z_{ex}=T_{ex}/T=F_{ex}/F$. The obtained $|Z_{ex}|$ and $\mathrm{Arg}(Z_{ex})$ are plotted in Fig. 4 as a function of f. The response function Z_{ex} obtained using the continuous method, which uses a single-frequency acoustic wave in our previous paper,[11] are also plotted in Fig. 4 as open circles. An excellent agreement was found between them. This fact gives us the validity of the determination of Z_{ex} using the acoustic pulse wave. It is worth noting that Z_{ex}, shown by dots, was determined only by a single pulse, whereas Z_{ex}, shown by open circles, was determined by many experiments with different f in the continuous method. The short measurement time is a clear advantage of the pulse method over the continuous method.

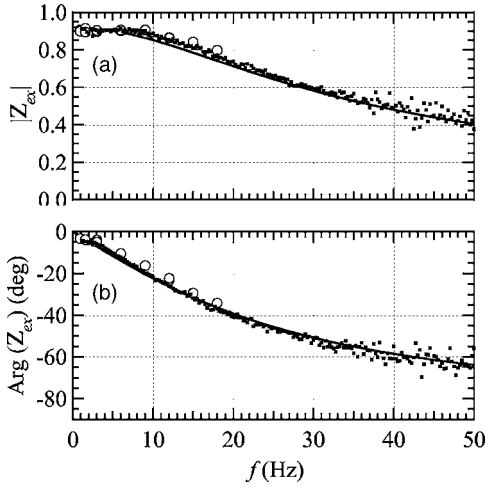


FIG. 4. Frequency dependence of $|Z_{ex}|$ and $\mathrm{Arg}(Z_{ex})$. Dots and open circles respectively represent Z_{ex} with the pulse and continuous (see Ref. 11) methods. Solid lines represent Z' in Eq. (4) with $\varepsilon=0.08$ and $\tau=5$.

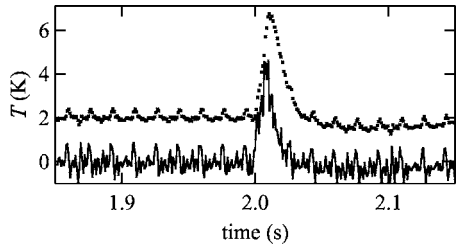


FIG. 5. True temperature T (solid line) deduced from T_{ex} (dotted line), where T_{ex} is plotted with an offset with 2 K.

Figure 4 shows that $|Z_{ex}|$ approac 0.92, whereas $\mathrm{Arg}(Z_{ex})$ approaches zero in the limit of f to 0. From this distinctive feature, we proposed a modified response function Z' in the previous paper,[11] which is written as

$$Z' = \frac{1-\varepsilon}{1+i\omega\tau}, \tag{4}$$

where ε is a parameter representing the degree of heat leakage through the thermocouple wires. In the case of $\varepsilon=0$, showing no heat leak through the wires, Z' is reduced to Z used in a first-order lag system. Solid lines in Figs. 4(a) and 4(b) were drawn by inserting $\varepsilon=0.08$ and $\tau=5$ ms into Eq. (4).[11] The modified response function Z' well fits Z_{ex}, which further validates Z' in the wider frequency region than that obtained using the continuous method.

Since T_{ex}/T can be denoted as Z_{ex} or Z', the amplitude and phase of the true temperature T as a function of ω may well be estimated from the product of $1/Z'$ derived from Eq. (4) and the Fourier transform of T_{ex} for any acoustic wave involving more than one frequency component. Now the time dependence of the true temperature can be calculated by taking the inverse Fourier transform of T thus obtained. This procedure was applied to T_{ex} measured in this experiment, and T thus obtained is plotted in Fig. 5 along with T_{ex}. It is shown that the correction by Z' contributes to narrowing and advancing the pulse of T. The difference between T and T_{ex} represents the importance of the correction. The 60 Hz ac power line noise is enhanced in T, but it can be intentionally cut off by eliminating the peak at 60 Hz from the amplitude spectrum of T_{ex}. As can be understood above, an oscillating temperature T in an acoustic wave can be reliably determined from the product of the Fourier transform of T_{ex} and Z'. By combining temperature measurements with the velocity measurements, we can reliably determine the total energy flow in the acoustic wave.

IV. SUMMARY

We employed a short acoustic pulse wave and measured the time-dependent temperature and the pressure. The measured temperature T_{ex} was compared with the reference temperature deduced from the measured pressure P. From the amplitude and phase spectra, the frequency dependence of the response function of the present thermocouple was obtained in the region below 50 Hz. The response function obtained with the pulse method agreed with that obtained using the conventional method with the use of a continuous acous-

tic wave. We were able to obtain the time-dependent temperature from the measured temperature and the response function.

ACKNOWLEDGMENT

This study was supported by JSPS Research Fellowships for Young Scientists.

[1]A. Tominaga, "Thermodynamic aspect of thermoacoustic theory," Cryogenics **35**, 427 (1995).
[2]G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145 (1988); *Thermoacoustics: Unifying Perspective for Some Engines and Refrigerators*, Acoustical Society of America, Sewickley, PA (2002).
[3]We neglected the contributions of thermal conduction and nonzero mean flow such as acoustic streaming to entropy flow.
[4]T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128 (1998).
[5]Y. Ueda, T. Biwa, U. Mizutani, and T. Yazaki, "Acoustic field in a thermoacoustic Stirling engine having a looped tube and resonator," Appl. Phys. Lett. **81**, 5252 (2002); "Experimental studies of a thermoacoustic Stirling prime mover and its application to a cooler," J. Acoust. Soc. Am. **115**, 1134 (2004).
[6]T. Biwa, Y. Ueda, T. Yazaki, and U. Mizutani, "Thermodynamic mode selection rule observed in thermoacoustic oscillations," Europhys. Lett. **60**, 363 (2002).
[7]T. Biwa, Y. Tashiro, M. Kozuka, T. Yazaki, and U. Mizutani, "Experimental demonstration of thermoacoustic energy conversion in a resonator," Phys. Rev. E **69**, 066304-1 (2004).
[8]G. Huelsz and E. Ramos, "Temperature measurements inside the oscillatory boundary layer produced by acoustic waves," J. Acoust. Soc. Am. **103**, 1532 (1998).
[9]M. Tagawa, T. Shimoji, and Y. Ohta, "A two-thermocouple probe technique for estimating thermocouple time constants in flows with combution: In situ parameter identification of a first-order lag system," Rev. Sci. Instrum. **69**, 3370 (1998).
[10]M. Tagawa, K. Kato, and Y. Ohta, "Response compensation of temperature sensors: Frequency-domain estimation of thermal time constants," Rev. Sci. Instrum. **74**, 3171 (2003).
[11]Y. Tashiro, T. Biwa, and T. Yazaki, "Calibration of a thermocouple for measurement of oscillating temperature," Rev. Sci. Instrum. **76**, 124901 (2005).
[12]P. Merkli and H. Thomann, "Thermoacoustic effects in a resonance tube," J. Fluid Mech. **70**, 161 (1975).

Modeling plasma loudspeakers

Ph. Béquin[a] and K. Castor
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613 F72085, Le Mans cedex 9, France

Ph. Herzog
Laboratoire de Mécanique et d'Acoustique, UPR 7051 F13402, Marseille cedex 20, France

V. Montembault
Laboratoire d'Acoustique de l'Universite du Maine, UMR CNRS 6613 F72085, Le Mans cedex 9, France

(Received 12 May 2006; revised 22 January 2007; accepted 22 January 2007)

This paper deals with the acoustic modeling and measurement of a needle-to-grid plasma loudspeaker using a negative Corona discharge. In the first part, we summarize the model described in previous papers, where the electrode gap is divided into a charged particle production region near the needle and a drift region which occupies most of the inter-electrode gap. In each region, interactions between charged and neutral particles in the ionized gas lead to a perturbation of the surrounding air, and thus generate an acoustic field. In each region, viewed as a separate acoustic source, an acoustical model requiring only a few parameters is proposed. In the second part of the paper, an experimental setup is presented for measuring acoustic pressures and directivities. This setup was developed and used to study the evolution of the parameters with physical properties, such as the geometrical and electrical configuration and the needle material. In the last part of this paper, a study on the electroacoustic efficiency of the plasma loudspeaker is described, and differences with respect to the design parameters are analyzed. Although this work is mainly aimed at understanding transduction phenomena, it may be found useful for the development of an audio loudspeaker. [DOI: 10.1121/1.2697201]

PACS number(s): 43.38.Ar, 43.28.Ra, 43.35.Kp [AJZ] Pages: 1960–1970

I. INTRODUCTION

In plasma loudspeakers, the ionized gas resulting from discharges transforms electrical signals into acoustic perturbations. The development of a loudspeaker of this kind resulted from the need for transducers with no moving parts, giving a wide frequency range. Excellent descriptions of the historical background, the basic principles, and the application of gas discharges to loudspeakers are to be found in Refs. 1 and 2. The first description of the physics involved in audio devices was published in 1982 by Bondar[3,4] and Deraedt[5] has described some devices developed in this context. Many publications have also dealt with the physics of energy transfer occurring in ionized gases (e.g., Refs. 6–10), including the processes addressed here. Last, a strong background about electric discharges in gases can be found in Refs. 11–14.

The plasma loudspeakers described so far in the literature can be divided into two classes with significantly different physical properties, corresponding to the so-called "hot-plasma" and "cold-plasma" loudspeakers, in which the main acoustic source is a heat or a force source, respectively, and their directivity is mainly monopolar or dipolar, respectively.[3,4] In most plasma loudspeakers developed so far, an ionized gas is obtained by applying a high voltage (dc or ac) between electrodes having different curvature radii (e.g., a point and a plane). The complex processes which occur in the electrode gap are called discharges. These discharges show different physical behavior or regimes, depending on the polarity of the small-radius electrode, its radius, the gap length, and the gas.[1,11,15–18]

By choosing appropriate geometric and electric electrode configurations, but also depending on the position relative to the electrodes, the interactions between charged and neutral particles in the ionized gas can produce either a predominant heat transfer (building a Joule heat source, which is almost isotropic) or a predominant momentum transfer resulting in a gas flow, the so-called "electric wind" (this builds a force source having a specific axis). The changes with time in the heating of an air volume lead to corresponding pressure changes. The hot-plasma loudspeakers based on this principle are mainly flame sources,[19–21] the thermophone used for calibrating microphones,[22] the glow discharges,[2,23] the spark discharges,[24,25] and the ionophone.[26–28] By contrast, cold plasma loudspeakers are based on momentum transfer, which was described for the first time in 1972 by Matsuzawa,[29] where a negative voltage is applied to the points of a multipoint-grid electrode system, generating a corona discharge in air. Negative ions are attracted by the positive electrode (grid), and thus tend to drift along the electrical field lines. This gives rise to the electric wind, which can be modulated by combining an alternating audio-frequency component with the dc one.

In the present study, it is therefore first proposed to describe the processes involved in negative discharges, with a view to understanding how these interact, giving rise to the two source mechanisms (heat and force sources) mentioned

[a]Author to whom correspondence should be addressed. Electronic mail: philippe.bequin@univ-lemans.fr

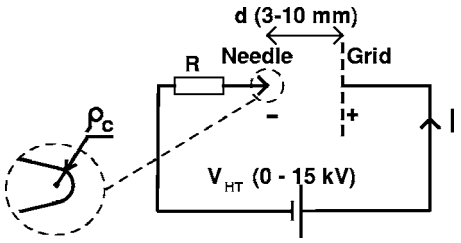


FIG. 1. Diagram of a needle-to-grid discharge. V_{HT}, I, R, d and ρ_c are the voltage supplied, the electric current, the ballast resistance, the gap length and the curvature radius, respectively.

above. Since these mechanisms are highly correlated, we deal with them simultaneously, and point out that in usual situations, they are of the same order of magnitude. Some effects of varying the electrode shapes and discharge parameters are then discussed, and tested experimentally on sample prototypes. This work is part of a research program about transduction principles, with no intent to optimize a commercial device. Our aim is to show the influence of various construction parameters of "plasma" transducers, so this paper deals only with models as simple as possible, and small-scale laboratory experiments. Our results seem, however, to provide useful lines which could be used for the development of actual laboratory or audio systems.

II. DESCRIPTION OF A CORONA LOUDSPEAKER

This section summarizes the results presented in previous papers,[30,31,48,49] focusing on the specific case of the discharges in air under normal atmospheric conditions, which can be obtained by applying a high negative voltage to a needle facing a grid (Fig. 1). The gas between the two electrodes is assumed to be weakly ionized. The energy exchanges between charged particles and neutral particles are complex and can take several forms. The charged particles in a weakly ionised gas acquire energy from the electric field and lose this energy during collisions with the neutral gas particles encountered, whether these collisions are of random or organized form. The macroscopic description of the neutral gas present within the weakly ionized gas involves the assumption that the collisions between neutral particles predominate, and that only the small disturbances generated by charged particles perturb the equilibrium of the system.

The interactions between charged and neutral particles involve both collisional exchanges (momentum and thermal energy), which are introduced into the linear acoustic equations (the equations for the mass, momentum and energy conservation) using two source terms[30,31,48,49] a heat source H and a force source F. Assuming the gas of neutral particles to be an ideal gas, the thermodynamic transformations to be adiabatic, and taking a first-order expansion of the acoustic quantities, a Helmholtz equation can be obtained from the set of linearized acoustic equations (continuity, Euler and Fourier equations). This equation can be separated into a pair of uncoupled equations, each of which is associated with one source. The sound pressures p_h and p_f associated with the heat source and the force source, respectively, can then be obtained by calculating the integral solution of the Helmholtz equation, using the free-space Green function.[48,49]

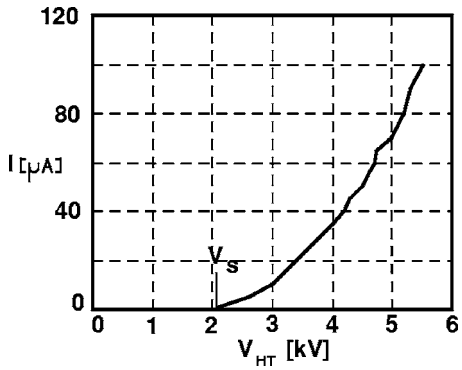


FIG. 2. Typical current–voltage measurement ($I-V_{HT}$) characteristic of needle-to-grid air gap (d=3 mm, ρ_c=20 μm).

Discharges of this kind intrinsically possess both heat and momentum transfer processes, as can be seen from the above description of the basic physics. A previous study by two of the present authors[30] showed that both sources have a significant influence on the pressure radiation. This was checked experimentally on a simple negative polarity point-to-grid electrode system. The directivity curves measured fitted to a supercardioid pattern, giving the sum of a monopole and a dipole with approximately twice the amplitude of the monopole.[31] This finding is in accordance with a note by Tombs and Shirley[32,33] who noted the lack of symmetry in the radiation of their prototypes, although they did not give any explanation for this characteristic.

A. Electrical behavior

When a high voltage is applied to a small-radius electrode, the air in the vicinity of the electrode becomes ionized. The charged particles thus created form a locally concentrated cloud, which drifts along the field lines. The local presence of the cloud distorts the electric field, which therefore varies with time as successive clouds drift from one electrode to the other. This process creates an electric current in the external circuit, consisting of successive "Trichel's pulses" (see Refs. 11, 15, 18, and 34–37), which are characteristic of low-current behavior.

The frequency of the pulses increases with the voltage applied, and successive pulses gradually overlap, leading to a dc current which also increases, and can even predominate when the "pulseless" regime is reached. Higher voltages can even lead to a spark. The geometrical and electrical configurations of the needle-to-grid system can be adapted to provide a wide range of currents, while avoiding the production of sparks. Figure 2 shows the nonlinear relation between the current I and the applied voltage V_{HT}, showing the lower limit V_s of the voltage required to force the current through the discharge. In the corona regime, an empirical relation between I and V_{HT} can be written as follows (see Refs. 15, 16, 18, 34, and 38)

$$I = CV_{HT}(V_{HT} - V_S),$$

where C is a factor which depends on the gap length d and the ionic mobility. If we take a time-averaged distribution of the electric field in the gap, two regions can be distinguished

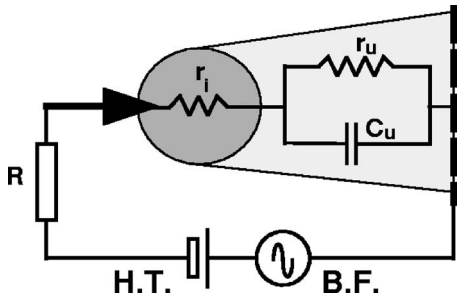


FIG. 3. Diagram of a needle-to-grid discharge. r_i, r_u and C_u denote the resistance of each region and the capacity of the drift region.

(Fig. 3)

- the first one, termed the ionization region, is located at the tip of the point and is characterized by a field value greater than the "critical" value of the electric field E_c, where the electron production resulting from ionization is exactly compensated for by the attachment of the electrons to neutral particles ($E_c \simeq 27$ kV/cm in air);
- the second region, termed the drift region, is located between the ionization region and the plane, and is characterized by a weaker, almost uniform value of the electric field.[39]

On average over the Trichel's pulse period, it can therefore be assumed that charged particles will be created inside the ionization region, where their temperature is high, and that electrons and negative ions will drift along the field lines across the drift region, where the collisions between charged particles and neutral particles result in the air movement known as electric wind or ionic wind (see Refs. 40–46).

B. Electrical model

Kekez, Savic, and Lougheed,[47] have described the point-to-plane discharges in terms of an equivalent circuit composed of three elements (Fig. 3) a resistance r_i, which is taken to be inversely proportional to the electron density, and represents the ionization region; and a resistance r_u, which is taken to be inversely proportional to the negative ion density, shunted by a capacity C_u formed between the plane and ionization region, and represents the drift region. Previous measurements made by the present authors[30,31,48–50] showed that this model accurately simulates the impedance of a single needle-to-grid system up to about 100 kHz, and the sum of the values estimated for r_i and r_u by dynamic impedance measurements showed good agreement with the static resistance calculated from the local slope of the $I-V_{\mathrm{HT}}$ curve. The values of the resistances r_u, r_i and the capacity C_u were found to be around 20 MΩ, 2 MΩ, and 0.2 pF, respectively. The main features of these resistances were the fact that they both decreased with increasing dc current I values, and that they increased with increasing distance d values at constant dc current. If the current flowing through the electrodes is modulated at acoustic frequencies by an external electronic circuit, the two energy transfer mechanisms behave like two coupled acoustic sources, whereas the functioning regime is modulated around an operating point and travels along the $I-V_{\mathrm{HT}}$ curve. The behavior of the transducer can thus be expected to be nonlinear, which cannot be completely avoided by using current drive, although this might be more appropriate than using voltage drive, which exhibits intrinsically quadratic behavior.

C. Heat source model

In the ionization region, the energy with which the charged particles are provided by the electric field is largely transferred to surrounding neutral particles in the form of thermal energy. To develop an expression for the power per unit volume gained by the neutral particles, Bayle *et al.*[7,8,10] have investigated the hydrodynamic equations for each particle type in a spherical volume and obtained the straightforward relation

$$H = \mathcal{K}\mathbf{JE}, \tag{1}$$

where $\mathbf{J}$ denotes the total current density and $\mathbf{E}$ the total electric field in the ionization region, and the parameter $\mathcal{K}$ is the energy transfer rate between charged particles and the surrounding neutral particles.

Taking the ionization region to be a small sphere (with a radius of about 0.1 mm) centered on the point, with an uniform heat source inside this volume, we now focus on the total heat flux leaving the volume. The linearized part of H is therefore the power per unit volume dissipated by the Joule effect, and can be expressed as a function of the electric and geometric models for the loudspeaker. Although the relations between the voltage, the current, and the source term are nonlinear, only small current variations around an average value are considered,[48,49] and the corresponding sound pressure p_h can be expressed as

$$p_h(r,\omega) \approx j\omega \frac{\gamma - 1}{c_0^2} \mathcal{K}[r_i I + (V_i - V_a)] \frac{e^{-jkr}}{4\pi r} i(\omega), \tag{2}$$

where $k = \omega/c_0$ (ω and c_0 are the pulsation and the adiabatic sound speed) and γ, V_a, V_i, I, and i stand for the specific-heat ratio, the electric potentials at the point and at the interface between the two regions, the discharge current and the linearized current variation, respectively.

The amplitude of the pressure depends on the electrical properties of the loudspeaker and shows a +6 dB/octave increase with the frequency. Since the source is assumed to be isotropic and sufficiently small, the pressure is described here as resulting from a pure monopole. In the above expression, only the quantity $\mathcal{K}(V_i - V_a)$ is unknown; it is a part of the total voltage applied to the electrodes, but the exact proportion cannot be determined theoretically without having a suitable model for the electric field, which is beyond the scope of this paper. Its value therefore has to be obtained experimentally by adjusting the predicted acoustic levels.

D. Force source model

In the drift region, only the negative particles drift to the collecting grid along the electric field lines and exchange momentum with the neutral particles during collisions. The electrons are considerably lighter than negative ions, and it is therefore assumed that the effects of the electrons on the

neutral particles will be negligible in comparison to the effects of the negative ions. In addition, the heating of the neutral particle gas is assumed to be negligible. Assuming the mass ratio between negative ion and neutral particle to be approximately equal to unity, the vector force **F** per unit volume applied to the neutral particles can be written as follows:[48,49]

$$\mathbf{F} = N_i q_i \mathbf{E}, \tag{3}$$

where the quantities N_i, q_i, **E** are the density of the negative ions, their charge, and the vector electric field inside the drift region, respectively.

For geometrical reasons, a cylindrical coordinate system was adopted. The drift region is modeled by a cylinder (with radius ρ_d and length d) in which the electric field is assumed to be uniform along and around the axis of the point. In the case where all the dimensions of the drift region are small in comparison with the wavelength λ, the far field sound pressure can be written[49]

$$p_f(r,\omega) \approx \frac{1}{\mu_i(\beta+1)} \frac{i(\omega)}{(1+j\omega r_u C_u)} \frac{(1+jkr)d\cos\theta}{r} \frac{e^{-jkr}}{4\pi r}, \tag{4}$$

where θ, $\beta = I_e/I_i$, μ_i and $i(\omega)$, are the observation angle, the ratio between the electric currents I_e and I_i carried by the electrons and the ions, the mobility of the negative ions ($\approx 1.8 \cdot 10^{-4}$ V m^2/s), and the linearized current variation, respectively.

The pressure expression (4) predicts a null slope at very low frequencies (due to the proximity), and a +6 dB/octave slope then as $(1+jkr)/r \approx jk = j\omega/c$, up to a first-order pole depending on the electric impedance of the drift region, as shown by the denominator term $(1+j\omega r_u C_u)$. The directivity pattern is dipolar in the case of the above assumption, but at higher frequencies (above 10 kHz with the usual configurations), interferences occur throughout the gap, leading to a higher-order directivity pattern, and a sharp cutoff when $kd \approx \pi/2$. As in the heat source model, there is an unknown quantity, the ratio $\beta = I_e/I_i$, for which only the order of magnitude has been mentioned in previous studies about electric discharges. This factor results from the hypothesis that only heavy ions interact significantly with neutral particules, so that only the corresponding part of the current leads to the source term. The electron current is therefore not considered in the transduction equations. The value of β must be deduced from experimental results, keeping in mind that it may then include both the physical parameter value, and some corrections to the above hypothesis.

E. Total pressure

The total pressure p_t is the sum of the pressures p_h [Eq. (2)] which is written assuming a pole to be located at the tip of the point, and taking the pressure p_f [Eq. (4)] with a dipole centered in the gap. These incompatible choices are analytically convenient because the symmetry of the related problem is maintained with each source, which makes it easier to analyze their respective directivities. In practice, during directivity measurements, the origin of the needle-to-grid system is almost centered on the gap. The distance between the heat source and the microphone therefore varies, depending on the observation angle. However, this angle dependence of the observation distance can reasonably be assumed to be negligible in the far field, at least at low enough frequencies for the corresponding phase lag to be of no importance.

Keeping in mind the above remarks, the total pressure p_t can be modeled as the sum of the analytical expression for p_h and p_f, giving a frequency response having a +6 dB/octave slope in most of the audio spectrum, and a general cardioid directivity pattern resulting from the sum of a monopole pressure p_h and a dipole pressure p_f. At lower frequencies (or when moving closer to the source), the near-field term of the dipole predominates, and at very high frequencies, the monopole behavior takes over as the gap resistance is shunted by its capacitance.

The model summarized above takes only into account the minimum complexity necessary for comparison with experimental results. More detailed expressions have been provided elsewhere,[30,31,48–50] and could be used in place of Eqs. (2) and (4). For the small gaps considered in this paper, we have assumed the linear sum of two discrete sources (monopole and dipole), because:

- The extent of the heat source is somewhat larger than the ionization zone, especially if we take into account thermal diffusion in the air; its dimension can be evaluated around 1 mm. The length of the force source is almost equal to the electrode distance, i.e., several mm. Actual responses and directivities may therefore differ from the discrete model, but only at frequencies above about 20 kHz.
- The two expressions for the source terms are not reciprocal, so we cannot take into account the reaction of the fluid (analogous to the radiation impedance in the case of a vibrating surface). We have investigated this topic recently, and found no evidence that it could be significant for the cases considered here.
- Although both sources are very close (and may probably overlap), linear acoustics seems still valid at the very low levels considered in this paper. Measurements at much higher modulation indices led to velocities of a few m/s (on the discharge axis), and marginal harmonic distortion.

Experiments show, however, that the discharge produces a significant air flow, which may interact with the source terms mentioned above. We neglect here the possible energy exchanges related to such interaction, assuming that the air flow is mainly connected with the dc current. Removing this hypothesis would require a much more complex model, which is beyond the scope of our work.

Since both source terms are therefore linearly summed in a volume integral, this simple analytical model can easily be extended to the case of multiple points. Assuming that they all have similar electrical behavior, a simple convolution of the Green function can be performed over the spatial distribution of the points. With small enough transducers, under far-field conditions, this sum can lead to an equivalent transducer combining the strength of all the points, if the path differences between points can be neglected. Although

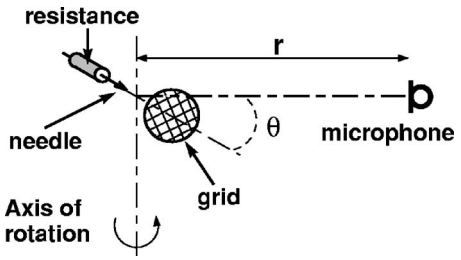


FIG. 4. Electrodes in the experimental needle-to-grid system.

this simple method can be used to combine several points from the acoustical standpoint, one should be aware that even quite slight modifications of the electrical field lines are likely to change the discharge regime completely, and thus the source terms. An initial experimental study on this phenomenon[49] showed that the values of the $\mathcal{K}(V_i - V_a)$ and $\beta = I_e / I_i$ factors differ considerably when multiple points are placed close together, from those obtained in the case of a single point.

III. EXPERIMENTAL TESTS ON DIFFERENT CONFIGURATIONS

In this section, the effects of the most significant parameters of a plasma loudspeaker are analyzed, based on experimental studies of a single point unit, which requires a dedicated setup because of the low acoustic pressures involved.

A. Experimental setup

Since the experimental setup used here was mostly described in detail in a previous paper,[31] only specific parts are recalled below (when necessary to interpret the response data obtained using the experimental setup). Experiments were performed using very simple electrode configurations, in order to keep as closely as possible to the modeling assumptions adopted in the previous part. The electrode system used consisted of a stainless steel needle connected in series with a 1 MΩ resistance (Fig. 4). The needle axis was placed perpendicularly to a circular flat steel wire gauze (mesh 0.05 $\times 0.05$ mm^2; $\varnothing_{\text{wire}} = 0.03$ mm), which was assumed to be acoustically transparent. Most of the measurements were carried out using needles having different tip radii of curvature and materials. The needle (with series resistor) and the gauze were held in a thin plastic frame designed to have little effect on the acoustic field. With this frame, the needle-to-grid gap can be varied from 3 to 10 mm.

The supply voltage values V_{HT} were in the 3–10 kV range, and the resulting dc current value I ranged from 30 to 100 μA. The voltage modulation was about 10 V_{rms} which leads to a modulation degree of less than 0.5%. This low value limits the acoustic field radiated by the system, while avoiding any loudspeaker nonlinearity. All the measurements were carried out in the "pulseless" regime so that the pulse frequency was far greater than the audio frequencies, and the pulsed part of the current was small (the ripple was lower than 10%).

Two microphones were used to measure the acoustic pressure generated by the discharge : a Brüel and Kjaer 1/4 in. condenser microphone for measuring frequencies up to 10 kHz, and an 1/8 in. condenser microphone for higher frequencies. The 1/8 in. microphone could not be used over the entire frequency range because of its background noise level, whereas the 1/4 in. microphone induced too much diffraction at higher frequencies. In both cases, the microphone axes were at right angles with respect to the transducer axis; the effect of diffraction on the microphone bodies is then supposed to be lower than 1 dB in the whole frequency range, and is certainly negligible compared to other diffraction sources, such as the supporting frame. Given the low pressure levels measured (0–60 dB_{SPL}), the electrical signals delivered by the microphone cartridge were amplified and filtered by a microphone amplifier and a 1/3 octave filter, respectively. In addition, an electromagnetic shield assumed to be transparent to the acoustic waves was used to protect the microphones against the non-negligible electromagnetic radiation emitted by the transducer.

With the equipment available, the frequency range scanned was limited to [100 Hz–100 kHz]. In the frequency range [5–50 kHz], the pressure signal was measured at least 20 dB above the mainly coherent background noise due to parasitic electromagnetic coupling. The voltages v_i and v_μ, which were proportional to the ac current flowing through the air gap and the acoustic pressure, respectively, were measured using a lock-in amplifier (Stanford Research SR850) via a self-made multiplexer, so that the transfer function (v_μ / v_i) could be calculated. The demodulation parameters were carefully optimized to obtain accurate signal measurements within a relatively short time, thus avoiding the problems due to changes in the experimental conditions with time.

B. Frequency response and directivity

Figure 5 gives an example of the directivity curves measured and the on-axis acoustic pressure. These directivity curves, like those obtained under other experimental conditions, have a very similar pattern of directivity to that of a supercardioid, resulting from the sum of a monopole and a dipole.

The frequency response and the directivity curve give a global acoustic pressure response of the plasma loudspeaker. The present approach involves the use of the models described above [Eqs. (2) and (4)], in order to separate, the responses associated with each acoustic source. This approach starts with the acoustic pressure measured, which can be modeled as follows[31,49,50]

$$p(r,\omega) = p_m(r,\omega) + p_d(r,\omega) = (A + B\cos\theta)\frac{e^{-jkr}}{r} i(\omega), \quad (5)$$

where p_m and P_d have a directivity of a monopolar and dipolar kind, respectively.

By plotting the amplitude of $p(r,\omega)$ vs $\cos(\theta)$, and using a linear regression, factors A and B can be deduced from the directivity curve at each frequency. The validity of this method was checked by plotting the difference between the modeled and the measured pressures, which usually varies almost randomly with respect to the angle θ. The adjusted

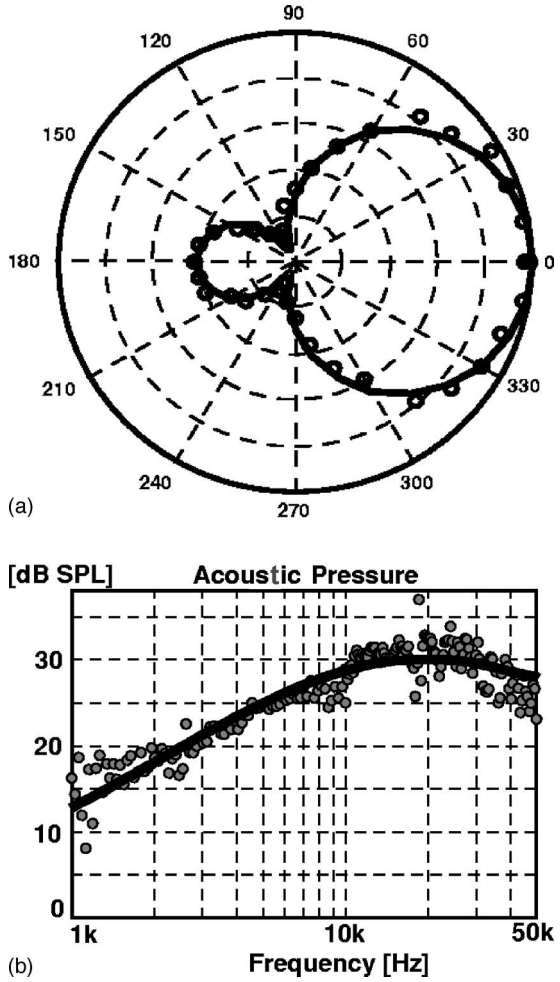


FIG. 5. (a) Directivity curve (f=5 kHz) the acoustic pressure is normalized with respect to the on-axis (θ=0) value, (b) On-axis acoustic pressure versus frequency, (∘∘∘) experimental results; (——) theoretical predictions [Eqs. (2) and (4)]. Configuration of the needle-to-grid system d=6 mm, I =60 μA, i_{mean}=0.3 μA, r=10 cm.

pressures p_m and p_d were then associated with the theoretical pressures p_h and p_f [Eqs. (2) and (4)], respectively, so that the two adjustable parameters $\mathcal{K}(V_i-V_a)$ and $\beta=I_e/I_i$ could be estimated from the pressure levels generated by the two source terms. Figure 5 shows the measured acoustic pressure of a needle-to-grid system, and that deduced from the models, with parameters $\mathcal{K}(V_i-V_a)$ and $\beta=I_e/I_i$ adjusted to obtain the best fit with the experimental data.

In the frequency range considered, the frequency response measured takes the form of a straight line with a slope of +6 dB/octave from about 3 to 15 kHz. At lower frequencies, the pressure is too low to be able to obtain a good estimate around the whole needle-to-grid setup, but the behavior of the transducer is likely to correspond to a simple asymptotic behavior. Above 30 kHz, the simple models used in this paper are no more valid, and fail to predict accurately the cutoff of both source terms. Moreover, diffraction occurs with various components of the measuring system, especially the frame and connecting wires (this is revealed by complex directivity patterns).

The agreement between the models and the measurements can therefore be said satisfactory up to 20 kHz. This measurement setup therefore provides a useful means of estimating the acoustic pressures associated with the heat and the force sources in most of the audio frequency range. It can therefore be used to determine how the response of each acoustic source is affected when the electric and geometric configurations of the plasma loudspeaker are changed.

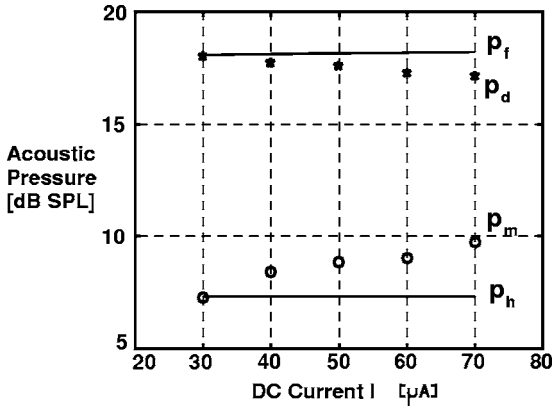


FIG. 6. Acoustic pressure level measured on the axis of the needle-to-grid systems at r=10 cm (d=3 mm, ρ_c=20 μm, i_{mean}=0.38 μA, and f=5 kHz). (⋆, ∘) Experimental results associated with p_d and p_m; (—) theoretical predictions associated with p_h and p_f where the parameters adjusted are $\beta \approx 1.9$ and $\mathcal{K}(V_i-V_a) \approx 1.4$ kV.

C. Influence of the electric and geometric configurations

The discharge current I, the electrode separation d, and the tip radius ρ_c are parameters which can affect the spatial distribution of the electric field inside a discharge, and these parameters in turn can influence the discharge regime, the acoustic behavior, and finally the electroacoustic efficiency of the plasma loudspeaker. Figures 6 and 7 show the acoustic pressure versus the dc current I and the gap length d, respectively, estimated for each source (heat and force) at r =10 cm from the gap center of a needle-to-grid system (with d=3 mm, ρ_c=20 μm, f=5 KHz). In this geometric and electric configuration, the sound level due to the force source has

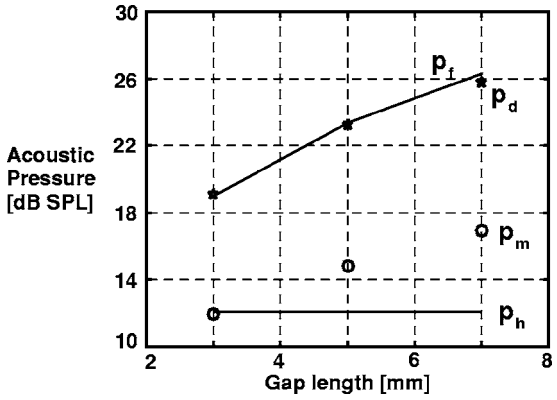


FIG. 7. Acoustic pressure level measured on the axis of the needle-to-grid systems at r=10 cm (ρ_c=20 μm, I=70 μA, i_{mean}=0.53 μA, and f =5 kHz). (⋆, ∘) Experimental results associated with p_d and p_m; (—) theoretical predictions associated with p_f and p_h, where the parameters adjusted are $\beta \approx 2.6$ and $\mathcal{K}(V_i-V_a) \approx 1.8$ kV.

been reported to be about 6 dB greater than the level due to the heat source in the (2–20 kHz) frequency range.[31,49,50]

In Figs. 6 and 7, the two parameters β and $\mathcal{K}(V_i-V_a)$ of the theoretical results p_h, p_f [Eqs. (2) and (4)] are adjusted to fit the first experimental result obtained with p_m, p_d [Eq. (5)], and kept constant when varying d or I. This leads to increasingly large discrepancies between the predictions p_f and experimental results p_d as dc current I increases (Fig. 6). Likewise, the discrepancies between the predictions p_h, p_f and experimental results p_m, p_d become increasingly large as the gap length d increases (Fig. 7). These findings clearly indicate that the two unknown parameters depend on d and I, although almost nothing has been published in the literature on this topic.

Many experimental values for the adjustable parameters have been obtained by Castor,[50] and only the main tendencies observed during these experiments will be summarized below. The parameter $\mathcal{K}(V_i-V_a)$ was found to increase with the gap length d or the dc current I. Furthermore, with a needle made of brass or a stainless material, this parameter decreases with increasing radii of curvature ρ_c. The opposite tendency is, however, observed with a needle made of copper or a steel. β is deduced and found to have a value of around 2, which increases with dc current I and is almost independent of the distance d. With a needle made of brass or a stainless material, the parameter β was found to increase with increasing radii of curvature ρ_c. The opposite occurs when the needle is made of copper or a steel material. This influence of the β parameter is rather surprising. It might be related to the intrinsic characteristics of the different materials (e.g., electron emissivity), but it can also reveal a hidden factor, as a different surface smoothness, which would modify the discharge behavior. This has to be further investigated. Last, it is worth noting that the smaller the gap length d is, the greater the dependence will be between the two parameters $[\beta, \mathcal{K}(V_i-V_a)]$ and the radius of curvature ρ_c.

IV. POWER EFFICIENCIES

In order to design and optimize a plasma loudspeaker, it is necessary to modelize the evolution of the electroacoustic efficiency of each source with the geometric and electric configurations. To our knowledge, no studies have yet been carried out on these lines, apart from some earlier studies in which the electric modulation was not taken into account owing to the complexity of the energy exchanges occurring between charged and neutral particles, the fraction of the electric energy transformed into thermal form and kinetic form (wind) was roughly estimated to be around 20%[7,8,10] and 5%,[5,51] respectively. In this section, the methods used to measure the dynamic input (electric) and output (acoustic) powers involved in the plasma loudspeaker will therefore be described.

A. Electric power

The dynamic part of the electric power associated with the needle-to-grid system can be expressed as[50,52]

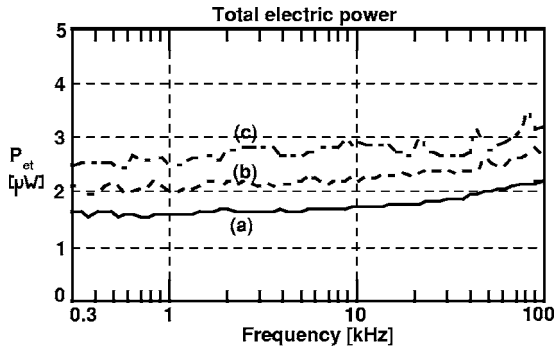


FIG. 8. Electrical power measured on the plasma loudspeaker with the geometrical configuration $d=3$ mm, and $\rho_c=20$ μm, and with three electric configurations (a) $I=30$ μA, $V=4$ kV; (b) $I=50$ μA, $V=4.6$ kV; (c) $I=70$ μA, $V=5$ kV.

$$P_e = \frac{1}{2}\Re\mathfrak{e}[v(\omega)\cdot i^*(\omega)] = \frac{1}{2}\Re\mathfrak{e}[Z(\omega)]\cdot|i(\omega)|^2, \tag{6}$$

where v, i, and Z are the ac voltages between the electrodes, the ac current flowing through the electrodes, and the electrical impedance. The asterisk denotes complex conjugation.

This electrical power can be divided in two parts. The one part is associated with the ionization region

$$P_{ei} = r_i\cdot|i_{\mathrm{rms}}(\omega)|^2 \tag{7}$$

and the other part is associated with the drift region

$$P_{\mathrm{ed}} = \frac{r_u}{1+(\omega r_u C_u)^2}\cdot|i_{\mathrm{rms}}(\omega)|^2. \tag{8}$$

From these expressions, it is deduced that the total electrical power is constant at low frequencies and equal to

$$P_{\mathrm{et}} \approx (r_i + r_u)\cdot|i_{\mathrm{rms}}(\omega)|^2. \tag{9}$$

The values of r_u, r_i and C_u measured by the present authors[31,49] were found to be around 20 MΩ, 2 MΩ and 0.2 pF, respectively, for needle-to-grid gap varying from 3 to 10 mm. This pattern of evolution is observed in Fig. 8, where the electrical power measured is around a few μW, and remains relatively constant with frequencies of up to 20 kHz. Its value then increases slightly at higher frequencies. Similar frequency-dependent behavior of the electrical power is observed whatever the geometrical and the electrical configurations of the plasma loudspeakers.[49]

B. Acoustic power

The intensity vector **I**, which describes the energy flow in an acoustic field, is defined as the time average of the product of the acoustic pressure p and the particle velocity **v**. After being integrated over an area S, the intensity gives the acoustic power flow through this area. In the present case, the plasma loudspeaker is taken to be placed in the center of a sphere with radius r, the surface area S of which thus completely surrounds the acoustic sources. Assuming far-field conditions, the pressure and the velocity are related via the characteristic impedance of air, and the intensity is a function of the pressure, which can be integrated over the sphere to obtain the radiated acoustic power.

Like the electrical power, the acoustic power is divided into two parts. On the one hand, we assume that the acoustic power delivered by the heat source corresponds to the monopolar component of the pressure response, and on the other hand, we assume that the dipolar component must be associated with the force source term. The acoustic power expressions associated with the ionization region and the drift region are then, respectively,[50,52]

$$P_{\mathrm{ah}} \approx \frac{1}{4\pi}\frac{1}{\rho_0 c_0}\left(\frac{\omega}{c_0}\right)^2\left(\frac{\gamma-1}{c_0}\right)^2.$$

$$[\mathcal{K}(V_i - V_a)]^2 |i_{\mathrm{rms}}(\omega)|^2 \tag{10}$$

and

$$P_{af} \approx \frac{1}{12\pi}\frac{1}{\rho_0 c_0}\left(\frac{\omega}{c_0}\right)^2\left(\frac{d}{\mu_i(\beta+1)}\right)^2.$$

$$\frac{1}{1+(\omega r_u C_u)^2}|i_{\mathrm{rms}}(\omega)|^2 \tag{11}$$

with $kr \gg 1$.

These two relations show the dependence on the frequency, and the geometric and electric configurations of the plasma loudspeaker. An experimental study by Castor[50] on needle-to-grid plasma loudspeakers has shown that the measured acoustic power associated with each source is very low; the very low values calculated, which were equal to only a few pW, were partly due to the low level of electric modulation degree ($v_{\mathrm{modulation}}/V_{\mathrm{HT}} \ll 1$) allowed by the experimental setup.

C. Efficiencies

The electroacoustic efficiency η_h (respectively η_f) associated with the heat source (respectively the force source) is the ratio between the time-averaged acoustic power P_{ah} (respectively P_{af}) and the time-averaged electric power P_{ei} (respectively P_{ed})

$$\eta_h = \frac{P_{\mathrm{ah}}}{P_{\mathrm{ei}}} = \frac{1}{4\pi}\frac{1}{\rho_0 c_0}\left(\frac{\omega}{c_0}\right)^2\left(\frac{\gamma-1}{c_0}\right)^2\frac{[\mathcal{K}(V_i - V_a)]^2}{r_i}. \tag{12}$$

$$\eta_f = \frac{P_{\mathrm{af}}}{P_{\mathrm{ed}}} \approx \frac{1}{12\pi}\frac{1}{\rho_0 c_0}\left(\frac{\omega}{c_0}\right)^2\left[\frac{d}{\mu_i(\beta+1)}\right]^2\frac{1}{r_u}. \tag{13}$$

Both electroacoustic efficiencies increase with the square of the frequency (+12 dB/octave). They also depend on many parameters associated with the weakly ionised air (γ, ρ, c_0 and μ_i), the electric configuration [$r_i, r_u, \mathcal{K}(V_i - V_a)$ and β] and the gap length d. However, since all the electrical parameters are themselves dependent on the geometric configuration of the electrodes, the actual number of parameters is in fact much lower than in the above expressions. For a given needle (i.e., a given material and radius of curvature ρ_c), the only effective "tunable" parameters are therefore I and d.

D. Measurements and trends

Experimental efficiency estimates are given in Fig. 9, showing the practical importance of the main parameters defining the needle and the configuration. The two efficiencies associated with each source term were first determined separately, identifying the monopole and dipole radiated fields, and the efficiency of the transducer was then studied as a whole. Comparisons between these curves can be discussed from the point of view of the influence of parameters, and an optimum configuration can be proposed.

Since the electric current crossing the ionization and drift regions is the same, the electric power associated with the electric modulation is mainly transferred to the drift region, where the electric impedance is higher [Eq. (8)]. Although the electric power transferred to the heat source is relatively low [Eq. (7)], its contribution to the total acoustic power [Eq. (10)] is generally found to be 25% higher than the force source contribution [Eq. (11)].

The efficiency associated with the heat source η_h is therefore generally about 20 times higher than that associated with the source force η_f (Figs. 9(a) and 9(b)). However, even with a much higher efficient heat source, the global energy behavior of the transducer is observed to be governed mainly by the drift region, which absorbs most of the modulation input power [Fig. 9(c)].

Whatever the value of the electric current I in the 20–100 μA range, the total electroacoustic efficiency (like the source force efficiency) increases with increasing gap lengths [d=6 and 8 mm; Fig. 9(c)]. This behavior depends on the needle material with needle made of titanium, steel or brass, a roughly linear relation is observed between the efficiency and the gap length, whereas with stainless needles this linear relation disappears.[50]

Figure 9(d) gives the total electroacoustic efficiency associated with the geometric configuration of the electrodes [d=6 mm, ρ_c=100 μm] allowing a comparison of several needle materials. Whatever the needle material the total electroacoustic efficiency increases with increasing electric current I; with this electrode configuration copper and steel needles gives the higher efficiency and higher currents can be applied to stainless needle.

With a needle made of brass or a stainless material, the total electroacoustic efficiency decreases with increasing radius of curvature (ρ_c=50 and 100 μm). The opposite tendency is observed for a needle made of copper or steel.[50]

Based on the above considerations, in order to obtain a needle-to-grid plasma loudspeaker with an optimum efficiency, a system involving the use of a steel or stainless needle with a small radius curvature placed 1 cm from the grid and fed with a dc current in the 70 to 90 μA range is recommended. Note that with a greater gap length or dc current, electric instabilities occur in the discharges, leading to an undesirable random acoustic emission.

V. CONCLUSION

The electrode gap in negative needle-to-grid corona discharge loudspeakers is divided into an ionization region located near the point of the needle and a drift region. In each

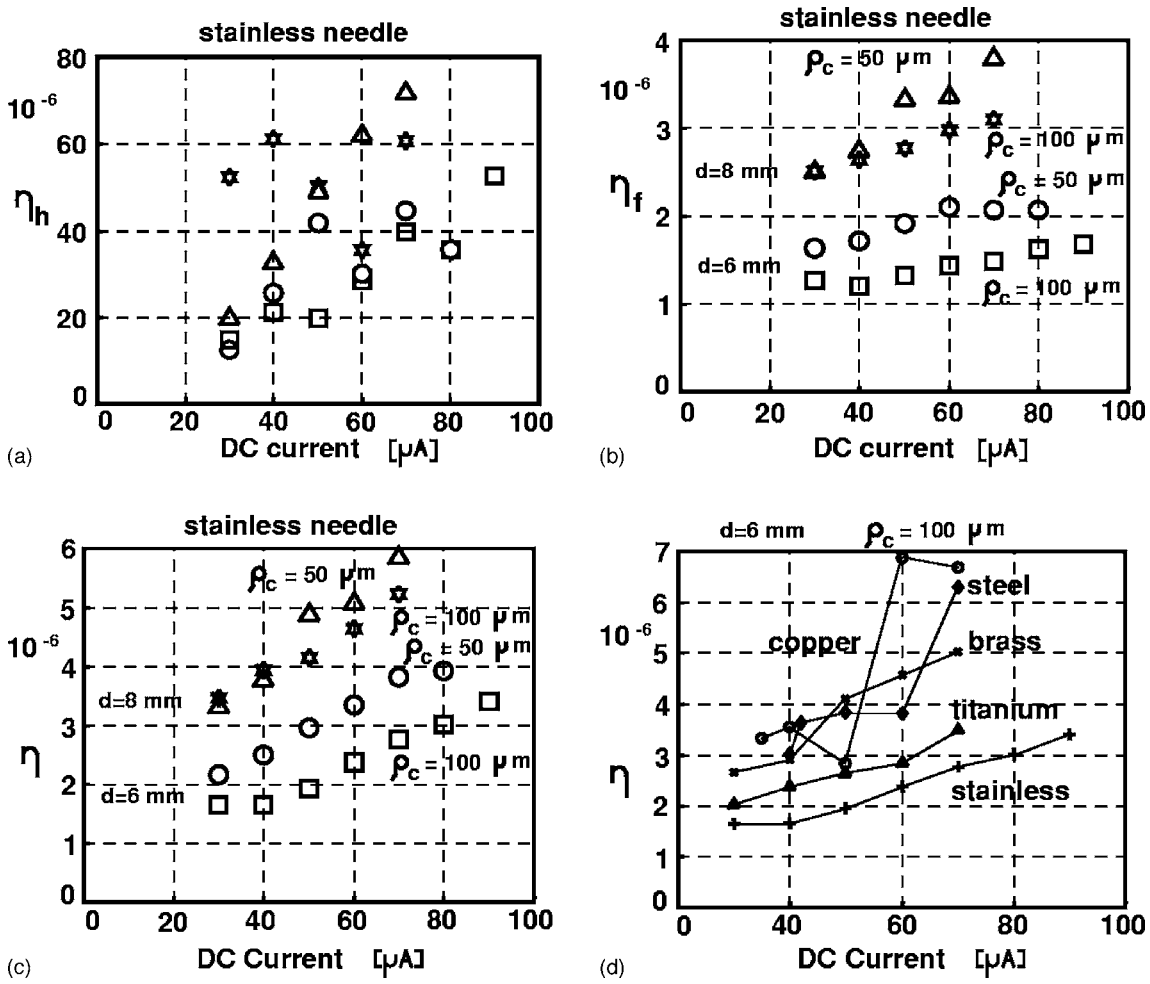


FIG. 9. Electroacoustic efficiency associated with (a) the heat source η_h, (b) the force source η_f, (c) the total electroacoustic efficiency η (Refs. 50 and 52): configuration of the needle-to-grid system with stainless needle (□, ○) d=6 mm, (⋆, △) d=8 mm with (△, ○)ρ_c=50 μm, (□, ⋆)ρ_c=100 μm. (d) The total electroacoustic efficiency η; configuration of the needle-to-grid system d=6 mm, ρ_c=100 μm.

region, interactions between charged and neutral particles present in the ionized gas lead to a perturbation of the surrounding air and thus generate an acoustic field. In each region, viewed as an independent acoustic source, a specific acoustical model is developed with monopolar and dipolar characteristics associated with the ionization and drift regions, respectively. The main limitation of these models is that in both cases it is necessary to adjust a parameter associated with the electrical configuration. A study was carried out on the electrical and acoustic power, and the electroacoustic efficiency of the loudspeaker. The results of this study were as follows:

1. The plasma loudspeaker has directivity curves that resemble a supercardioid pattern, and are almost constant over the audio frequency range.
2. The amplitude of the acoustic pressure associated with the source force (in the drift region), which shows dipolar behavior, was found to be approximately twice as high as the acoustic pressure associated with the heat source (in the ionization region), which shows monopolar behavior.
3. The main physical properties of the plasma loudspeaker which determine the acoustic pressure amplitudes associated with each source are the needle material and the radius of curvature, and the distance between the two electrodes. The dc current feeding the discharge can be adjusted within a small range, in a given electrode configuration.
4. Although both sources have similar strengths in terms of the radiated pressure, the heat source has a much greater efficiency, but the total efficiency is still governed by the force source term.

These conclusions are still preliminary ones, as the range of possible materials, tip geometries, and electrode configurations is obviously much wider than that covered here. Our results seem, however, to be in agreement with most of those published up to now, as far as negative dc Corona discharges are concerned.

Based on the results obtained here on plasma loudspeakers, the acoustic pressure generated by this system is directly proportional to the modulation of the electric current flowing between the electrodes. When an oscillating electric field is obtained by voltage modulation of the needle-to-grid system, current harmonic distortion and consequently acoustic pressure distortion are, however, to be expected, because of the nonlinear current voltage $I-V_{\mathrm{HT}}$ characteristic (see Fig. 2).

In the previous section, this nonlinear behavior is minimized by using a low voltage modulation ($v_{\text{modulation}}/V_{\text{HT}} < 0.5\%$), but the acoustic power generated in this case by the needle-to-grid system is then very limited.

The following topics now require further study, in order to assess the actual performances of a future unit:

1. In order to obtain higher acoustic pressure levels and avoid the effects of the nonlinear current–voltage ($I - V_{\text{HT}}$) characteristic on the pressure frequency response, a current modulation will have to be used. This will make a higher level of modulation possible.
2. The model will have to be extended, and the nonlinearities in the acoustic sources (heat and force) will have to be analyzed, and compared with the results of experimental studies on the distortions occurring in plasma loudspeakers.
3. Theoretic and experimental studies on plasma loudspeakers with other geometric configurations (e,g., a wire-to-plane system) and multiple electrode systems could also be envisaged, provided that a suitable model for, the electric behavior is developed.

ACKNOWLEDGMENTS

The authors would like to thank Guy Tournois and Alain Brunet for their assistance with the experiments and their technical advice.

[1]F. Bastien, "Acoustic and gas discharges applications to loudspeakers," J. Phys. D **20**, 1547–1557 (1987).
[2]M. S. Mazzola and G. M. Molen, "Modeling of a dc glow plasma loudspeaker," J. Acoust. Soc. Am. **81**(6), 1972–1978 (1987).
[3]H. Bondar, "Haut-parleur vers une ère nouvelle" ("Loudspeaker toward a new era"), Nouv. Rev. son **58**, 71–79 (1982).
[4]H. Bondar, "Un haut-parleur à plasma froid" ("A cold plasma loudspeaker"), Nouv. Rev. son **59**, 73–80 (1982).
[5]A. Deraedt, "Electroacoustic transducer using corona effect," in *Proceeding of the 90th Audio Engineering Society Convention*, volume preprint 3037 (F-2), 1–19 (A.E.S., Paris, 1991).
[6]M. Fitaire and T. Mantei, "Some experimental results on acoustic wave propagation in plasma," Phys. Fluids **15**, 464–469 (1972).
[7]P. Bayle, M. Bayle, and G. Forn, "Neutral heating in glow to spark transition in air and nitrogen," J. Phys. D **18**, 2395–2415 (1985).
[8]P. Bayle, M. Bayle, and G. Forn, "Blast wave propagation in glow to spark transition in air," J. Phys. D **18**, 2417–2432 (1985).
[9]Ph. M. Morse and K. U. Ingard, *Plasma Acoustics* (Princeton University Press, Princeton, 1986) Chap. 12.
[10]O. Eichwald, M. Jugroot, P. Bayle, and M. Yousfi, "Modeling neutral dynamics in pulsed helium short-gap spark discharges," J. Appl. Phys. **80**(2), 694–709 (1996).
[11]M. Goldman and A. Goldman, in *Gaseous Electronics*, edited by M. Hirsch and H. Oskam (Academic, New York, 1978), Vol. **I**, Chap. 4.
[12]E. E. Kunhardt and L. H. Luessen, *Electrical Breakdown and Discharges in Gases* (Plenum, New York, 1982).
[13]E. Nasser, *Fundamentals of Gaseous Ionisation and Plasmas Electronics* (Wiley, New York, 1971).
[14]Y. P. Raiser, *Gas Discharge Physics* (Springer-Verlag, Berlin, 1991).
[15]W. L. Lama and C. F. Gallo, "Systematic study of the electrical characteristics of the Trichel current pulses from negative needle-to-plane coronas," J. Appl. Phys. **45**(1), 103–113 (1974).
[16]R. S. Sigmond, "Simple approximate treatment of unipolar space-charge-dominated coronas, The Warburg law and the saturation current," J. Appl. Phys. **53**(2), 891–898 (1982).
[17]R. S. Sigmond, "The residual streamer channel return strokes and secondary streamers," J. Appl. Phys. **56**(5), 1355–1370 (1984).
[18]G. F. L. Ferreira, O. N. Oliveira, and J. A. Giacometti, "Point-to-plane corona current-voltage characteristics for positive and negative polarity with evidence of an electronic component," J. Appl. Phys. **59**(6), 3045–3049 (1986).
[19]W. R. Babcock, K. L. Baker, and A. G. Cattaneo, "Music flames," Nature (London) **216**, 676–678 (1967).
[20]M. Fitaire and D. Sinitean, "Excitation d'ondes acoustiques par une flamme" ("Acoustic waves generation by flame") Czech. J. Phys., Sect. B **B**(22), 394–397 (1972).
[21]M. S. Sodha, V. K. Tripathe, and J. K. Sharma, "Flame loudspeakers," Acustica **40**, 68–69 (1978).
[22]P. Riety, "Retour sur la théorie du thermophone à feuilles d'or" ("Look back on thermo-phone theory") Cahiers d'Acoustique **70**, 169–201 (1955).
[23]F. J. Fransson and E. V. Jansson, "The STI-Ionophone transducer properties and construction," J. Acoust. Soc. Am. **58**(4), 910–915 (1975).
[24]L. D. Lafleur, J. J. Matesse, and R. L. Spross, "Acoustic refraction by a spark discharge in air," J. Acoust. Soc. Am. **81**(1), 606–610 (1987).
[25]M. Akram and E. Lundgren, "The evolution of spark discharges in gases Macroscopic models," J. Phys. D **29**, 2129–2136 (1996).
[26]S. Klein, "L'ionophone" ("The ionophone") Onde Electr. **26**, 314–320 (1946).
[27]S. Klein, "Cellule thermionique de grande puissance a atmosphère gazeuse et ions positifs" ("High power thermo-ionic device with gaseous atmosphere and positive ions") Onde Electr. **26**, 367–373 (1946).
[28]S. Klein, French patent specification No. 79 09450 (1979).
[29]K. Matsuzawa, "Sound sources with corona discharges," J. Acoust. Soc. Am. **54**(2), 494–498 (1972).
[30]Ph. Béquin and Ph. Herzog, "Model of acoustic sources related to negative point-to-plane discharges in ambiant air," Acta Acust. **83**, 359–366 (1997).
[31]Ph. Béquin, V. Montembault, and Ph. Herzog, "Modeling of negative point-to-plane corona loudspeaker," Eur. Phys. J.: Appl. Phys. **15**, 57–67 (2001).
[32]D. M. Tombs, Nature (London) **176**, 923 (1955).
[33]G. Shirley, "The corona wind loudspeaker," J. Audio Eng. Soc. **5**, 29–37 (1957).
[34]M. Boutlendj and N. L. Allen, "Current-density distribution on a plane cathode in dc glow and streamer corona regime in air," IEEE Trans. Electr. Insul. **28**(1), 86–92 (1993).
[35]Y. S. Akishev, I. V. Kochetov, A. P. Napartovich, and N. I. Trushkin, "The generation-zone structure in negative corona discharges," Plasma Phys. Rep. **21**(2), 179–183 (1995).
[36]A. P. Napartovich, Y. S. Akishev, A. A. Deryugin, I. V. Kochetov, M. V. Pan'kin, and N. I. Trushkin, "A numerical simulation of Trichel-pulse formation in a negative corona," J. Phys. D **30**, 2726–2736 (1997).
[37]Y. S. Akishev, M. E. Grushin, I. V. Kochetov, A. P. Napartovich, and N. I. Trushkin, "Establishment of regular Trichel pulses in a negative corona in air," Plasma Phys. Rep. **25**(8), 922–927 (1999).
[38]A. E. Seaver, "An engineering equation for Corona devices," IEEE Industry Applications Magazine 30–35 (1995).
[39]R. Morrow, "Theory of negative corona in oxygen," Phys. Rev. A **32**(1), 1799–1808 (1986).
[40]M. Robinson, "Movement of air in the electric wind of the corona discharge," Trans. Am. Inst. Electr. Eng. **80**, 143–152 (1961).
[41]L. C. Thanh, "Similitude between ionic wind discharge pattern and corona current," Electron. Lett. **15**(2), 57–58 (1979).
[42]R. S. Sigmond, "Mass transfer in corona discharges," Nonlinear Dyn. **25**, 201–206 (1989).
[43]R. S. Sigmond, A. Goldman, and M. Goldman, "Ring vortex gas flow in negative point coronas," in *Proceedings of the 10th International Conference on Gas Discharge and Their Applications*, 330–333 Swansea, U.K., 1992.
[44]R. S. Sigmond and I. H. Lågstad, "Mass and species transport in corona discharges," Plasma Phys. Rep. **2**, 221–229 (1993).
[45]J. Batina, F. Noël, S. Lachaud, R. Peyrous, and J. F. Loiseau, "Hydrodynamical simulation of the electric wind in a cylindrical vessel with positive point-to-plane device," J. Phys. D **34**, 1510–1524 (2001).
[46]Ph. Béquin, K. Castor, and J. Scholten, "Electric wind characterization in negative point-to-plane corona discharges in air," Eur. Phys. J.: Appl. Phys. **22**, 41–49 (2003).
[47]M. M. Kekez, P. Savic, and G. D. Lougheed, "A novel treatment of Trichel type phenomena with possible application to stepped-leader phenomena," J. Phys. D **15**, 1963–1973 (1982).
[48]Ph. Béquin, Modèles de sorces acoustiques à gaz ionisé ("Model of acoustic sources using ionised gas") (Ph.D. dissertation, Université du Maine,

Le Mans, France) (1994).
[49]V. Montembault, Etude des sources acoustiques associées aux décharges corona négatives ("Study of acoustic sources related to negative corona discharges") (Ph.D. dissertation, Université du Maine, Le Mans, France) (1997).
[50]K. Castor, Caractérisation des sources acoustiques associées aux décharges couronnes négatives ("Characterization of acoustic sources related to negative corona discharges") (Ph.D. dissertation, Université du Maine, Le Mans, France) (2001).
[51]H. Bondar and F. Bastien, "Effect of neutral fluid velocity on direct conversion from electrical to fluid kinetic energy in an electro-fluid-dynamics (EFP) device," J. Phys. D **19**, 1657–1663 (1986).
[52]K. Castor and Ph. Béquin, "Rendement d'un haut-parleur à décharges corona négatives" ("Efficiency of negative corona discharges loudspeaker") in *Proceedings of the 5th International French Congress on Acoustics*, Lausanne, 676–679 (2000).

Theoretical foundations of apparent-damping phenomena and nearly irreversible energy exchange in linear conservative systems

A. Carcaterra[a)]
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza" Via Eudossiana, 18, 00184, Rome, Italy, and Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

A. Akay
Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 22 June 2006; revised 17 January 2007; accepted 19 January 2007)

This paper discusses a class of unexpected irreversible phenomena that can develop in linear conservative systems and provides a theoretical foundation that explains the underlying principles. Recent studies have shown that energy can be introduced to a linear system with near irreversibility, or energy within a system can migrate to a subsystem nearly irreversibly, even in the absence of dissipation, provided that the system has a particular natural frequency distribution. The present work introduces a general theory that provides a mathematical foundation and a physical explanation for the near irreversibility phenomena observed and reported in previous publications. Inspired by the properties of probability distribution functions, the general formulation developed here is based on particular properties of harmonic series, which form the common basis of linear dynamic system models. The results demonstrate the existence of a special class of linear nondissipative dynamic systems that exhibit nearly irreversible energy exchange and possess a decaying impulse response. In addition to uncovering a new class of dynamic system properties, the results have far-reaching implications in engineering applications where classical vibration damping or absorption techniques may not be effective. Furthermore, the results also support the notion of nearly irreversible energy transfer in conservative linear systems, which until now has been a concept associated exclusively with nonlinear systems. [DOI: 10.1121/1.2697030]

PACS number(s): 43.40.At, 43.40.Kd, 43.40.Jc [ADP] Pages: 1971–1982

I. INTRODUCTION

[1]In complex built-up structures, comprised of many individual structural components, the spatial redistribution of vibratory energy throughout the structure, in many ways, seems similar to reduction of vibration due to classical dissipation mechanisms and "appears" as damping, but distinct from dissipation of vibratory energy as heat.

In a complex structure that consists of a set of satellite oscillators attached to a common vibrating master, even a light damping in the oscillators can be amplified, as observed in the master impulse response.[1–3] The observed damping in such cases, however, is a combination of classical and apparent damping, with increased rate of vibration energy flow from the master to the attached set which makes the master motion appear to be strongly damped, although its energy is not dissipated as heat at the same rate.

In cases where both the master and the satellites are conservative,[4–11] in general, energy flows rapidly and almost completely from the master to the satellites followed by a complete return to the master, and the cycle repeats itself.[4,8] This is consistent with the notion that in linear undamped systems, energy transfer from the master to the satellite oscillators is reversible. The duration during which energy is trapped in the attached set can be estimated[8] and is shown to change with the frequency distribution within the set of oscillators. Furthermore, as shown in several recent numerical simulations and experimental studies, certain frequency distributions of the attached oscillators prevent any energy return from the satellites to the master.[9,10] Such frequency distributions make the time it takes for the energy to return to the master structure so long that energy return is not observed as a result of inherent losses in physical systems, making the energy transfer "nearly irreversible."

Analogous behavior exists in continuous systems characterized by an impulse response that has a sharp decrease of amplitude with time, as shown for a beam-like undamped structure.[11] This particular, and counter-intuitive, response appears closely related to the discrete systems described above and shows how the presence of a condensation point in the natural frequency distribution of a continuous structure can also lead to apparent damping.

This paper provides a theoretical foundation for determining frequency distributions that can provide apparent damping properties to structures, such as those described above, which were obtained through optimization and experimental studies.

[a)]Author to whom correspondence should be addressed. Electronic mail: a.carcaterra@dma.ing.uniroma1.it

Although most conservative systems do not exhibit a decaying trend in their impulse response, as shown in this paper, there exists a class of conservative systems with apparent damping properties based on nearly irreversible energy transfer. The analysis developed in this paper provides the theoretical foundation for apparent damping and the special characteristic that a linear conservative system must possess to exhibit strong apparent damping and near irreversibility.

The paper, similar to others that treat linear undamped structures,[4–13] provides a different perspective on the basic mechanisms of energy sharing among complex subsystems.[7–13] Specifically, the absence of dissipation in the systems considered in the present paper leads to new insights in the general mechanism of near-irreversible energy redistribution among the degrees of freedom of large structures. Concepts of irreversibility and energy redistribution continue to appear in other areas of physics, for example, in the context of atomic chains and lattices.[14–16] Although it is widely accepted that the root of irreversible energy redistribution rests on the presence of nonlinear interactions among the atoms, recently another possible scenario for irreversibility in linear systems has been offered.[17] The present work brings a contribution to this discussion outlining a general mechanism for the development of nearly irreversible processes in complex linear conservative systems.

II. COMPARISON OF DISCRETE AND INTEGRAL SUMMATIONS

In many problems of mechanics, a combination of harmonic functions represents the impulse response $S(t)$ of a linear dynamical system, which commonly yields a discrete eigenfrequency spectrum

$$S(t) = \frac{1}{N}\sum_{i=1}^{N} G_i \sin \omega_i t, \tag{1}$$

where G_i, ω_i, N represent the modal participation factor, system natural frequencies and the number of the modes involved in the system response. Except for the well known case where frequencies ω_i are integer multiples of a fundamental frequency, the properties of a harmonic series such as $S(t)$ can be quite complicated. As suggested in Refs. 8 and 10 certain relevant properties of the series $S(t)$ emerge when compared with its integral form

$$I(t) = \int_0^{\xi_{\max}} G[\omega(\xi)] \sin \omega(\xi) t d\xi, \tag{2}$$

where $G(\omega)$ and $\omega(\xi)$ are two arbitrary functions that relate to their discrete counterparts as

$$G(\omega_i) = G_i, \quad \omega(\xi_i) = \omega_i, \quad \xi_i = \frac{i}{N}\xi_{\max}, \quad i = 1,2, \ldots, N.$$

Analysis presented in Refs. 8 and 10 showed that if the function $\omega(\xi)$ has a nonvanishing derivative or vanishes at a finite number of points, in the interval $\xi \in [0, \xi_{\max}]$, then $I(t)$ has the following fundamental asymptotic property:

$$\lim_{t\to\infty} I(t) = 0. \tag{3}$$

Although $S(t)$ can be considered a discrete approximation of $I(t)$, the asymptotic property in Eq. (3) may not be directly generalized to $S(t)$, since the two functions $I(t)$ and $S(t)$ differ by a remainder term.[8,10] In general, an impulse response of the type $S(t)$ for conservative systems does not vanish asymptotically but behaves as an *almost periodic function*. The theory outlined in this paper investigates the properties of the remainder term between the integral $I(t)$ and the series $S(t)$ to identify the conditions under which $S(t)$ develops the asymptotic property in Eq. (3) as closely as possible. In particular, the following sections describe the conditions under which the series $S(t)$ approaches the integral $I(t)$ and show that a criterion of minimum distance $D(t)$

$$D^2(t) = \int_C [S - I]^2 W dC = \overline{(S - I)^2}$$

can be satisfied with a suitable weighting function W in a prescribed domain C within a certain space Σ.

The results show that for a given $G(\omega)$, there exists a class of functions $\omega(\xi)$ that minimizes the distance between $S(t)$ and $I(t)$. In such cases, the series $S(t)$ tends to closely match the asymptotically vanishing trend of the integral $I(t)$, producing apparent damping effects through nearly irreversible energy transfer processes in conservative linear systems. The next section reviews the definitions and properties necessary to form the basis for the ensuing theoretical development.

III. AVERAGES IN THE SPACE OF HARMONICS: DEFINITIONS AND PROPERTIES

The analysis eveloped in the subsequent sections involves averaged values of discrete summations and integrals.

For a set of functions $s_i = G_i \sin \omega_i t$, $i = 1,2, \ldots, N$ at any time t, $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$ defines a point (or a vector) in the space Σ of harmonics; $\mathbf{s}$ exists within the hypercube $C \equiv \{E \times E \times \ldots \times E\}$, with $E \equiv [-G_{\max}, G_{\max}]$, $G_{\max} = max\{G_1, G_2, \ldots, G_N\}$.

Let $f(\mathbf{s})$ be an arbitrary function defined over $C \subset \Sigma$ with the vector $\mathbf{s} \in \Sigma$. In general, the average value $\bar{f}$ of $f(\mathbf{s})$ over C can be expressed using a weighting function $P(\boldsymbol{s}, I)$ as

$$\bar{f}(I) = \int_C f(\mathbf{s}) P(\mathbf{s}, I) dC,$$

where the integration acts on the variable $\mathbf{s}$, while I plays the role of a parameter. Time appears through I

$$\bar{f}[I(t)] = \int_C f(\mathbf{s}) P[\mathbf{s}, I(t)] dC$$

or, more concisely

$$\bar{f}(t) = \int_C f(\mathbf{s}) P(\mathbf{s}, I) dC, \quad dC = \prod_{k=1}^{N} ds_k. \tag{4}$$

The weighting function P is selected to depend on $\mathbf{s}$ and I as described below.

As a consequence, scalar product of the two functions $f(\mathbf{s})$ and $g(\mathbf{s})$ in C follows as

$$\overline{f \cdot g} = \int_C f(\mathbf{s})g(\mathbf{s})P(\mathbf{s},I)dC. \tag{5}$$

Similarly, the distance $D(t)$ between $f(\mathbf{s})$ and $g(\mathbf{s})$ follows:

$$D^2(t) = \overline{(f-g)\cdot(f-g)} = \overline{(f-g)^2} = \int_C [f(\mathbf{s}) - g(\mathbf{s})]^2 P(\mathbf{s},I)dC. \tag{6}$$

In this context, the two functions are said to be orthogonal if $\overline{f \cdot g} = 0$ and parallel if $\overline{f \cdot g} = \sqrt{\overline{f \cdot f}}\sqrt{\overline{g \cdot g}}$.

A. Weighting function, $P(\boldsymbol{s},\boldsymbol{I})$

The weighting function $P(\mathbf{s},I)$ in Eq. (6) is selected to have the form

$$P(\mathbf{s},I) = \prod_{k=1}^{N} p(s_k,I), \tag{7}$$

where $p(s_k,I)$ is an arbitrary function that satisfies the conditions

$$\int_{-G_{\max}}^{G_{\max}} \sigma p(\sigma,I)d\sigma = I(t). \tag{8}$$

$$\int_{-G_{\max}}^{G_{\max}} p(\sigma,I)d\sigma = 1. \tag{9}$$

Equation (8) offers a comparison with the integral in Eq. (2) for $\sigma = G(\omega)\sin \omega t$ and through a change of integration variables in Eq. (8) (first from $d\sigma$ to $d\omega$, then to $d\xi$)

$$I(t) = \int_{-G_{\max}}^{G_{\max}} \sigma p(\sigma,I)d\sigma = \int_{\omega_{\min}}^{\omega_{\max}} \sigma(\omega)p[\sigma(\omega),I]\frac{d\sigma}{d\omega}d\omega,$$

$$I(t) = \int_0^{\xi_{\max}} \sigma[\omega(\xi)]p\{\sigma[\omega(\xi)],I\}\frac{d\sigma}{d\omega}\frac{d\omega}{d\xi}d\xi \tag{10}$$

provided that in the interval $[\omega_{\min},\omega_{\max}]$, $\sigma = G(\omega)\sin \omega t$ is single valued and $\sigma \in [-G_{\max}, G_{\max}]$. A comparison of Eqs. (2) and (10) implies that the function $p(\sigma,I)$ must satisfy the following compatibility condition:

$$p(\sigma,I)\frac{d\sigma}{d\omega}\frac{d\omega}{d\xi} = 1. \tag{11}$$

The condition expressed by Eq. (11) also implies a dependence between $p(\sigma,I)$ and $\omega(\xi)$ for $\sigma = G(\omega)\sin \omega t$.

It follows that substituting Eq. (11) in Eq. (9) yields the upper bound of ξ as

$$\int_{-G_{\max}}^{G_{\max}} p(\sigma,I)d\sigma = \int_0^{\xi_{\max}} p\{\sigma[\omega(\xi)],I\}\frac{d\sigma}{d\omega}\frac{d\omega}{d\xi}d\xi = \xi_{\max}$$

yielding $\xi_{\max} = 1$, which leads to the conclusion about the bounds of ξ as: $\xi \in [0,1]$.

B. Averages in the Σ space

The average of the function $S(t)$ in Eq. (1) can be expressed by substituting for $\bar{s}_i$ the averaging property expressed for $\bar{f}(t)-$ in Eq. (4)

$$\bar{S}(t) = \frac{1}{N}\sum_{i=1}^{N} \bar{s}_i = \frac{1}{N}\sum_{i=1}^{N} \int_C s_i P(\mathbf{s},I)dC.$$

Further substitution for $P(\boldsymbol{s},I)$ from Eq. (7) and for dC from Eq. (4) yields

$$\bar{S}(t) = \frac{1}{N}\sum_{i=1}^{N} \int_{-G_{\max}\ s}^{G_{\max}} s_i p(s_i,I)ds_i \int_{C^{(N-1)}} \prod_{k\neq i}^{N} p(s_k,I)ds_k.$$

The condition (9) leads the mutiplication series in the second integral to produce identity and applying condition (8) to the first integral shows that

$$\bar{S}(t) = I(t). \tag{12}$$

By invoking the definition of average value in Eq. (4) leads to the fundamental result

$$\bar{S}(t) = \int_C SP(\mathbf{s},I)dC = I(t) \tag{13}$$

provided that Eq. (11) is satisfied.

Similarly, substituting for $P(\mathbf{s},I)$ from Eq. (7) and for dC from Eq. (4) and invoking the condition in Eq. (9) it can be show that

$$\int_C P(\mathbf{s},I)dC = 1. \tag{14}$$

The first derivative of Eq. (14) with respect to I can be expressed as

$$\int_C \frac{\partial}{\partial I}P(\mathbf{s},I)dC = 0 \rightarrow \int_C \left[\frac{1}{P(\mathbf{s},I)}\frac{\partial}{\partial I}P(\mathbf{s},I)\right]P(\mathbf{s},I)dC = 0.$$

Thus

$$\int_C \frac{\partial}{\partial I}[\log P(\mathbf{s},I)]P(\mathbf{s},I)dC = 0$$

which is equivalent to stating

$$\overline{\frac{\partial}{\partial I}\log P} = 0. \tag{15}$$

Following the same approach, the first derivative of Eq. (13) with respect to I produces a similar expression

$$\int_C S\frac{\partial}{\partial I}P(\mathbf{s},I)dC = 1 \rightarrow \int_C S\left[\frac{1}{P(\mathbf{s},I)}\frac{\partial}{\partial I}P(\mathbf{s},I)\right]P(\mathbf{s},I)dC = 1$$

$$\overline{S\frac{\partial}{\partial I}\log P} = 1. \tag{16}$$

In order to find the distance between $S(t)$ and $I(t)$, an equivalent expression for Eq. (16) is developed for $I(t)$ by multiplying Eq. (15) by the factor I, which is independent of the integration variable s

$$\overline{I\frac{\partial}{\partial I}\log P} = 0. \tag{17}$$

IV. FUNDAMENTAL INEQUALITY FOR THE DISTANCE BETWEEN INTEGRAL AND SUMMATION

This section discusses a method to find the best choice for the frequency distribution $\omega(\xi)$ to minimize the distance between $S(t)$ and $I(t)$ in the space of harmonics, Σ. Finding P such as the distance D is minimum is a problem of functional analysis. Two approaches are employed to provide the solution. The first, illustrated in this section, makes direct use of the Schwartz inequality, while the second, detailed in Appendix A, is based on the calculus of variations.

The difference between Eqs. (16) and (17) produces a condition on the scalar product between the two functions $(S-I)$ and $(\partial/\partial I \log P)$ as

$$\overline{(S-I)\left(\frac{\partial}{\partial I}\log P\right)} = 1.$$

The Schwartz inequality provides that

$$\overline{(S-I)\cdot\left(\frac{\partial}{\partial I}\log P\right)^2} \leqslant \overline{(S-I)^2}\,\overline{\left(\frac{\partial}{\partial I}\log P\right)^2}. \tag{18}$$

Thus, the distance D between S and I satisfies the inequality

$$\overline{(S-I)^2} = D^2 \geqslant \frac{1}{\overline{\left(\frac{\partial}{\partial I}\log P\right)^2}}.$$

The right-hand side of previous inequality presents a lower bound for D, depending on the weighting function P or, equivalently, depending on the frequency distribution $\omega(\xi)$. Among the possible frequency distributions, those that provide the minimum distance between S and I match the lower bound exactly, such that

$$D^2 = \frac{1}{\overline{\left(\frac{\partial}{\partial I}\log P\right)^2}}.$$

The functions P provide the *minimum distance* between S and I and, thus, in a sense, satisfy an optimum condition. The next section describes how to determine the analytical form of this special family of weighting functions.

V. WEIGHTING FUNCTIONS FOR MINIMUM DISTANCE

As the minimum distance bound is reached, the Schwartz inequality (18) becomes an equality. This happens when the scalar product on the left-hand side consists of two parallel functions, that means the functions $(S-I)$ and $(\partial/\partial I \log P)$ involved in the scalar product that lead to Eq. (18) become proportional, i.e.,

$$\frac{\partial}{\partial I}\log P = a(I)(S-I), \tag{19}$$

where $a(I)$ is a proportionality constant, independent of the variable s, and depends only on I.

An identical solution can be obtained by the calculus of variation looking at the P that produces a minimum for the functional D^2 constrained by the condition $(S-I)\cdot(\partial/\partial I \log P)=1$ (see Appendix A).

Condition (19) represents a differential equation in terms of P, and its solution leads to a family of exponential functions $P(\mathbf{s},I)=\prod_{k=1}^{N} p(s_k,I)$. The solution to Eq. (19), originally obtained by Pitman and Koopman in the context of the theory of estimators,[18–20] is given as

$$p(\sigma,I) = \exp\{\alpha(I)\varepsilon(\sigma) + \zeta(\sigma) + \beta(I)\},$$

where $\alpha(I)$, $\beta(I)$, $\varepsilon(\sigma)$, *and* $\zeta(\sigma)$ are arbitrary functions of their respective arguments.

Gauss function also belongs to this family of solutions and provides an excellent example that can be easily verified by substituting into Eq. (19)

$$p(\sigma,I) = \frac{1}{r\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(\sigma-I)^2}{r^2}\right\}. \tag{20}$$

The solution $p(\sigma,I)$, with $\sigma(G,\omega)$, has a shape that depends on the function $I(t)$ and on the parameter r.

Together with Eq. (20), the compatibility Eq. (11) becomes a nonlinear differential equation

$$\frac{1}{r\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(\sigma-I)^2}{r^2}\right\}\frac{d}{d\omega}[G(\omega)\sin\omega t]\frac{d\omega}{d\xi} = 1. \tag{21}$$

and its solution provides the frequency distribution $\omega(\xi)$ that minimizes the distance D. Equation (21) can be solved for $\omega(\xi)$ numerically; however, an alternative approach using density of harmonic functions, analogous to modal density in a dynamical system, produces a closed-form expression. Since $d\xi/d\omega N$ represents the harmonic density, $\delta(\omega)$, that counts the number of harmonics, or modes, contained within the frequency band $d\omega$, Eq. (21) directly leads to an expression for $\delta(\omega)$. With $d\xi=1/Ndn$, N being the total number of harmonics, s_i, for $\xi\in[0,1]$, and dn the number of harmonics for $\xi\in[\xi,\xi+d\xi]$, it follows that $Nd\omega/d\xi\propto d\omega/dn=1/\delta(\omega)$. Substituting in Eq. (21) produces

$$\delta_{\text{opt}}(t) = \frac{1}{N}\frac{1}{r\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(\sigma-I)^2}{r^2}\right\}\frac{d}{d\omega}[G(\omega)\sin\omega t]. \tag{22}$$

Equations (21) and (22) show that time appears as a parameter in the frequency distribution that minimizes the difference between S and I. Oscillators with time-dependent parameters or, equivalently, with time-varying natural frequencies, imply the presence of parametrically controlled

resonators or nonlinear resonators. The problem under consideration in this paper addresses linear time-invariant dynamical systems and, thus, Eqs. (21) and (22) cannot be satisfied for all times t. Thus, the approach taken here uses the frequency distribution $\omega(\xi)$ that results from Eqs. (21) and (22) for a particular time t_0 to solve the compatibility equation

$$p(\sigma,I)\frac{d\sigma}{d\omega}\frac{d\omega}{d\xi}\bigg|_{t=t_0} = 1. \tag{23}$$

The choice for t_0, selection of the frequency interval $[\omega_{\min},\omega_{\max}]$ within which $\omega(\xi)$ falls, which also depends the choice of t_0, and the implication of their selection are discussed with examples in the next sections.

Normally, the form of Eq. (20) satisfies Eqs. (8) and (9) automatically for an integration domain $[-\infty,+\infty]$; however, since the actual domain is finite $E\equiv[-G_{\max},G_{\max}]$, r and $I(t_0)$ must satisfy the additional constraints

$$r \ll G_{\max}, \quad I(t_0) \in E. \tag{24}$$

These constraints guarantee that the function represented by Eq. (20) has its primary distribution within the interval E and therefore (approximately) satisfying Eqs. (5) and (6).

VI. APPLICATION OF THE THEORY AND EXAMPLES

The examples given in this section illustrate application of the theory described above. Each case demonstrates how to minimize the difference between a sum of harmonic functions and the corresponding integral summation. The first example consists of a simple summation of sine functions for which $G(\omega)\equiv 1$. In the second example, $G(\omega)\equiv\omega$ represents the reaction force of a set of undamped resonators on a common rigid base.[10] Sections VII and VIII examine more complex examples.

A. Simple sine series, $G(\omega)\equiv 1$

Summation of a series of $N=100$ sine functions with frequencies ω_i results from Eq. (1) by substituting for $G(\omega)\equiv 1$

$$S(t) = \frac{1}{N}\sum_{i=1}^{N} \sin \omega_i t$$

and the corresponding integral from Eq. (2) becomes

$$I(t) = \int_0^1 \sin \omega(\xi) t d\xi.$$

For this case, the nonlinear differential Eq. (21) becomes

$$\frac{1}{r\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(\sigma-I_0)^2}{r^2}\right\} t_0 \cos \omega t_0 \frac{d\omega}{d\xi} = 1, \tag{25}$$

where σ and I_0 from Eqs. (23) and (2) become

$$\sigma = \sin \omega(\xi) t_0, \quad I_0 = \int_0^{\xi_{\max}} \sin \omega(\xi) t_0 d\xi.$$

In this case, $E\equiv[-1,1]$. Restricting the selection to monotonic frequency distributions $\omega(\xi)$, so that $d\omega/d\xi>0$ for $\omega \in [\omega_{\min},\omega_{\max}]$, implies that, according to Eq. (25), $d\sigma/d\omega=\cos \omega(\xi) t_0$ must always be positive for $\omega \in [\omega_{\min},\omega_{\max}]$. It follows that assigning, for example, $t_0=\pi/4$, yields $d\omega/d\xi>0$ for $\omega\in[0,1]$. Under these conditions, the values for I_0 may be arbitrary, except that they must satisfy the inequalities in Eq. (24) and p satisfies the conditions (7) and (11).

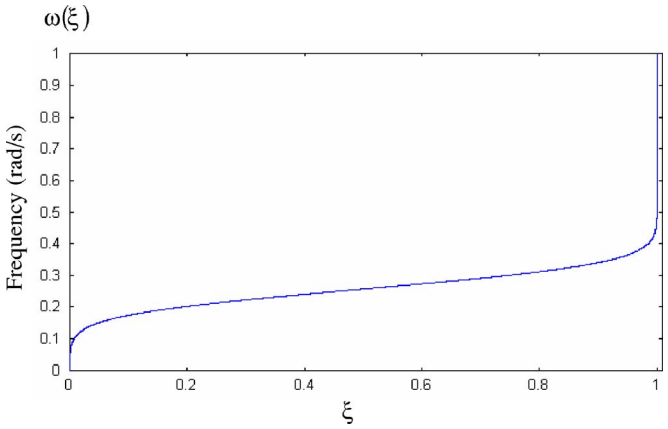


FIG. 1. Optimal frequency distribution for $N=100$, $t_0=\pi/4$, $r=0.05$, $\omega\in[0,1]$.

Figures 1–5 illustrate the results obtained for $N=100$, $t_0=\pi/4$, $r=0.05$, $\omega\in[0,1]$ and with the choice of $I_0=0.2$ and $r\ll 1$, both of which satisfy Eq. (24). Figure 1 represents the frequency distribution $\omega(\xi)$ determined by a numerical integration of Eq. (25) from which the set $\omega_i(i=1,..100)$ is determined by sampling 100 points equally spaced along ξ. Figure 2 represents the harmonic density and Fig. 3 the time history of $S(t)$. As shown in Figs. 4 and 5 for different time scales, in the time history of the series obtained using the theory developed here the strong periodicity disappears when compared with the corresponding series consisting of a linear frequency distribution with period $2\pi N/\omega_{\max}$ for two different time scales.

Figures 6–9 show a case analogous to the previous one except for $r=0.01$, representing a higher harmonic density around its peak resulting in a somewhat better performance.

Finally, the third example, shown in Figs. 10–12, uses $t_0=\pi/8$ and $N=100$, $r=0.05$, $\omega\in[0,1]$.

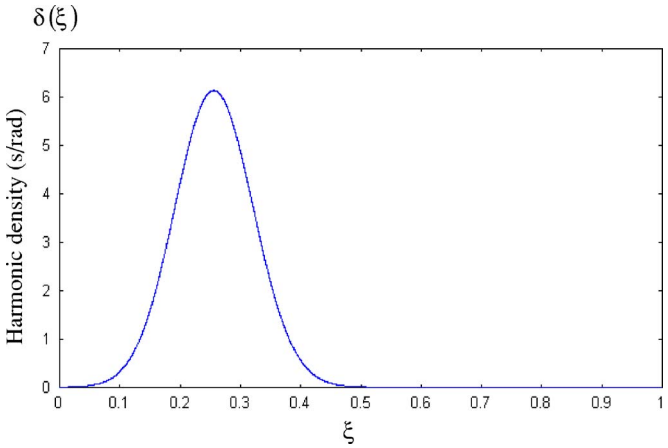


FIG. 2. Modal density for $N=100$, $t_0=\pi/4$, $r=0.05$, $\omega\in[0,1]$.

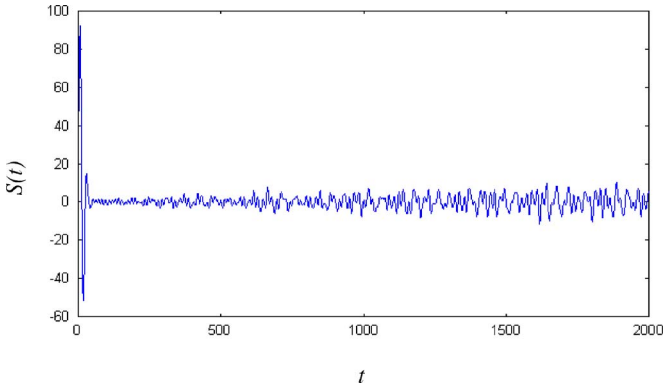


FIG. 3. Time history of the series (1) for $N=100$, $t_0=\pi/4$, $r=0.05$, $\omega \in [0,1]$.

The results of the examples above show that the frequency distributions satisfying the minimum distance bound requirements produce time histories that bring the summation $S(t)$ very close to $I(t)$, without recurrence or periodicity in its time history unlike, say, the case of a linear frequency distribution. The envelope of the summation in Eq. (1) decays significantly with respect to its early oscillations and without regaining its initial amplitude, following closely the same trend that its integral counterpart $I(t)$ exhibits in Eq. (3).

Finally, it is interesting to underline how the choice of the frequency distribution cannot indeed affect the time average energy of the function $S(t)$. In fact

$$\langle S^2(t)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T S^2(t)dt$$

$$= \lim_{T\to\infty}\frac{1}{T}\frac{1}{N^2}\sum_{i,j=1}^{N} G(\omega_i)G(\omega_j)\int_0^T \sin\omega_i t \sin\omega_j t dt$$

$$\langle S^2(t)\rangle = \frac{1}{2N^2}\sum_{i=1}^{N} G^2(\omega_i)$$

that in the present case, $G(\omega)=1$, simplifies as $\langle S^2(t)\rangle = 1/2N$. This means that the time average energy of $S(t)$ is in this case invariant and independent of the frequency distribution. This makes clear how, as it appears, for example, from Figs. 4 and 5, the energy of the two compared signals is the same and the choice of the frequency distribution only affects the way this energy distributes along the time axis.

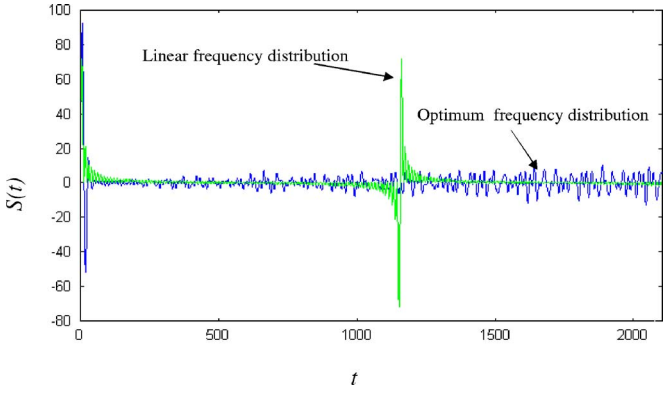


FIG. 4. Comparison between time histories obtained by the optimal and the linear frequency distribution, short time range.

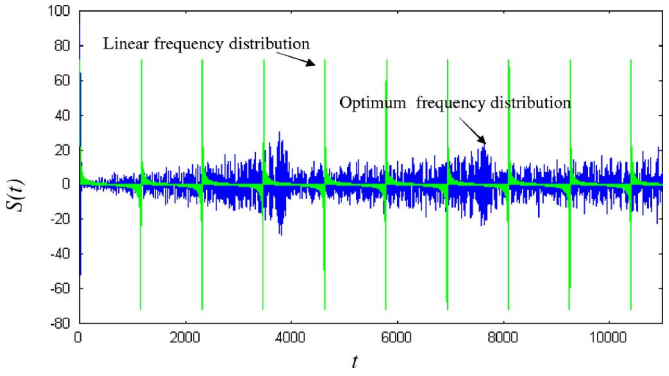


FIG. 5. Comparison between time histories obtained the optimal and the linear frequency distribution, long time range.

B. Reaction force of a set of oscillators attached to a rigid base, $G(\omega)\equiv\omega$

Consider a set of N parallel oscillators attached to a common rigid base. Oscillators have equal mass m and uncoupled natural frequencies $\omega_i=\sqrt{k_i/m}$, where k_i represents the stiffness of oscillator i. Impulse response of each oscillator is expressed as

$$h_i(t) = \frac{1}{m\omega_i}\sin\omega_i t.$$

The total reaction force exerted on the base by a set of N oscillators can be represented as

$$S(t) = \sum_{i=1}^{N} k_i h_i(t) = \sum_{i=1}^{N} \omega_i \sin\omega_i t.$$

$S(t)$ has the same form as the series in Eq. (1) with $G\equiv\omega$. In this case, for time $t=t_0$, Eq. (23) together with Eq. (20) for $p(\sigma)$, provides

$$\frac{1}{r\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(\sigma-I_0)^2}{r^2}\right\}[\sin\omega t_0 + t_0\cos\omega t_0]\frac{d\omega}{d\xi} = 1$$

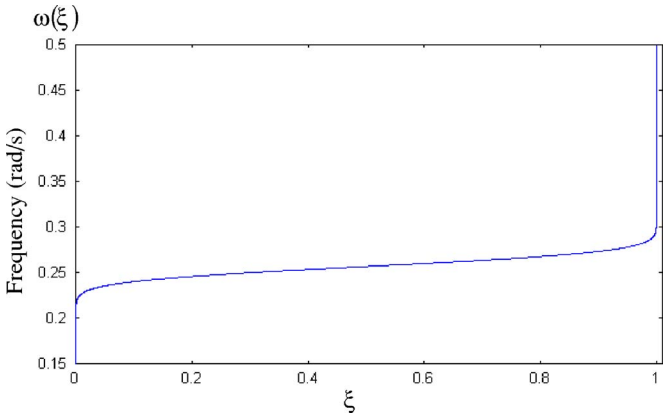


FIG. 6. Optimal frequency distribution for $N=100$, $t_0=\pi/4$, $r=0.01$, $\omega \in [0,1]$.

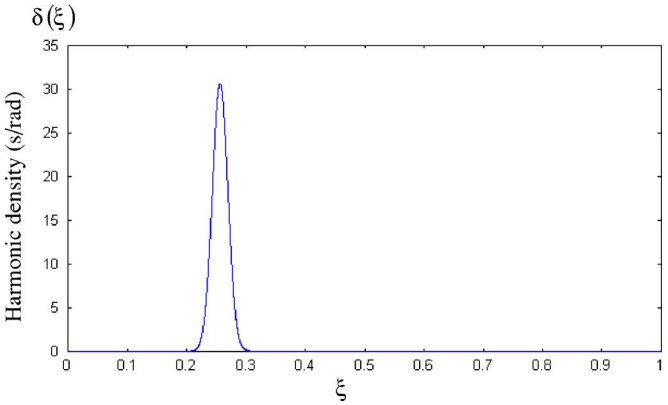


FIG. 7. Modal density for $N=100$, $t_0=\pi/4$, $r=0.01$, $\omega\in[0,1]$.

$$\sigma=\omega(\xi)\sin\,\omega(\xi)t_0,\quad I_0=\int_0^{\xi_{\max}}\omega(\xi)\sin\,\omega(\xi)t_0 d\xi. \quad (26)$$

As before, restricting the analysis only to monotonic frequency distributions $\omega(\xi)$, such that $d\sigma/d\omega=\sin\,\omega t_0+t_0\cos\,\omega t_0>0$, and choosing, for example, $t_0=\pi/4$, leads to the condition that in the frequency interval $\omega\in[0,2]$, $d\sigma/d\omega>0$ and $\sigma\in[0,2]$. Again, r and I_0 can be assigned arbitrarily, but consistent with inequalities (24); in this case, $r=0.1$ and $I_0=0.8$.

Figure 13 displays the frequency distribution obtained by solving Eq. (26) and Fig. 14 shows the corresponding optimal modal density from Eq. (22). The time history of the reaction force on the rigid base, shown in Fig. 15, exhibits a rapid decay and remains at a negligibly low amplitude.

VII. APPARENT DAMPING IN CONSERVATIVE CONTINUOUS STRUCTURES $G(\omega)\equiv 1/\omega$

A continuous linear undamped dynamic system, excited by a unit impulse at point x_0, satisfies the equation of motion

$$L[w(x,t)]+m'\frac{\partial^2 w(x,t)}{\partial t^2}=0$$

with initial conditions $w(x,0)=0$, $\dot{w}(x,0)=\delta(x-x_0)$, where δ is the Dirac's distribution. $L[\,]$, $w(x,t)$, m' represent the system operator, the displacement response and the mass density, respectively. The general response of such a linear system can be represented by its orthonormal modes $\Phi_i(x)$ and principal coordinates $q_i(t)$, as

$$w(x,t)=\sum_{i=1}^{N}\Phi_i(x)q_i(t)$$

$$q_i(t)=A_i\sin\,\omega_i t,\quad A_i=\frac{m'}{\omega_i}\Phi_i(x_0).$$

Then its impulse response at x_0 can be represented by the series expression $S(t)$ in Eq. (1), with $G_i=Nm'/\omega_i\Phi_i^2(x_0)$.

In general, in the absence of damping, this finite series, a superposition of pure sine functions, exhibits an almost-periodic trend. For example, the case of a Fourier series of sine functions with linearly distributed frequencies $\omega_i=i\omega_0$, where ω_0 is the fundamental frequency, becomes periodic. As before, a decaying trend in $S(t)$ is expected only in the presence of energy dissipation. However, as shown in previous studies[10] that in cases where condensation points exist within the frequency distribution or, equivalently, natural frequencies accumulate around a particular frequency, impulse response of that linear system exhibits a decaying characteristic even in the absence of dissipation sources, a phenomenon referred to here as near irreversibility or apparent damping.

Application of the theory developed in this paper to the continuous system described above provides a theoretical ba-

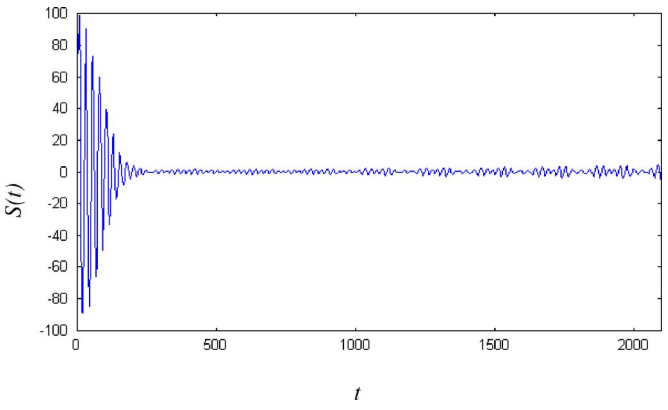


FIG. 8. Time history of the series (1) for $N=100$, $t_0=\pi/4$, $r=0.01$, $\omega\in[0,1]$.

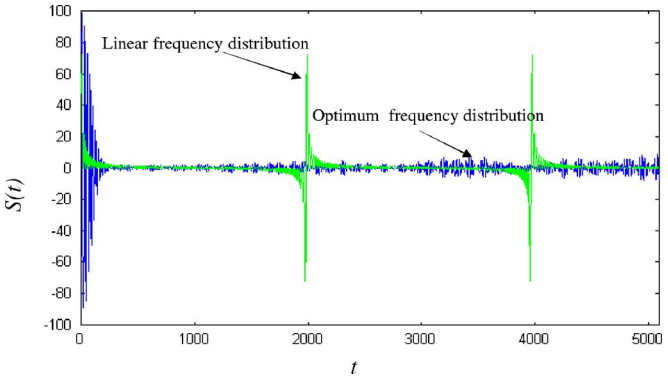


FIG. 9. Comparison between time histories obtained by the optimal and the linear frequency.

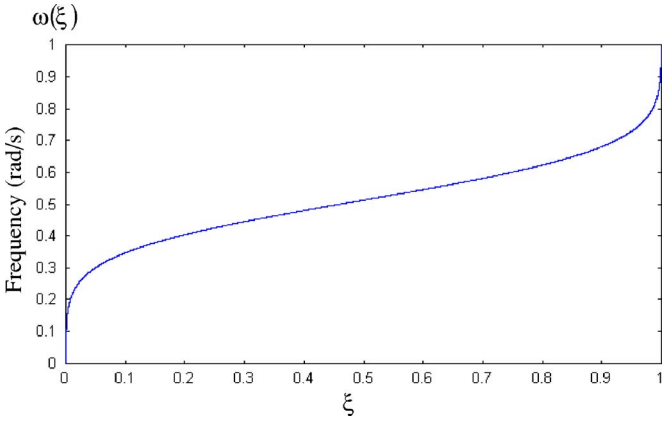


FIG. 10. Optimal frequency distribution for $N=100$, $t_0=\pi/8$, $r=0.05$, $\omega\in[0,1]$.

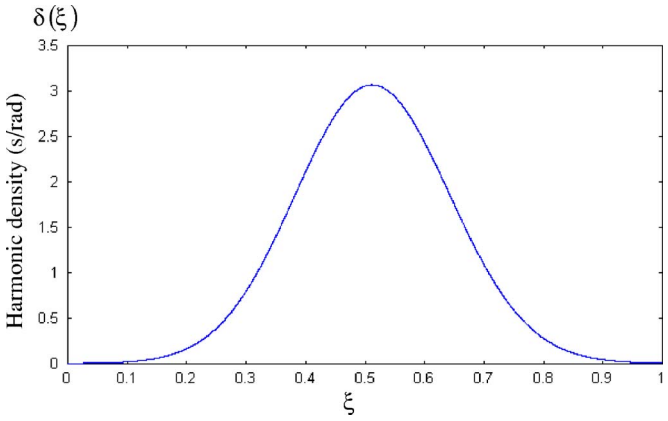


FIG. 11. Modal density for $N=100$, $t_0=\pi/8$, $r=0.05$, $\omega\in[0,1]$.

sis to the numerically obtained results in earlier studies and demonstrates how a class of frequency distributions ω_i can produce apparent damping.

As an example, consider a simply supported beam as a prototypical linear system for which $\Phi_i(x_0)=\sqrt{2/m'L}\sin\pi i x_0/L$, $x_0/L=1/2$. Substituting for $G_i=1/\omega_i 2/L(\sin\pi i/2)^2$ in Eq. (1), and retaining only the odd terms:

$$S(t)=\sum_{i=1}^{N/2}\frac{2}{L}\frac{1}{\omega_{2i-1}}\sin\omega_{2i-1}t.$$

Selecting $t_0=\pi/4$, the function $\sigma=1/\omega\sin\omega t_0$ has a monotonically increasing trend, for example, within the interval $\omega\in$.[8,9] Choosing values $I_0=0.05$, $r=0.005$, which satisfy the conditions (24), the frequency distribution and the corresponding modal density can be obtained from Eqs. (21) and (22), as shown in Figs. 16 and 17. The impulse response of the beam with such a frequency distribution is shown in Fig. 18 for $N=200$ (but includes only the 100 odd modes).

VIII. NEARLY IRREVERSIBLE ENERGY TRANSFER BETWEEN A SIMPLE RESONATOR AND A SET OF PARALLEL OSCILLATORS $G(\omega)\equiv\omega^3$

Figure 19 depicts the system under consideration in this section, which consists of set of resonators with natural frequencies $\omega_i(i=1,2\ldots N)$, that are connected in parallel to a

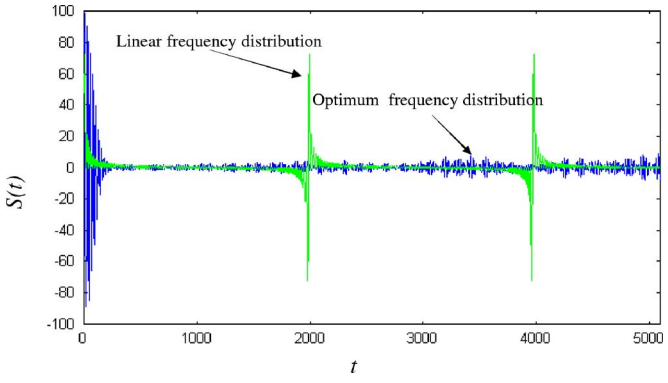


FIG. 12. Comparison between the time histories of the optimal and the linear frequency distribution.

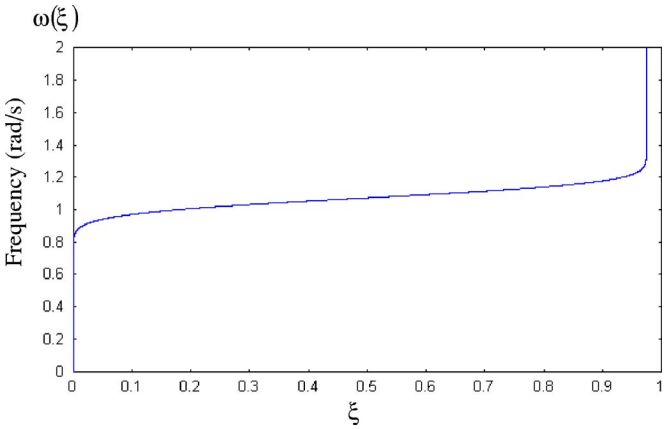


FIG. 13. Natural frequency distribution of the resonator within the set for $N=100$, $t_0=\pi/4$, $r=0.1$, $\omega\in[0,2]$.

common principal structure. The system does not possess any means of energy dissipation. For a very large number of oscillators N, approaching infinity, but with a constant total mass, the attached oscillators can be considered as a continuous distribution with the equations of motion for the coupled system expressed as

$$\begin{cases} M\ddot{x}_M(t)+K_M x_M(t)+\displaystyle\int_0^1 k(\xi)(x_M(t)-x(\xi,t))d\xi=0, \\ m\ddot{x}(\xi,t)-k(\xi)(x_M(t)-x(\xi,t))=0, \end{cases} \tag{27}$$

where M,K_M,x_M are the mass, stiffness and displacement of the master structure, respectively; m,k,x represent the same quantities of the distributed oscillators in the attached set.

Several studies have shown that such a distribution of oscillators produces a damping effect on the principal mass[1–8] as N approaches infinity.

An alternative derivation of this result, presented in the Appendix B, shows that the impulse response of the principal oscillator progressively decays and vanishes asymptotically with time. Energy initially imparted to the principal structure migrates to the attached set of infinite number of oscillators that have frequencies that fall within a finite bandwidth, where it remains indefinitely. As discussed earlier, it is commonly accepted that, in general, such near-irreversible

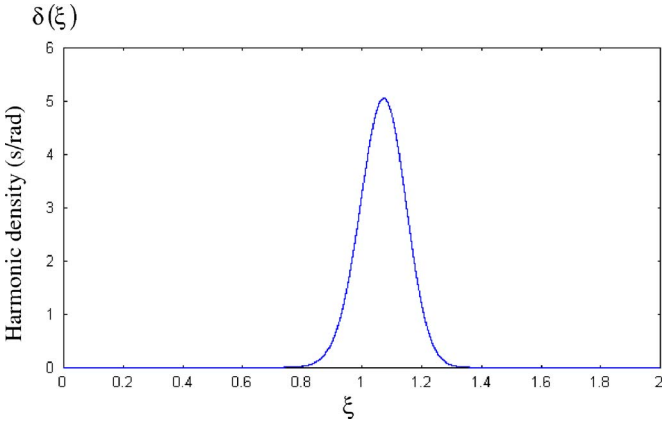


FIG. 14. Modal density for $N=100$, $t_0=\pi/4$, $r=0.1$, $\omega\in[0,2]$.

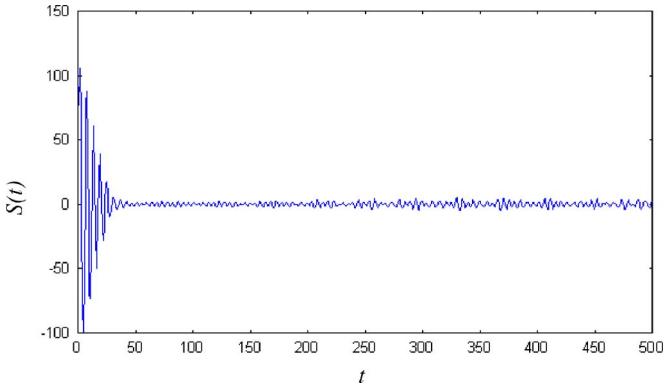


FIG. 15. Time history of the reaction force on the rigid base.

energy transfer does not hold for a finite N.[8] However, as the following application of the theory developed in this paper shows, there exist particular frequency distributions which afford a nearly irreversible energy transfer even for a finite set of oscillators.

Considering the second of Eqs. (26), the displacement of the continuous set of resonators in terms of the master response can be expressed by the convolution integral

$$x(\xi,t) = \omega_n(\xi)\int_0^t x_M(\tau)H(t-\tau)\sin\,\omega_n(\xi)(t-\tau)d\tau,$$

where H is the Heaviside function. Introducing this expression into the first part of Eq. (26), an integro-differential equation results in terms of x_M

$$M\ddot{x}_M(t) + K_M x_M(t) + x_M(t)\int_0^1 k(\xi)d\xi$$

$$-\int_0^t x_M(\tau)\int_0^1 m\omega_n^3(\xi)H(t-\tau)\sin\,\omega_n(\xi)(t-\tau)d\xi d\tau = 0,$$

which can be also expressed as

$$M\ddot{x}_M(t) + (K_M + \bar{k})x_M(t) - x_M(t) * [I(t)H(t)] = 0, \qquad (28)$$

where $I(t)H(t)$ is the kernel of the integral part of the previous equation and

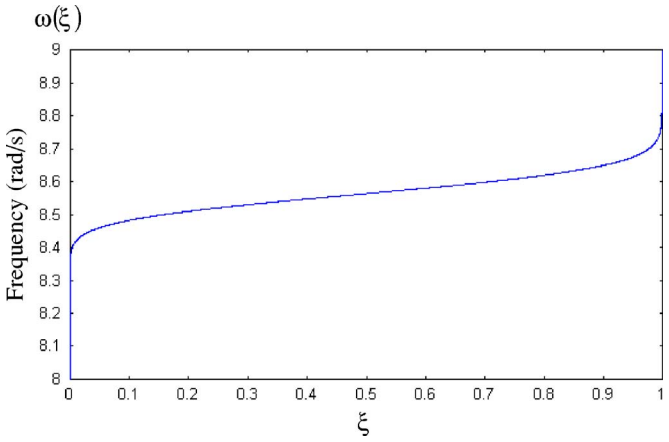


FIG. 16. Natural frequency distribution of the resonator within the set for $N=200$, $t_0=\pi/4$, $r=0.005$, $\omega\in[8,9]$.

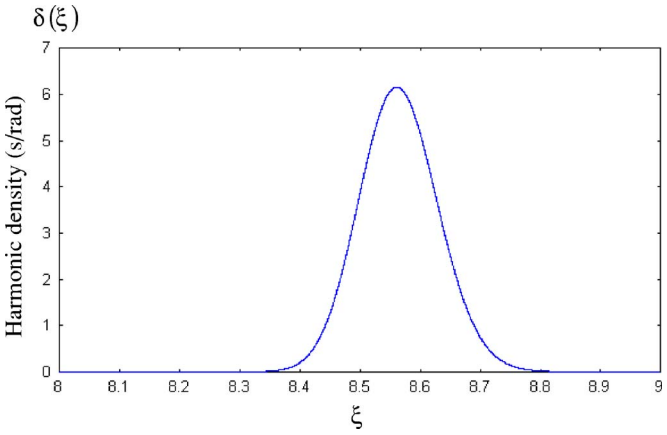


FIG. 17. Modal density for $N=200$, $t_0=\pi/4$, $r=0.005$, $\omega\in[8,9]$.

$$\bar{k} = \int_0^1 k(\xi)d\xi, \quad I(t) = \int_0^1 m\omega_n^3(\xi)\sin\,\omega_n(\xi)td\xi. \qquad (29)$$

In the case of a finite set of N resonators, the equation of motion takes a different form where integrals over ξ are substituted by summations. Thus, Eq. (28) remains applicable provided that $\bar{k}=\sum_{i=1}^{N} k_i$ and $I(t)$ is replaced by its discrete counterpart $S(t)=1/N\sum_{i=1}^{N} m\omega_i^3 \sin\,\omega_i t$

$$M\ddot{x}_M(t) + (K_M + \bar{k})x_M(t) - x_M(t) * [S(t)H(t)] = 0. \qquad (30)$$

The apparent damping and near irreversibility as manifested by the decay characteristics of the impulse response result from the application of the present theory by considering $G(\omega)=m\omega^3$ with $\sigma=m\omega^3\sin\,\omega t_0$ and searching for the optimum frequency distribution.

As an example, consider a master structure, with an uncoupled natural frequency $\omega_M=1$, with $N=100$ attached oscillators. Assuming $t_0=\pi/4$ and searching for a monotonic frequency distribution $\omega(\xi)$, it follows that, within the frequency interval $\omega\in[0,2]$, $d\sigma/d\omega>0$ and $\sigma\in[0,8]$. The values of r and I_0 ($r=0.4$ and $I_0=0.6$) are selected to be consistent with inequalities (24), and to assure that the function represented by Eq. (20) has its peak around $\omega=\omega_M=1$.

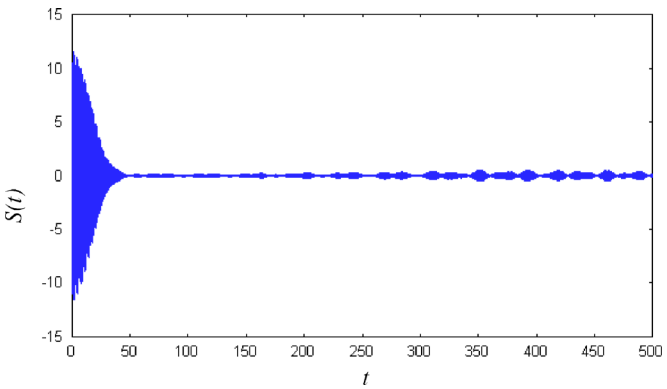


FIG. 18. Impulse response of the one-dimensional structure.

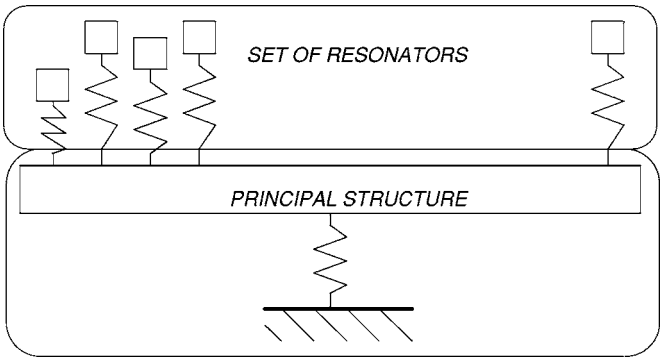


FIG. 19. Schematic illustration of the system.

Figures 20 and 21 show the frequency distribution and the frequency density of the attached oscillators determined by solving Eqs. (21) and (23). Figure 22 shows the master response following an impulse applied at $t=0$, which illustrates how a significant part of its energy is transferred to the set of oscillators and remains there without returning back to the master, producing a near-irreversible energy transfer.

IX. DISCUSSION ON A PROBABILISTIC INTERPRETATION OF THE THEORY

The theory presented above refers to a set of oscillators whose distribution of natural frequencies is deterministically given by $\omega(\xi)$. However, these results also apply to the case of resonators with randomly distributed natural frequencies and can be interpreted in the context of the Theory of Estimators.

Consider $\sigma=G(\omega)\sin \omega t$ as a function of the random variable ω with a probability density function p_ω. Expected value, $E\{\sigma\}$ of σ:

$$E\{\sigma\} = \int_{-\infty}^{+\infty} p_\omega(\omega)G(\omega)\sin \omega t d\omega. \tag{31}$$

In terms of the probability density function of σ, p_σ, an equivalent expression for its expectation has the form

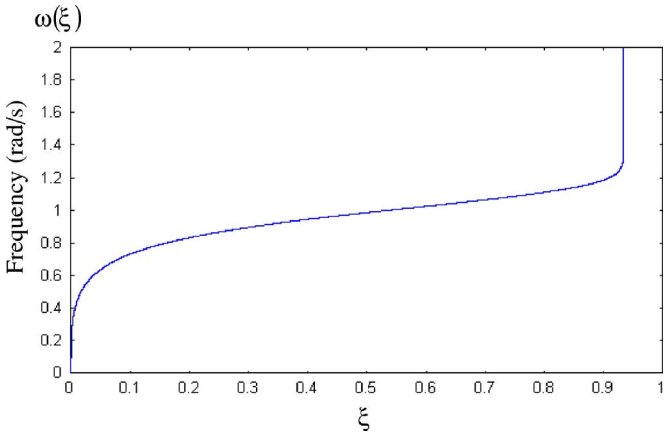


FIG. 20. Frequency distribution within the set of oscillator.

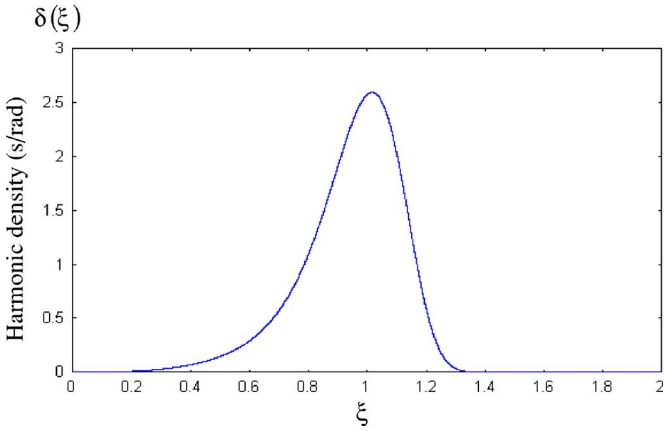


FIG. 21. Natural frequency density within the set of oscillators.

$$E\{\sigma\} = \int_{-\infty}^{+\infty} p_\sigma(\sigma)\sigma d\sigma.$$

From the well known condition $p_\sigma(\sigma)d\sigma=p_\omega(\omega)d\omega$

$$p_\sigma(\sigma)\frac{d\sigma}{d\omega} = p_\omega(\omega). \tag{32}$$

An asymptotic expansion of the integral (31) shows it obeys the asymptotic property

$$\lim_{t\to\infty} E\{\sigma\} = 0. \tag{33}$$

Considering N independent samples of σ, obtained from N random samples of the variable ω, and expressed as $s_i = G(\omega_i)\sin \omega_i t$, $i=1,2,\ldots,N$, the expected value of σ, $E\{\sigma\}$, can be estimated as

$$S = \frac{1}{N}\sum_{i=1}^{N} s_i. \tag{34}$$

The estimator S given by Eq. (34) represents an approximation of $E\{\sigma\}$ given by Eq. (30) and improves for larger values of N.

Having the estimator S to best represent $E\{\sigma\}$ in a statistical sense, makes $E\{\sigma\}$ analogous to $I(t)$. In addition, if the conditional probability $p_\sigma(\sigma|E\{\sigma\})$ replaces the weighting function $p(\sigma,I)$ in the previous analysis, $P(\mathbf{s}|E\{\sigma\})$ defined by equation (4) becomes the *likelihood function* asso-

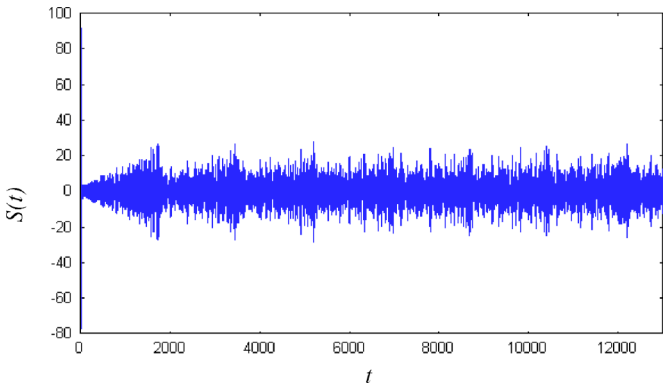


FIG. 22. Master response subjected to an initial impulse.

ciated with the estimator S.[18] Then, in this case, Eq. (9) provides the expected value of the general function $\boldsymbol{f}$ depending on the set of independent random samples $s_1, s_2, \ldots, s_N$. Finally, the inequality (18) provides the minimum variance bound theorem, known as the Cramer-Rao inequality.[18] The minimum variance $\overline{(S-E\{\sigma\})^2}$ is obtained for that class of probability density functions $p_\sigma(\sigma|E\{\sigma\})$,which form the solutions to Eq. (19), developed by Pitman and Koopman. The Gauss distribution given by Eq. (20) is an example of functions that belong to this class, which in statistical terms, exhibits *sufficient statistics*.

X. CONCLUDING REMARKS

The general expression for the response of a linear conservative system, in terms of superposition of pure harmonic functions, suggests an almost periodic time response. Such an expression implies that the impulse response of a finite conservative system cannot possess a decaying time history, a property commonly associated with dissipative systems. This study offers a theory that provides the foundation and explanations for the earlier studies which have shown the existence of a class of conservative linear systems that exhibit apparent damping characteristics and nearly irreversible energy exchange.

Although still open, the question of irreversibility and thermalization of energy has been widely investigated for nonlinear systems, showing nonlinearity as the main cause forcing a Hamiltonian system to thermalize its energy through transition to chaos. There is a paucity of analogous investigations that consider the possibility of irreversibility in linear Hamiltonian systems, as those examined in this paper, except studies that consider the presence of fractional integrals in the equation of motion.

APPENDIX: MINIMIZATION OF DISTANCE USING A VARIATIONAL APPROACH

The problem of minimization of the functional D^2 can be treated also using a variational approach. In this case, the method of the Lagrange's multipliers must be necessarily used to take into account the additional restraint condition $\overline{(S-I)(\partial/\partial I \log P)}=1$ introduced in Sec. IV.

Thus, P must produce the minimum of the modified functional:

$$\tilde{D}^2 = \overline{(S-I)^2} + \lambda\overline{(S-I)\left(\frac{\partial}{\partial I}\log P\right)},$$

where λ is the Lagrange multipler. The variation of $\tilde{D}^2$ with respect to I (or equivalently with respect to $\omega(\xi)$) leads to the Euler-Lagrange equation

$$-2(S-I)P + (S-I)^2\frac{\partial P}{\partial I} - \lambda P\frac{\partial}{\partial I}\log P + \lambda(S - I)P\frac{\partial^2}{\partial I^2}\log P + \lambda(S-I)\frac{\partial}{\partial I}\log P\frac{\partial P}{\partial I} = 0.$$

Dividing for $P(S-I)$ and rearranging it produces

$$-2 + (S-I)\frac{\partial}{\partial I}\log P - \frac{\lambda}{(S-I)}\frac{\partial}{\partial I}\log P + \lambda\frac{\partial}{\partial I}\left(\frac{\partial}{\partial I}\log P\right) + \lambda\left(\frac{\partial}{\partial I}\log P\right)^2 = 0 \quad \text{(A1)}$$

This differential equation in terms of P admits a solution that matches the one determined by the Schwartz inequality. In fact, as it can be easily verified, substitution for $\partial/\partial I \log P = -(S-I)/\lambda$ in Eq. (A1) produces an identity.

APPENDIX B: EQUIVALENT DAMPING OF A CONTINUOUS DISTRIBUTION OF OSCILLATORS

The Fourier transform of the impulse response $\hat{I}(t)$ is expressed as

$$\mathfrak{I}\{\hat{I}\} = \mathfrak{I}\{I\} * \mathfrak{I}\{H\}, \quad \text{(B1)}$$

Where H represents Heaviside function. The first in the convolution is

$$\mathfrak{I}\{I\} = \int_0^1 m\omega^3(\xi)\int_{-\infty}^{+\infty} e^{-j\Omega t}\sin\omega(\xi)t\,dt\,d\xi = -j\frac{\pi}{2}\int_0^1 m\omega^3(\xi)[\delta(\Omega+\omega(\xi)) + \delta(\Omega-\omega(\xi))]d\xi.$$

Replacing $d\omega = \omega'(\xi)d\xi$ yields

$$\mathfrak{I}\{I\} = -j\frac{\pi}{2}\int_0^1 \frac{m\omega^3}{\omega'}[\delta(\Omega+\omega) + \delta(\Omega-\omega)]d\omega.$$

Writing the distribution $\omega(\xi)$ as the solution of an equation $\omega' = f(\omega)$, for an arbitrary function f, the previous integral produces the expression

$$\mathfrak{I}\{I\} = -j\Omega\left[m\frac{\pi}{2}\frac{\Omega^2}{f(\Omega)}\right] \quad \text{for } \Omega \in [\omega(0);\omega(1)]$$

$$\mathfrak{I}\{I\} = 0 \quad \text{elsewhere.}$$

The second term in the convolution Eq. (B1) is

$$\mathfrak{I}\{H\} = \frac{1}{2}\delta(\Omega) + \frac{1}{j\Omega}.$$

Therefore

$$\mathfrak{I}\{\hat{I}\} = -j\Omega\left[m\frac{\pi}{4}\frac{\Omega^2}{f(\Omega)}\right] + \int_{-\infty}^{+\infty} m\frac{\pi}{2}\frac{\zeta^3}{f(\zeta)(\zeta-\Omega)}d\zeta. \quad \text{(B2)}$$

This expression provides the frequency domain counterpart of the term $-x_M(t)*[I(t)H(t)]$ in Eq. (26), as $-X_M\cdot\mathfrak{I}\{\hat{I}\}$. The imaginary part in $\mathfrak{I}\{\hat{I}\}$, i.e., $\Omega[m\pi/4\Omega^2/f(\Omega)] = \Omega C(\Omega)$, suggests that the set of oscillators with a continuous frequency distribution introduces a frequency dependent viscous damping $C(\Omega)$ on the master motion.

The real part corresponds to the reactive part of the impedance, which, for small m, is generally not very important compared with the inertial and stiffness effects intrinsically related to the master oscillator.

[1] A. D. Pierce, V. W. Sparrow, and D. A. Russel, "Fundamental structural-acoustic idealization for structure with fuzzy internals," J. Vibr. Acoust.

117, 339–348 (1995).
[2]M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344 (1996).
[3]G. Maidanik, "Induced damping by a nearly continuous distribution of nearly undamped oscillators: Linear analysis," J. Sound Vib. **240**, 717–731 (2001).
[4]R. L. Weaver, "The effect of an undamped finite degree of freedom 'fuzzy' substructure: Numerical solution and theoretical discussion," J. Acoust. Soc. Am. **101**, 3159–3164 (1996).
[5]R. J. Nagem, I. Veljkovic, and G. Sandri, "Vibration damping by a continuous distribution of undamped oscillators," J. Sound Vib. **207**, 429–434 (1997).
[6]C. E. Celik and A. Akay, "Dissipation in solids: Thermal oscillations of atoms," J. Acoust. Soc. Am. **108**, 184–191 (2000).
[7]R. L. Weaver, "Equipartition and mean square response in large undamped structures," J. Acoust. Soc. Am. **110**, 894–903 (2001).
[8]A. Carcaterra and A. Akay, "Transient energy exchange between a primary structure and a set of oscillators: Return time and apparent damping," J. Acoust. Soc. Am. **115**, 683–696 (2004).
[9]I. Murat Koç, A. Carcaterra, Zhaoshun Xu, and A. Akay, "Energy sinks: Vibration absorption by an optimal set of undamped oscillators," to appear, J. Acoust. Soc. Am. **118**(5), 3031-3042 (2005).
[10]A. Carcaterra, A. Akay, and I. M. Koc, "Near-Irreversibility and damped response of a conservative linear structure with singularity points in its modal density," J. Acoust. Soc. Am. **119**(4), 2141-2149 (2006).
[11]A. Akay, Z. Xu, A. Carcaterra, and I. Murat Koç, "Experiments on vibration absorption using energy sinks," J. Acoust. Soc. Am. **118**(5), 3043-3049 (2005).
[12]A. Carcaterra, "An entropy formulation for the analysis of energy flow between mechanical resonators," Mech. Syst. Signal Process. **16**(5), 905–920 (2002).
[13]A. Carcaterra, "Ensemble energy average and energy flow relationships for nonstationary vibrating systems," J. Sound Vib. Special Issue, Uncertainty in Structural Dynamics, **288**(3), 751–790 (2005).
[14]T. Y. Petrosky, "Chaos and irreversibility in a conservative nonlinear dynamical system with a few degrees of freedom," Phys. Rev. A **29**(4), 2078–2091 (1984).
[15]P. K. Datta and K. Kundu, "Energy transport in one-dimensional harmonic chains," Phys. Rev. B **51**(10), 6287–6295 (1995).
[16]R. Livi, M. Pettini, S. Ruffo, M. Sparpaglione, and A. Vulpiani, "Equipartition threshold in nonlinear large Hamiltonian systems: The Fermi-Pasta-Ulam model," Phys. Rev. A **31**(2), 1039–1045 (1985).
[17]R. R. Nigmatullin and A. Le Mehaute, "To the nature of irreversibility in linear systems," Magn. Reson. Solids **6**(1), 165–179 (2004).
[18]M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Charles Griffin Limited, London, 1961).
[19]E. J. G. Pitman, "Sufficient statistics and intrinsic accuracy," Proc. Cambridge Philos. Soc. **32**, 576 (1936).
[20]B. O. Koopman, "On distributions admitting a sufficient statistics," Trans. Am. Math. Soc. **39**, 399 (1936).

On the existence of localized shear horizontal acoustic waves in a piezoelectric plate with two semi-infinite same/different coatings

Shi Chen,[a] Tiantong Tang, and Zhaohong Wang
Xi'an Jiaotong University, Xi'an 710049, China

(Received 21 May 2006; revised 22 October 2006; accepted 11 January 2007)

In this paper, the existence theorem of localized shear horizontal acoustic waves in a piezoelectric plate with two semi-infinite same/different coatings is established. Some properties of the waves in the waveguide structures are also discussed. The results show that the waveguides have some advantages and provide more choice for the designs of acoustic devices. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2641812]

PACS number(s): 43.40.Cw, 43.20.Mv, 43.20.Bi [PEB] Pages: 1983–1986

I. INTRODUCTION

There are many ways to confine acoustic waves in the appointed regions in a solid, among them three commonly used and extensively studied ways are: surface acoustic wave waveguides (SAWGs),[1–3] interfacial acoustic wave waveguides (IAWGs),[4] and plates or rods.[1,5] Some drawbacks exist in the acoustic devices based on the acoustic waveguides. For examples, the devices, based on SAWGs, plates, and rods, must be enveloped, since acoustic waves in the waveguides are easily influenced by the outer surroundings. Thus the scale of the devices becomes very large. When the frequency of acoustic waves tends to extreme large, the penetration depth of the surface or interfacial acoustic waves (SAW/IAW) becomes very small. Thus the devices consisting of the waveguides become ineffective at the extremely high frequency regime. Furthermore the existence conditions of the IAWs are very rigorous, which leads to the materials used to form the IAWGs are limited.

In this paper, an acoustic waveguide structure consisting of a piezoelectric plate coated by two semi-infinite same/different materials is studied. It is detected that all the drawbacks mentioned previously can be overcome. Furthermore the structure may have some other advantages. For an example, optical waves may be guided in the same structure, so some acousto-optic interaction devices may be fabricated easily based on the structure.

The focus of this paper is to investigate the existence theorem of localized shear horizontal (SH) acoustic waves in the structure mentioned earlier. On the existence theorem, only some simple structures, such as SAWs[6–8] on a semi-infinite media and IAWs[9,10] on the interface between two semi-infinite medias, have been studied comprehensively. It is very difficult or may be impossible nowadays to establish the existence theorem for complex structures thoroughly and generally. So only materials with special crystal symmetry are considered here. The materials in class *6mm* are selected for the plate, and the materials in class *6mm* or some nonpiezoelectrics are selected as the coatings. It is also noticeable that similar structures have been investigated in Ref. 11 with different focus, which emphasizes the meaning of studying the structure.

II. SH IAWs ON THE INTERFACE BETWEEN TWO SEMI-INFINITE PIEZOELECTRIC CRYSTALS IN CLASS *6MM*

For the sake of convenience later we discuss localized SH acoustic waves in the piezoelectric plate with two semi-infinite coatings. Some known properties of SH IAWs are described simply in this section. The knowledge can be found in Ref. 4.

Let A and B be two semi-infinite piezoelectric crystals (in class 6*mm* and whose *C* axes are parallel to the x_3 direction) in contact, the plane $x_2=0$ being the interface. There may exist a SH interfacial wave on the interface, which is described by the mechanical displacement u_3 and the electrical potential ϕ (on the quasistatic assumption). The wave propagates in the x_1 direction.

Let $c=c_{44}$, $e=e_{15}$, and $\varepsilon=\varepsilon_{11}$ denote the elastic, piezoelectric, and dielectric tensor components, respectively, and ρ the mass density. For a realizable material, a stability condition[12] must be satisfied, i.e., the energy density must be a positive quantity for arbitrary strains and electric fields. The condition is expressed as follows:

$$\bar{E} = c_{ijkl}s_{ij}s_{kl} + \varepsilon_{ij}E_iE_j > 0, \tag{1}$$

where c_{ijkl}, s_{ij}, and ε_{ij} denote the elastic, strain, and dielectric tensor components, E_i are the components of the electric field vector, and $\bar{E}$ denotes the energy density. The stability condition Eq. (1) ensures that the material does not undergo spontaneous deformation or electric polarization.

For the materials in class 6*mm* or nonpiezoelectric materials, some useful properties can be derived[4,13] from Eq. (1). These to be used are listed as follows:

$$c > 0, \quad \varepsilon > 0, \quad c > e^2/\varepsilon. \tag{2}$$

The existence theorems of the IAWs are listed as follows:

Existence theorem 2.1: If the interface between media

[a]Electronic mail: chenshi@mail.xjtu.edu.cn

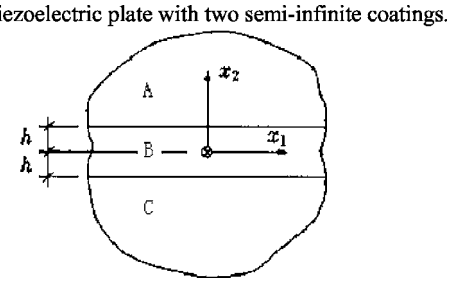


FIG. 1. A piezoelectric plate with two semi-infinite coatings.

A and media B is nonmetallized (open circuit), the existence condition is

$$\bar{C}_A\left(1-\frac{v_B^2}{v_A^2}\right)^{1/2}<\frac{(e_A/\varepsilon_A-e_B/\varepsilon_B)^2}{1/\varepsilon_A+1/\varepsilon_B}, \tag{3}$$

where the subscripts A and B denote the quantities relative to media A and media B, respectively, $\bar{C}=c+e^2/\varepsilon$, $v=(\bar{C}/\rho)^{1/2}$, $v_A>v_B$.

Existence theorem 2.2: If the interface is metallized (short circuit), i.e., an infinitely thin metal layer is inserted in the interface, the existence condition is

$$\bar{C}_A\left(1-\frac{v_B^2}{v_A^2}\right)^{1/2}<\frac{e_A^2}{\varepsilon_A}+\frac{e_B^2}{\varepsilon_B}, \tag{4}$$

where $v_A>v_B$.

Existence theorem 2.3: If media A and media B are identical with codirectional C axes, and the interface is nonmetallized, there are no SH interfacial waves.

Existence theorem 2.4: If media A and media B are identical with codirectional C axes, and the interface is metallized, there is a SH interfacial wave always.

III. LOCALIZED SH ACOUSTIC WAVES IN THE PIEZOELECTRIC PLATE WITH TWO SEMI-INFINITE SAME COATINGS

The structure of a piezoelectric plate with two semi-infinite coatings is shown in Fig. 1. The thickness of the plate is $2h$. The principal axes of media A, B, and C are identical with the coordinate axes. In this section, media A is identical to media C.

The media B is a piezoelectric crystal in class *6mm*, and the media A and media C are piezoelectric crystals in class 6*mm* or nonpiezoelectric crystals. The waves propagate in the x_1 direction.

Assuming that the waves are monochromatic and propagate in the x_1 direction, then all field variables have a common factor $\exp(-j\omega t)\exp(jk_1x_1)$ which is omitted hereafter. Then field equations[14,15] in the system are

$$c\nabla_\tau^2u_3+e\nabla_\tau^2\phi+\rho\omega u_3=0, \tag{5}$$

$$e\nabla_\tau^2u_3-\varepsilon\nabla_\tau^2\phi=0, \tag{6}$$

where $\nabla_\tau^2=\partial_{x_1}^2+\partial_{x_2}^2$.

The general solutions of Eqs. (5) and (6) can be expressed as follows:[14]

$$u_3=F_1\exp(k_2x_2)+F_2\exp(-k_2x_2), \tag{7}$$

$$\phi=\frac{e}{\varepsilon}[F_1\exp(k_2x_2)+F_2\exp(-k_2x_2)]+F_3\exp(k_1x_2)+F_4\exp(-k_1x_2), \tag{8}$$

where $k_2=k_1\sqrt{1-V^2/v^2}$. $V=\omega/k_1$ denotes the phase velocity of the waves. F_1, F_2, F_3, and F_4 are unknown constants. The normal stress is: $T=T_{23}=c(\partial/\partial x_2)u_3+e(\partial/\partial x_2)\phi$, and the normal electric displacement is: $D=D_2=e(\partial/\partial x_2)u_3-\varepsilon(\partial/\partial x_2)\phi$. They can be expressed as follows:

$$T=F_1\bar{C}k_2\exp(k_2x_2)-F_2\bar{C}k_2\exp(-k_2x_2)+F_3ek_1\exp(k_1x_2)-F_4ek_1\exp(-k_1x_2), \tag{9}$$

$$D=-F_3\varepsilon k_1\exp(k_1x_2)+F_4\varepsilon k_1\exp(-k_1x_2). \tag{10}$$

For getting localized modes, the fields in the coatings must be attenuated to zero when $x_2\to\pm\infty$. Since the structure considered here is symmetrical, there are two kinds of the solutions: symmetrical solutions and antisymmetric solutions. For the problem of the existence, the symmetrical solutions are chosen here. The convenience of the choice will be verified by the results derived from it.

The boundary conditions are: the T, D, u_3, and ϕ must be continuous at the interfaces. The dispersion equation for the symmetrical modes can be derived directly,

$$\frac{(e_B/\varepsilon_B-e_A/\varepsilon_A)^2}{1/\varepsilon_A+1/\varepsilon_BG(k_1,h)}k_1=\bar{C}_Ak_2^A+\bar{C}_Bk_2^BH(k_2^B,h), \tag{11}$$

where the subscripts or superscripts A and B denote the quantities relative to media A and media B, respectively. $G(k_1,h)=[\exp(2k_1h)+1]/[\exp(2k_1h)-1]$, $H(k_2^B,h)=[\exp(2k_2^Bh)-1]/[\exp(2k_2^Bh)+1]$.

Some simple existence theorems can be discussed without considering the physical detail of the system.

Existence theorem 3.1: If all the interfaces are open circuit, and there exists an IAW on the interface (which is open circuit) of the structure consisting of a semi-infinite media A and a semi-infinite media B (which is called AB IAW hereafter), then in a frequency interval localized SH waves will exist in the structure in Fig. 1.

When the frequency ω of the AB IAW tends to infinite large, the penetration depth of it will tend to zero. It is obvious that at extremely high frequency, an IAW can exist on the upper interface of the structure in Fig. 1 and is not influenced by the nether interface. Thus a localized SH acoustic wave can exist in the structure in a frequency interval, and the frequency interval has no upper limit where "low frequency cutoff theorem" is used. In other words, if there is a localized wave at ω_1 in a system, then at least a localized wave at $\omega_2(\omega_2>\omega_1)$ must exist in the same system.

At the case, when $\omega\to\infty$, all the partial waves in the plate B are inhomogeneous. So the waveguide structures at the case of the existence theorem 3.1 may be ineffective at extremely high frequency. If $v_B>v_A$, all the partial waves in the plate B are inhomogeneous for all the frequencies.

Existence theorem 3.2: If one of the interfaces in Fig. 1

is short circuit, then for all the frequency localized SH waves will exist in the structure in Fig. 1.

When the frequency tends to zero, the influence of the plate B will be ignored. Thus the existence theorem 2.4 ensures that an IAW will exist in the structure when the frequency tends to zero. From the low frequency cutoff theorem, localized SH acoustic waves will exist in the structure for all frequencies.

At the case of the existence theorem 3.2, if there is no AB IAW for the open circuit and short circuit interfaces, when $\omega \to \infty$, homogeneous partial waves must exist in plate B, which implies that at that case the waveguide structures may be effective at the extremely high frequency.

Existence theorem 3.3a: If all the interfaces in Fig. 1 are open circuit, and there is no AB IAW, the necessary condition of the existence of localized SH acoustic waves in the structure in Fig. 1 is $v_A > v_B$.

At the case of the existence theorem 3.3a, when the frequency tends to infinite large, there must exist partial waves in plate B which are homogeneous in the x_2 direction, or else boundary conditions cannot be satisfied. If $v_A < v_B$, then at extremely high frequency, when homogeneous partial waves exist in plate B, there must also exist homogeneous partial waves in the coating A,[4] thus no localized SH acoustic waves can exist in the structure.

It will be verified that $v_A > v_B$ is also the sufficient condition of the existence of localized waves at the case of the existence theorem 3.3a.

When $\omega \to \infty$ and $v_B < V < v_A$, Eq. (11) can be transformed into the following form:

$$\frac{(e_B/\varepsilon_B - e_A/\varepsilon_A)^2}{1/\varepsilon_A + 1/\varepsilon_B} - \bar{C}_A(1 - V^2/v_A^2)^{1/2} = -\bar{C}_B(V^2/v_B^2 - 1)^{1/2}\tan[(V^2/v_B^2 - 1)^{1/2}hk_1], \quad (12)$$

where the symbol tan( ) denotes the tangent function. In the velocity interval $v_B < V < v_A$, the value of the left-hand side of Eq. (12) changes in a finite range. That of the right-hand side of Eq. (12) changes in $(-\infty, \infty)$. It is obvious that localized waves must exist. Now the existence theorem 3.3a can be expressed as follows:

Existence theorem 3.3b: If all the interfaces in Fig. 1 are open circuit, and there is no AB IAW, the necessary and sufficient condition of the existence of localized SH acoustic waves in the structure in Fig. 1 is $v_A > v_B$.

Existence theorem 3.3b verifies that the choice of the symmetrical solutions is convenient. Some properties of the localized SH waves at the case of the existence theorem 3.3b will be discussed in the following.

When $0 < V \le v_B$, Eq. (11) is transformed into the following form:

$$\frac{(e_B/\varepsilon_B - e_A/\varepsilon_A)^2}{1/\varepsilon_A + 1/\varepsilon_B G(k_1,h)} = \bar{C}_A(1 - V^2/v_A^2)^{1/2} + \bar{C}_B(1 - V^2/v_B^2)^{1/2}H(k_2^B,h). \quad (13)$$

Because there is no AB IAW, so Eq. (3) must not be satisfied. Because $G(k_1,h) > 1$, $H(k_1,h) > 0$, the value of the left-hand side of Eq. (13) is always less than that of the right-hand side of Eq. (13). It is deduced that there are no localized SH acoustic waves when $0 < V \le v_B$.

The following conclusions can be summarized: At the case of the existence theorem 3.3b, the velocities of the localized waves are larger than v_B and less than v_A, which implies that homogeneous partial waves must exist in plate B. So when the frequency tends to infinite large, the waveguide structures are also effective. Furthermore it is also observed that the existence condition of the localized waves is weaker than that of the IAWs.

IV. LOCALIZED SH ACOUSTIC WAVES IN THE PIEZOELECTRIC PLATE WITH TWO SEMI-INFINITE DIFFERENT COATINGS

The structure considered here is shown in Fig. 1, and media A is different from media C here. Some simple existence theorems can also be discussed without considering the physical details of the system.

Existence theorem 4.1: If all the interfaces are open circuit, and there exists an AB IAW or BC IAW, then in a frequency interval localized SH waves will exist in the structure in Fig. 1.

Existence theorem 4.2: If there exist an AC IAW (which is the IAW on the interface of the structure consisting of a semi-infinite media A and a semi-infinite media C), then for all the frequency localized SH waves will exist in the structure in Fig. 1.

Existence theorem 4.3a: If all the interfaces are open circuit, and the AB IAW, the BC IAW, and the AC IAW do not exist, then the necessary conditions of the existence of localized SH acoustic waves are $v_B < v_A$ and $v_B < v_C$.

The existence theorems can be verified by similar deductions to the ones used in Sec. III. So it is omitted here.

It will be verified that $v_B < v_A$ and $v_B < v_C$ are also the sufficient conditions of the existence of localized waves at the case of the existence theorem 4.3a. In order to do it, the dispersion equation of the system must be gotten.

The boundary conditions are: The fields in the coatings must be attenuated to zero when $x_2 \to \pm\infty$. The T, D, u_3, and ϕ must be continuous at the interfaces.

The derivation of the dispersion equation is tedious and direct, and the form of the equation is complex. But for the problem of the existence of localized waves, only a limit case needs be considered necessarily (i.e., $\omega \to \infty$). It is fortunate that the dispersion equation at the limit case can be derived easily with the symmetrical and antisymmetric base function method.[15] The process of the derivation is omitted here, only the results are presented.

When $\omega \to \infty$ and $v_B < V < \min(v_A, v_C)$, the dispersion equation of the system is

$$(M_A + M_C)\bar{C}_B\sqrt{V^2/v_B^2 - 1}[\tan(\Delta) - \tan^{-1}(\Delta)] = M_A M_C - \bar{C}_B^2(V^2/v_B^2 - 1), \quad (14)$$

where

$$\Delta = k_1 h\sqrt{V^2/v_B^2 - 1}, \quad M_A = \bar{C}_A\sqrt{1 - V^2/v_A^2} - (e_A/\varepsilon_A - e_B/\varepsilon_B)^2/(1/\varepsilon_A + 1/\varepsilon_B),$$

$$M_C = \bar{C}_C\sqrt{1 - V^2/v_C^2} - (e_C/\varepsilon_C - e_B/\varepsilon_B)^2/(1/\varepsilon_C + 1/\varepsilon_B).$$

In the velocity interval $v_B < V < \min(v_A, v_C)$, the value of the right-hand side of Eq. (14) changes in a finite range. That of the left-hand side of Eq. (14) changes in $(-\infty, \infty)$. It is obvious that localized waves must exist. Now the existence theorem 4.3a can be expressed as follows:

Existence theorem 4.3b: If all the interfaces are open circuit, and the AB IAW, the BC IAW, and the AC IAW do not exist, then the necessary and sufficient conditions for the existence of localized SH acoustic waves are $v_B < v_A$and $v_B < v_C$.

Some properties of the localized waves at the case of the existence theorem 4.3b will be discussed simply in the following.

When $\omega \to \infty$, there must exist homogeneous partial waves in plate B. So the waveguide structures are effective at extremely high frequency. It is also observed that the existence conditions of the localized acoustic waves are weaker than that of the IAWs.

V. CONCLUSIONS

In this paper, according to the low frequency cutoff theorem, the existence theorem of localized SH acoustic waves in a piezoelectric thin plate with two semi-infinite same/different coatings is established. The existence conditions and some properties of the localized waves in the structures are also discussed. The results show that the waveguide structures have some advantages: (1) The encapsulation is not necessary since the localized waves are not influenced by the outer surroundings. (2) The waveguides are effective when the frequency tends to extreme large. (3) The waveguides bring more choice for the designs of acoustic devices since the existence conditions of the localized waves in the waveguides are weaker compared with that of the IAWs. (4) Some acousto-optic interaction devices may be fabricated easily based on the structure.

[1] B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Malbar, FL, 1990), Vol. **1**.

[2] Lord Rayleigh, "On waves propagated along the plane surface of elastic solid," Proc. London Math. Soc. 17, 4–11 (1885).

[3] T. Sato and H. Abe, "Propagation properties of longitudinal leaky surface waves on lithium teraborate," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 136–151 (1998).

[4] C. Maerfeld and P. Tournois, "Pure shear elastic surface wave guided by the interface of two semi-infinite medias," Appl. Phys. Lett. **19**, 117–118 (1971).

[5] J. L. Bleustein, "Some simple modes of wave propagation in an infinite piezoelectric plates," J. Acoust. Soc. Am. **45**, 614–620 (1969).

[6] J. Lothe and D. Barnett, "Integral formalism for surface waves in piezoelectric crystals. Existence considerations," J. Appl. Phys. **47**, 1799–1807 (1976).

[7] J. Lothe and D. M. Barnett, "Further development of the theory for surface waves in piezoelectric crystals," Phys. Norv. **8**, 239–254 (1976).

[8] J. Lothe and D. M. Barnett, "On the existence of surface-wave solutions for anisotropic elastic half-space with free surface," J. Appl. Phys. **47**, 428–433 (1976).

[9] A. N. Darinskii and V. N. Lyubimov, "Shear interfacial waves in piezoelectrics," J. Acoust. Soc. Am. **106**, 3296–3304 (1999).

[10] A. N. Darinskii and M. Weihnacht, "Interface acoustic waves in piezoelectric bi-crystalline structures of specific types," Proc. R. Soc. **461**, 995–911 (2005).

[11] A. J. Niklasson, S. K. Datta, and M. L. Dunn, "On ultrasonic guided waves in a thin anisotropic layer lying between two isotropic layers," J. Acoust. Soc. Am. **108**, 2005–2011 (2000).

[12] R. Peach, "On the existence of surface acoustic waves on piezoelectric substrates," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1308–1320 (2001).

[13] F. Milstein, "Mechanical stability of crystal lattices with two-body interactions," Phys. Rev. B 512–518 (1970).

[14] Q. Wang and V. K. Varadan, "Wave propagation in piezoelectric bounded plates by use of interdigital transducer. Dispersion characteristics," Int. J. Solids Struct. **39**, 1119–1130 (2002).

[15] S. Chen, T. T. Tang, and Z. H. Wang, "Shear-horizontal acoustic wave propagation in piezoelectric bounded plates with metal gratings," J. Acoust. Soc. Am. **117**, 3069–3615 (2005).

Using cross correlations of turbulent flow-induced ambient vibrations to estimate the structural impulse response. Application to structural health monitoring[a)]

Karim G. Sabra[b)]
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Eric S. Winkel, Dwayne A. Bourgoyne, Brian R. Elbing, Steve L. Ceccio, Marc Perlin, and David R. Dowling
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109

(Received 26 September 2006; revised 30 January 2007; accepted 30 January 2007)

It has been demonstrated theoretically and experimentally that an estimate of the impulse response (or Green's function) between two receivers can be obtained from the cross correlation of diffuse wave fields at these two receivers in various environments and frequency ranges: ultrasonics, civil engineering, underwater acoustics, and seismology. This result provides a means for structural monitoring using ambient structure-borne noise only, without the use of active sources. This paper presents experimental results obtained from flow-induced random vibration data recorded by pairs of accelerometers mounted within a flat plate or hydrofoil in the test section of the U.S. Navy's William B. Morgan Large Cavitation Channel. The experiments were conducted at high Reynolds number ($\mathrm{Re} > 50$ million) with the primary excitation source being turbulent boundary layer pressure fluctuations on the upper and lower surfaces of the plate or foil. Identical deterministic time signatures emerge from the noise cross-correlation function computed via robust and simple processing of noise measured on different days by a pair of passive sensors. These time signatures are used to determine and/or monitor the structural response of the test models from a few hundred to a few thousand Hertz. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2710463]

PACS number(s): 43.40.Le, 43.40.Sk, 43.40.Hb [WMC] Pages: 1987–1995

I. INTRODUCTION

Structural health monitoring (SHM) of large mechanical structures is of key importance for maintenance and safety in aeronautics (e.g., aircraft wings, control surfaces, propulsion system components, etc.), civil engineering (e.g. bridges, buildings, offshore structures), and transport of commodities (e.g., pipelines). By measuring the vibration properties of those structures, such as resonant frequencies, mode shapes, and damping coefficients, tools for prognosis and diagnosis of vibratory characteristics can be developed to ensure long term maintenance of structural integrity. In practice the impulse response obtained by linearly relating a measured input (e.g., force) to a measured response (e.g., acceleration) is used to characterize structures. Typical commercial systems available today use controlled active sources (e.g., hammer, vibroseis truck, or laser).[1] Two drawbacks of these techniques are: (i) the required deployment of an active source, and (ii) the need for direct access to the inspected area. Such access often requires some disassembly of the structure and prevents its continuous use for its intended purpose during testing.

Ambient noise and passive monitoring provide a nonintuitive approach to SHM. Random and randomly scattered seismic or acoustic noise fields are often considered to be incoherent and of limited utility, but there is some coherence between two sensors that receive sounds or vibrations from the same source or scatterer. Experimental and theoretical analyses have shown that the arrival-time structure of the impulse response between a sensor pair can be estimated from the noise cross-correlation function where, in this case, the noise is ambient vibration. This method was investigated in various environments and frequency ranges: ultrasonics,[2–4] civil engineering[5–10] (also referred to as the natural excitation technique[11]), underwater acoustics,[12–14] seismology,[15–17] and helioseismology.[18] These results, when coupled with tomography, provide a means for passive imaging using only the ambient vibration field, without the use of active sources. The physical process underlying this noise cross-correlation technique is similar for all these environments. Initially, the coherent component of the noise field at each receiver is buried in the spatially and temporally incoherent field. However, after sufficient temporal averaging over noise events, the coherent impulse response arrival structure emerges in the noise cross-correlation function wave form. The correlation process accumulates contributions over time from noise components that propagate through both receiving sensors, along the impulse response paths.

The purpose of this paper is to document the SHM potential of this ambient-vibration cross-correlation method by reporting experimental results from structures subject to con-

[a)]Part of this work was presented at the 151th meeting of the Acoustical Society of American in Providence, RI.

[b)]Author to whom correspondence should be addressed. Electronic mail: ksabra@mpl.ucsd.edu

tinuous high-Reynolds-number turbulent-flow excitation. Here, two structures were investigated: a rigid steel skin-on-frame flat plate,[19] and a solid Ni–Al bronze hydrofoil.[20] Vibration measurements were made using pairs of accelerometers mounted inside each structure while it was undergoing hydrodynamic testing in the William B. Morgan Large Cavitation Channel (LCC).[21] The ambient-vibration cross-correlation method is investigated by: 1) comparing the computed noise cross-correlation function to the measured impulse response of the plate using an active point source, and 2) measuring the variations of the noise cross-correlation function due to changes in the mounting conditions of the hydrofoil before and after large-amplitude load fluctuations caused by unsteady sheet cavitation. Once the impulse response is known, standard mathematical global SHM methods can be used to extract a structure's modal parameters even when such modes are closely spaced.[5,6,22] For instance, the use of these modal-parameter-estimation methods, such as the eigensystem realization algorithm,[23] is well documented[5] and thus will not be the focus of this present article.

The remainder of this article is divided into five Sections. Section II is a summary of the theoretical basis for the ambient-vibration cross-correlation method. Section III presents experimental comparisons between the active-source-measured and noise cross-correlation function-determined impulse response of the test plate. The influence of the turbulent flow conditions on the emergence rate of the noise cross-correlation function is discussed also. Section IV is an illustration of this method for SHM of the hydrofoil model's mounting conditions. Section V summarizes the findings and conclusions drawn from this study.

II. ESTIMATING THE IMPULSE RESPONSE FROM CROSS CORRELATION OF AMBIENT VIBRATIONS: THEORETICAL BASIS

The theoretical relationship between the impulse response and the noise cross-correlation function have been presented previously (see references of Sec. I) so only a terse summary of the main assumptions and significant results is provided here. In a stationary medium, the expected value of the temporal noise cross-correlation function $\langle C_{1,2}(t)\rangle$ between the noise signals, $S_1(t)$ and $S_2(t)$, recorded by receivers 1 and 2 on the time interval $[-T_r/2, +T_r/2]$ is

$$\langle C_{1,2}(t)\rangle = \frac{1}{T_r}\int_{-T_r/2}^{+T_r/2} S_1(\tau)S_2(\tau+t)d\tau. \tag{1}$$

Although defined here in terms of a single temporal integration, the noise cross-correlation function may also be constructed from an ensemble average of shorter duration time averages as well. In addition, for the work reported here, the time shift, τ, is typically much smaller than the recording interval, T_r, and both negative and positive time delays are considered.[24] In practice, the nature of the two recorded signals $S_1(t)$ and $S_2(t)$ (e.g., forces, accelerations) depends on the type of the sensors used and this determines which impulse response will be recovered experimentally by cross correlations of the two signals. In this investigation, two accelerometers are used; thus $\langle C_{1,2}(\tau)\rangle$ will have units of acceleration squared.

The next step is to expand the expression for S_1 and S_2 in Eq. (1) as a sum (or infinite integral) overall all noise source contributions based on the principle of superposition.[25] Two key ingredients are needed: 1) *diffuse* ambient vibrations or noise field and 2) reciprocity or a modal orthogonality relationship for the true time-domain impulse response between the two receivers, denoted $G_{1,2}(t)$. The need for these two ingredients is explained in the next two paragraphs.

First, a fully diffuse ambient noise field is defined either as an incoherent superposition of plane waves of all directions and phases, or the field produced by an isotropic distribution of random sources, with spatial-delta-correlated amplitudes, in a homogeneous medium for acoustic[13,24] or elastic waves;[26,27] or more generally one with equipartitioned uncorrelated normal mode amplitudes in the case of heterogeneous or anisotropic or multi-mode environments (e.g., waveguides,[15,28] dynamic structures[29,30]). This assumption ensures the uniform spatial and temporal distribution of the noise sources so that all propagation paths between the two passive sensors are illuminated fully by the noise events along those paths. Hence the diffuse, noise field results from the spatial and temporal average over all noise sources.

Second, a reciprocity theorem, also referred to as the Maxwell-Betti or Rayleigh reciprocity relationship or the Ward identity, is typically used to relate two independent elastodynamic states (recorded wave fields and noise sources here) in a given linear system. Overall either these reciprocity theorems[29,15,30] or the orthogonality relationship for the normal modes of the medium[28,29,27] or the use of the stationary phase approximation,[26,13,24] are then applied to Eq. (1). These relationships are used to reduce the number of integrals and terms in the expansion of Eq. (1) and also to simplify cross terms in the impulse response, in particular for the spatial integration over the noise source distribution.

The formal relationship between the Fourier transform of the impulse response, $\widetilde{G_{1,2}}(\omega)$, and the expected value of the Fourier transform of the noise cross-correlation function, $\widetilde{C_{1,2}}(\omega)$, between the two sensors can be stated as[15,27,29,30]

$$\langle\widetilde{C_{1,2}}(\omega)\rangle = i\beta(\widetilde{G_{1,2}}(\omega) - \widetilde{G_{2,1}^{*}}(\omega)), \tag{2}$$

where the symbol * stands for complex conjugation. The constant β is related to the power spectrum of the noise excitation, i.e., the noise source strength or modal density of the structure, and the nature of the signals S_1 and S_2, and the propagating medium. Noting that phase conjugation in the frequency domain correspond to time reversal in the time domain, the two impulse response terms in the last equality of Eq. (2) are respectively: (1) the causal impulse response which comes from noise events that propagate from 1 to 2 and yield a nonzero correlation for a positive time delay, and (2) the time-reversed (or anti-causal) impulse response which comes from noise events that propagate from 2 to 1 and yield a nonzero correlation at a negative time delay. Thus, for an isotropic noise source distribution or a fully diffuse noise

field, the noise cross-correlation function is a symmetric function in time.

When inverse transformed to the time domain for the simple case of a homogeneous medium with attenuation,[26,24,13] Eq. (2) reduces to

$$\frac{d\langle C_{1,2}(t)\rangle}{dt} \approx Q\Omega a(G_{1,2}(t) - G_{2,1}(-t)) \tag{3}$$

where Q is the noise energy recorded by one sensor, Ω is a frequency dependent factor depending on the medium's attenuation, and a is the distance between the two sensors. The impulse response $G_{1,2}(t)$ in Eq. (3) is the solution of the standard wave equation for an impulsive (acoustic or elastic) source. To derive Eq. (3) from Eq. (2), it is assumed that the two signals $S_1(t)$ and $S_2(t)$, and the source term for the impulse response $G_{1,2}(t)$, have the same units. Here, the measured quantities are all accelerations; thus, the impulse response arrival times are estimated from the time derivative of the noise cross-correlation function.[13,24,26]

III. COMPARISON BETWEEN ACTIVE MEASUREMENT OF THE IMPULSE RESPONSE AND PASSIVE NOISE CROSS-CORRELATION FUNCTION FROM TURBULENCE-INDUCED FLOW VIBRATIONS

A. Experimental setup: Flat plate model tests

The structural vibration experiments were conducted at the U.S. Navy's William B. Morgan Large Cavitation Channel (LCC), the world's largest low-turbulence recirculating water tunnel.[21] The dimensions of the LCC test section are 13 m (length) and 3.05 m × 3.05 m (width and height). The test model[19] was constructed in three sections having stainless steel skins, nominally 12.7 and 19 mm thick, that were bolted and welded to box-beam frames, respectively. The three sections were coupled in the LCC's test section via slot-and-key mechanisms into a single structure 12.9 m long, 3.05 m wide, and 18.4 cm thick. These dimensions include the model's 4-to-1 elliptical leading edge and a 15° degree full-angle truncated-wedge trailing edge. The mass of the fully assembled model was approximately 17,000 kg. A schematic diagram of the test model including the accelerometer locations is provided as Fig. 1. The model was mounted with its center plane positioned 5 cm below the vertical centerline of the LCC test section and tests were conducted at free stream flow speeds from 3 to 20 m/s. Thus, the downstream-distance-based Reynolds numbers ($\mathrm{Re}_x = \int_0^x U dx/\nu$ where x is the distance from the plate's leading edge, U is the average free stream flow speed above the plate, and ν is the kinematic viscosity of water at the measured average water temperature of 20.4 °C for these tests) reached as high as 210 million. Hydrodynamic findings from these tests are available elsewhere[31,19] and are not repeated here.

For the present study, turbulent-flow induced plate vibrations were recorded at an average free stream flow speed, $U = 20.0$ m/s. For this test model at these flow speeds, the primary vibratory excitation comes from turbulent boundary layer pressure fluctuations which acted vertically, perpendicular to the plate. Such pressure fluctuations appear to convect over the plate surface at a speed that is 70–80% of the free stream speed. However, turbulent boundary layer pressure fluctuations decorrelate over distances greater than about one or two boundary layer thicknesses. Given that the measured boundary layer thickness in these experiments never exceeded 10 cm, the plate's vibratory excitation came from a large number of independent spatially distributed sources. The primary plate response to this nearly uniform and random vibration source distribution was also vertical (i.e., normal to the plate surface and to the free stream directions). Thus, accelerometers were mounted within the plate to measure vertical vibratory accelerations.

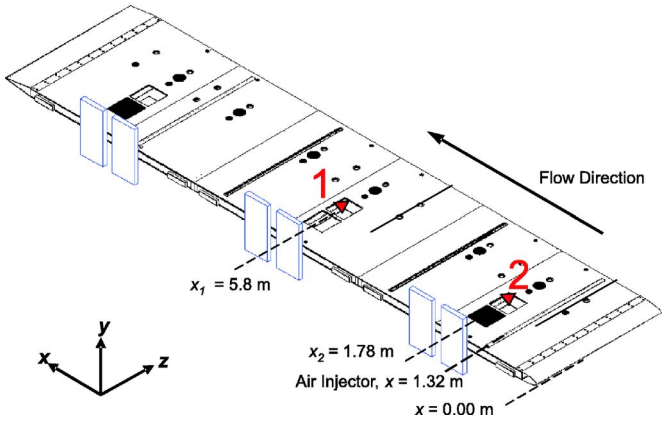


FIG. 1. (Color online) Geometry of the rigid flat plate. Accelerometer 2 is mounted in the upstream measurement box at a downstream distance $x_2 = 1.78$ m from the leading edge of the flat plate. Accelerometer 1 is mounted in the midstream measurement box at a downstream distance $x_1 = 5.8$ m.

For the noise cross-correlation function analysis presented here, the output from two single-component accelerometers (Wilcoxon 754-1) was used. These were located in the test-plate's interior at $x_1 = 5.80$ m and $x_2 = 1.78$ m (see Fig. 1). The ±3 dB frequency range of the accelerometers was 2 Hz–19 kHz. The acceleration signals were low pass filtered at 5 kHz, sampled at 10 kHz and divided into successive 10-s-long data sets. Multiple data sets were obtained for each flow condition.

B. Ambient vibration data processing and impulse response measurement

The ambient vibration cross-correlation technique works best when the noise-source distribution is uniform in space and time.[3,25] Hence the effects of high amplitude noise events should be minimized in the accelerometer recordings since they might otherwise dominate the arrival-time structure of the noise cross-correlation function. One straightforward way to discount high-amplitude events is to discard the amplitude completely by keeping only the sign of the signals.[3,32] However, such severe clipping creates artificial high-frequency spectral content that modifies the ambient vibration spectrum. Furthermore, 1 bit truncation accentuates the incoherent electronic noise component which was not negligible in these experiments. Thus, to minimize the effect of clipping, a clipping threshold was set for each accelerometer to three times the standard deviation of the ambient vibration after filtering the data in the frequency band

50–4950 Hz. In this manner, the effect of large events is reduced, but the high-frequency content of the ambient noise is less distorted and the impact of electronic noise remains fairly low.

To validate the passive extraction of the true impulse response $G_{1,2}(t)$ between accelerometer 1 and 2 from the noise cross-correlation function, active vibration source experiments without flow were conducted also. A compressed-air-cylinder impact source (Bimba FLAT-1 model, 80 psi) was placed 14 cm downstream of the first accelerometer ($x=5.94$ m). As discussed later in Sec. III C, the wavelengths of interest in this study are on the order of a few meters, hence the 14 cm separation between the impact source and the true location of the first accelerometer was neglected when using the knocker impact to determine the true structural impulse response $G_{1,2}(t)$ between accelerometers 1 and 2. The cylinder-actuation pressure was set as high as possible to ensure the best possible signal on accelerometer 2 while preventing mechanical overload (saturation) of accelerometer 1. Hence once deconvolved using the recording of accelerometer 1, the knocker impact recorded at accelerometer 2 yields an approximation of the true structural impulse response $G_{1,2}(t)$ which is denoted $H_{1,2}(t)$. The active-source impulse response estimate, $H_{1,2}(t)$, relates a vertical unit concentrated acceleration impulse at receiver 1 to the vertical acceleration response at receiver 2. Unfortunately, experimental limitations prevented reliable determination of $H_{1,2}(t)$ at frequencies below 500 Hz.

C. Comparison of the passive noise cross-correlation function to the active impulse response

The noise cross-correlation function $C_{1,2}(t)$ between the two accelerometers was computed for each of the 60 non-overlapping noise recording data sets and then averaged coherently (or stacked). This corresponds to a total recording duration $T_r=10$ min. Figure 2 is a time-frequency analysis of the noise cross-correlation function for a nominal flow speed of $U=20$ m/s. The noise cross-correlation function was filtered using multiple narrowband filters (fourth order Butterworth) of fixed bandwidth $B=[200\ \text{Hz}]$ and increasing center frequency f_c from 150 to 1500 Hz in increments of 20 Hz. In each frequency subband $[f_c-B/2, f_c+B/2]$ a normalized noise cross-correlation function denoted $NC_{1,2}(t;f_c)$ can be defined by dividing the filtered noise cross-correlation function with the square of the averaged energy of the filtered data in the same subband

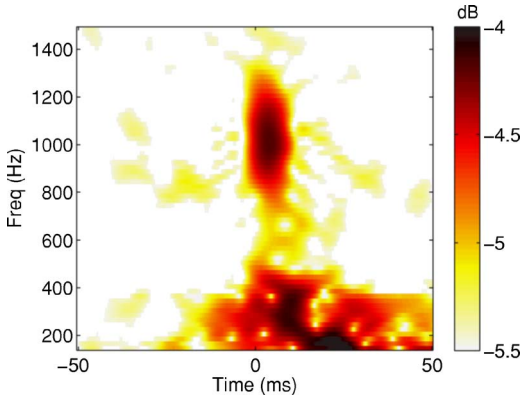


FIG. 2. (Color online) Time-frequency spectrogram of the normalized noise cross-correlation function (see Eq. (4)) between a pair of accelerometers mounted inside the plate test model (see Fig. 1) 402 cm apart. Ten minutes of ambient vibrations were correlated for a nominal flow speed of $U=20$ m/s. The color values of the correlation coefficient are logarithmic.

$$NC_{1,2}(t;f_c)=\frac{\sum_{k=1}^{k=N}\left(\int_{2\pi(f_c-B/2)}^{2\pi(f_c+B/2)}\tilde{S}_{1,k}^{*}(\omega)\tilde{S}_{2,k}(\omega)e^{-i\omega t}d\omega\right)}{\sqrt{\sum_{k=1}^{k=N}\left(\int_{2\pi(f_c-B/2)}^{2\pi(f_c+B/2)}|\tilde{S}_{1,k}(\omega)|^2d\omega\right)}\sqrt{\sum_{k=1}^{k=N}\left(\int_{2\pi(f_c-B/2)}^{2\pi(f_c+B/2)}|\tilde{S}_{2,k}(\omega)|^2d\omega\right)}},\tag{4}$$

where $\tilde{S}_{1,k}$ and $\tilde{S}_{2,k}$ are the Fourier transforms of the recorded signals of the kth data set. The denominator in Eq. (4) is a normalization factor that mitigates variations in sensor responses and helps in determining the actual coherence of the noise field between the stations. Indeed, the normalized noise cross-correlation function $NC_{1,2}(t)$ reaches a maximum closer to unity for correlated vibrations between the two sensors and a clear coherent peak can then be expected. In the case of decorrelated random vibrations (e.g., when the accelerometers are too widely separated), no coherent arrivals emerge from $NC_{1,2}(t;f_c)$). Instead, $NC_{1,2}(t;f_c)$ has a variance proportional to $1/2BT_r$.[25,33]

In the present case, a dominant coherent arrival emerges from the noise cross-correlation function in several frequency intervals in Fig. 2 (e.g., see $f=300$ Hz and $f=1$ kHz) with the dispersive nature of the strongest arrival most visible at low frequencies ($f<400$ Hz). Based on its frequency-dispersion behavior and slow group speed (see below), the coherent arrival of the noise cross-correlation function can be identified as the first anti-symmetric vibration mode of the test plate, commonly referred to as the A_0 mode.[34] The A_0 is the dominant arrival mainly because: 1) the relatively small thickness of the test model's skins and the low-frequency range recorded ($f<5$ kHz) do not allow for the efficient generation, propagation, and recording of

higher-order vibratory wave-propagation modes, and 2) the turbulent boundary layer pressure fluctuations primarily excite the mode A_0 and not the symmetric mode S_0 since they act as a surface distribution of random vertical-acting sources on only one side of each of the test plate's skins. The weaker coherent arrivals in the noise cross-correlation function at higher frequencies may also be due to the discontinuous structure of the test plate. For manufacturing and assembly disassembly reasons, the test plate was made of three different sections, each having a welded mild steel frame. The edges of the three frames were bolted together but each section may still appear distinct, structurally speaking, above some frequency. The two accelerometers were mounted inside separate measurement boxes located on the first and second section of the plate, respectively (see Fig. 1). Hence the increased attenuation of the higher frequency components of the coherent noise field may be due to these structural discontinuities in the test model.

For the case of a uniform noise source distribution or a fully diffuse noise field, the noise cross-correlation function is a symmetric function with respect to time, as stated in Eq. (2). However, in this experiment the lack of time symmetry of the noise cross-correlation function indicates that the random field in the test plates generated by the turbulent flow is not isotropic.[35] Indeed most of the coherent noise in the test plate appears to be propagating upstream. This is consistent with the fact that 55% of the model was downstream of accelerometer 1 while less than 15% of it was upstream of accelerometer 2. Thus, the prevalence of upstream propagating vibration waves that were recorded by both sensors should greatly exceed that of such downstream propagating vibrations waves. Therefore, the finite size of the model and the asymmetrical placement of the accelerometers are the most likely causes of the lack of time symmetry of the measured noise cross-correlation function.

To validate the passive extraction of the impulse response from the vibration noise recordings, Fig. 3 compares $dC_{1,2}/dt$ and $H_{1,2}(t)$, both bandpass filtered to the frequency range 900–1200 Hz, where the greatest coherence in the two accelerometer signals is present (see Fig. 2) at a nominal flow speed of $U=20$ m/s.

Here, both functions were normalized by their peak amplitude to ease comparison. The agreement between the time derivative of the noise cross-correlation function and active impulse response is good (see Fig. 3(a)) and can be even improved by equalizing their spectral amplitudes (see Fig. 3(c)). This spectral equalization helps to minimize the influence of spectral-amplitude disparities between turbulent boundary layer turbulence and the pneumatic knocker excitation. The two wave forms appear to be in phase, as predicted theoretically (see Eq. (2)) but the envelopes of the two wave forms differ. This may result from the directionality of the noise source radiation patterns which creates an amplitude shading of the noise cross-correlation function when compared to the true impulse response. Similar amplitude shading effects were reported in previous studies of cross-correlation function of surface-generated ocean noise, where the shallow noise sources act as dipole sources due to presence of the ocean free surface.[12,13] Furthermore, note that the

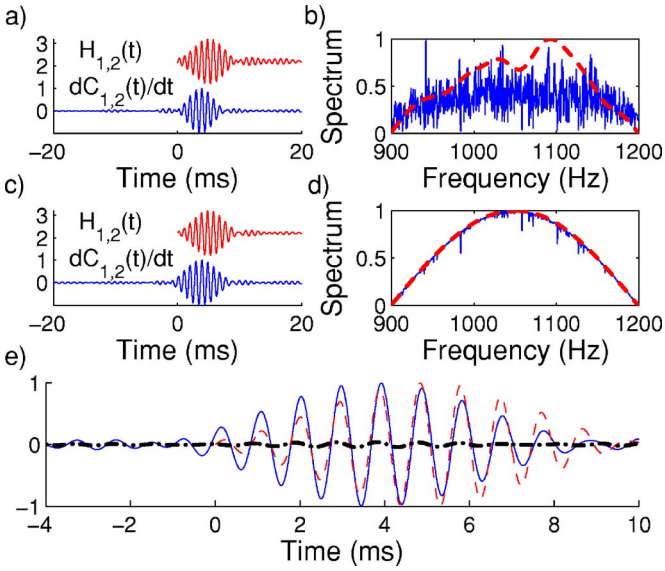


FIG. 3. (Color online) (a) Comparison of the estimated impulse response, $H_{1,2}(t)$, and the time derivative of the noise cross-correlation function, $dC_{1,2}(t)/dt$, between accelerometers 1 and 2 in the frequency band (900–1200 Hz). (b) Corresponding spectra (plain line: $dC_{1,2}(t)/dt$, dashed line: $H_{1,2}(t)$), (c) Same as (a) but after spectral whitening of both wave forms as shown in (d). (e) Zoom of the wave forms shown in (c) (plain line: $dC_{1,2}(t)/dt$, dashed line: $H_{1,2}(t)$). The resulting time derivative of the noise cross-correlation function obtained from recordings of incoherent vibrations from the two accelerometers is displayed also (thick dash-dot line).

impulse response was measured under a no-flow condition and several hours after $dC_{1,2}/dt$ was measured. Thus it shows that the primary coherent arrival of $dC_{1,2}/dt$ can be used as a robust and stable estimate of the main impulse response arrival time. The signal-to-noise ratio (SNR) of the wave from $dC_{1,2}/dt$ can be defined as the ratio of the maximum of the envelope of the main arrival to the standard deviation (STD) in an incoherent noise-only time window, (selected as $|t| \geqslant 0.8$ s) and is 47 dB in this case. The arrival time of the A_0 mode is 4 ms which corresponds to a group speed of around 1000 m/s at the center frequency of 1050 Hz given the sensor separation of 402 cm. Figure 3(e) shows the time derivative of the noise cross-correlation function (plain line) and the impulse response (dashed line) for time delays centered around the observed A_0 mode arrival only. In addition the time derivative of the incoherent noise cross-correlation function obtained from correlating the same duration of vibrations data on the two accelerometers but recorded at different times is also displayed (thick dash-dot line). In practice, this incoherent noise cross-correlation function was obtained using uncorrelated random vibrations data sets on each accelerometer recorded 30 s apart (i.e., by substituting $\tilde{S}_{2,k}(\omega)$ by $\tilde{S}_{2,k+3}(\omega)$ in Eq. (4)). As discussed previously, this incoherent noise cross-correlation function is a wave form having the same variance $1/2BT_r$ as $NC_{1,2}(t;f_c)$, but without any significant coherent arrival.[25,33] This confirms that the observed A_0 mode from the wave form can only result from propagation of coherent noise between the two accelerometers; such results do not arise fortuitously from the measurements. Overall, Fig. 3 supports the contention that an estimate of the structural impulse response be-

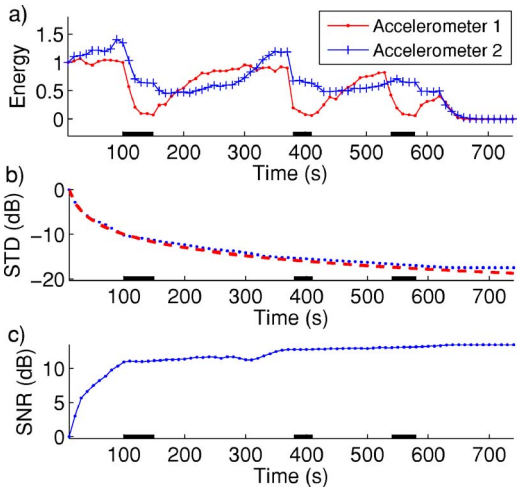


FIG. 4. (Color online) (a) Variations of the recorded energy by accelerometer 1 (dotted line) and accelerometer 2 (crossed line) in successive 10 s duration intervals in the frequency band (900–1200 Hz). The values were normalized to unity by the recorded energy during the first 10 s. The time intervals corresponding to the three successive air bubble injections events are indicated by thick black lines on the horizontal axis. (b) Corresponding decay of the standard deviation (STD) of the noise cross-correlation function $C_{1,2}(t)$ (computed for $t>0.8$ s) vs accumulated recording time T_r (dotted line) along with the predicted theoretical decay of $1/\sqrt{T_r}$ (dashed line). (c) Evolution of the signal-to-noise ratio (SNR) of the noise cross-correlation function. Each curve was normalized to unity at $t=0$.

tween the two accelerometers can be obtained passively from the time derivative of the noise cross-correlation function without the use of an active source.

D. Influence of the flow condition

In practical applications of this noise cross-correlation technique for structural monitoring, a fundamental question is to determine the extent to which the turbulent flow conditions influence the structural impulse response estimated from the noise cross-correlation function and its relationship to the true impulse response between the two accelerometers. An exhaustive answer is beyond the scope of this proof-of-concept paper and will be likely weakly dependent on the particular test models and flow types. However, a simple experiment is presented here which illustrates the influence of the flow condition. Here, the flow variation was provided by the injection of gas from a narrow spanwise slot inclined in the flow direction and located 1.32 m from the leading edge of the test plate (see Fig. 1). The intermittent injection of compressed air from the slot injector modifies the turbulent boundary layer on one side of the test model and consequently the random vibration's characteristics. The purpose and hydrodynamic effects of such air injection are described elsewhere.[19]

Figure 4 illustrates the influence of intermittent injection of gas at a rate of 200 standard cubic feet/min at nominal flow speed of $U=20.0$ m/s. The discontinuous thick black lines along the time axes indicate approximately the three time intervals when gas was injected into the flow. Figure 4(a) plots the variations of the normalized recorded vibration energy by the two accelerometers.

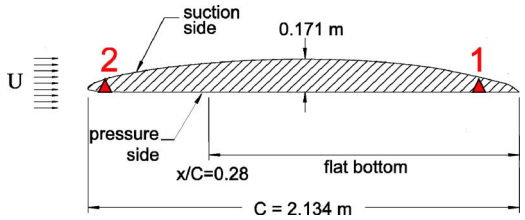


FIG. 5. (Color online) Cross sectional view of the hydrofoil geometry and accelerometer locations. The upstream flow direction is parallel to the flat portion of the foil's pressure side. All dimensions are in meters. Accelerometer 1 was mounted 15.24 cm from the trailing edge. Accelerometer 2 was mounted 12.18 cm from the leading edge. Accelerometers 1 and 2 were in different spanwise locations, respectively 13.3 and 61 cm from the test section wall.

In the absence of gas bubbles (e.g., from 0 to 100 s), the turbulent boundary layer induced vibrations of the test model are fairly constant. The first injection of air (around 100 s) disrupts the turbulent boundary layer, adds compressibility to the water flow, and reduces the level of recorded vibrations. When the bubble injection is discontinued (around 150 s), the vibration energy level recovers but not quite to the pre-injection level. Residual air bubbles distributed throughout the flow that alter its compressibility are believed to be the cause of the imperfect vibration energy recovery. This vibration-energy pattern repeats itself for the two subsequent gas injection periods (between 380 and 410 s and 540 and 580 s). After 620 s, the flow speed U was reduced from 20.0 to 0 m/s, and the vibration level for both accelerometers approached zero as expected.

Figure 4(b) shows the standard deviation STD of the time derivative of the uncorrelated portion of the noise cross-correlation function which converges like $1/\sqrt{T_r}$ as predicted theoretically,[25,33] assuming a distribution of uncorrelated noise sources. The SNR increases with recording time despite the variations in the flow conditions (see Fig. 4(c)): a longer recording interval simply helps to accumulate more coherent noise events between the two accelerometers. When the amount of coherent events recorded is reduced drastically due to bubbles injection, the SNR plateaus (between 100 and 140 s). Once air injection is discontinued, the growth of the SNR simply resumes, but is slower (e.g., 140–290 s) until the flow velocity U is decreases to zero (after 620 s).

The main experimental finding here is that the estimated impulse response is primarily a characteristic of the mechanical structure of the test model, and is not strongly influenced by the characteristics of the turbulent excitation.

IV. APPLICATIONS TO STRUCTURAL HEALTH MONITORING OF HYDROFOIL MOUNTING

A. Experimental setup: Two-dimensional hydrofoil tests

The second set of experiments was conducted in the LCC water tunnel with a different test model: a solid Ni–Al bronze hydrofoil with a 2.134 m chord(=C), a 3.05 m span, 17.1 cm maximum thickness (see Fig. 5). The basic foil section was a NACA 16 modified to have a pressure side that was nearly flat and a suction side that terminated in a blunt

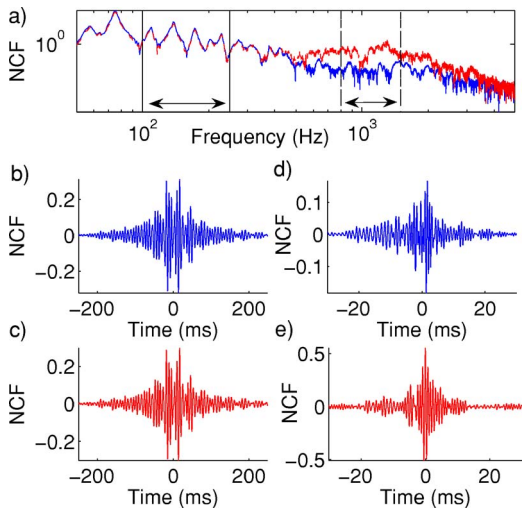


FIG. 6. (Color online) (a) Spectrum of the noise cross-correlation function vs frequency: plain (blue) line pre-cavitation-test mounting, and dashed (red) line post-cavitation-test mounting. (b) Normalized noise cross-correlation function wave form (see Eq. (4)) for the frequency band 100–250 Hz for pre-cavitation-test mounting. (c) Same as (b) but for postcavitation-test mounting. (d) Normalized noise cross-correlation function wave form for the frequency band from 800 to 1500 Hz for precavitation-test mounting. (e) Same as (b) but for postcavitation-test mounting. The duration of the random vibrations recording was 5 min. Note the different time axes for (b) and (c) vs (d) and (e).

trailing edge bevel. Further description of the experimental setup can be found in the previous literature.[20]

For the results presented here, the inlet flow speed upstream of the foil was $U=20$ m/s and the average water temperature was 32 °C, leading to a chord-based Reynolds number ($=UC/\nu$) of approximately 50 million. The instrumentation was the same as for the plate vibration experiments. Two single-component accelerometers were mounted within machined cavities in the hydrofoil, 15.24 cm from the trailing edge (No. 1) and 12.18 cm from the leading edge (No. 2). Both accelerometers were in the same vertical location, 3.8 cm above the flat pressure side of the hydrofoil, but in different spanwise locations, respectively 13.3 cm (No. 1) and 61 cm (No. 2) from the test section wall. Hence the three-dimensional distance between the two accelerometers was 214.7 cm. The data processing followed the description provided in Sec. III B.

B. Variations of the noise cross-correlation function

At one point in these experiments, the hydrofoil and its mountings were subjected to significant loading from unsteady sheet cavitation. These loads were large enough to change the mechanical characteristics of the foil's mounts, and, for the purposes of this investigation, these mechanical changes are apparent in noise cross-correlation function results obtained before and after the sheet cavitation tests. Figure 6(a) displays the cross power spectrum of the normalized noise cross-correlation function under these two mounting conditions based on multiple vibration recordings of 5 min cumulative length. The results shown in this figure are stable and unchanged for 1–10 min total vibration recording times. Below 400 Hz, the two spectra are essentially identical. In particular the location and relative magnitude of the resonance peaks remain unchanged at low frequencies. These resonance peaks correspond to vibration modes of the fluid-loaded mounted foil, hence these results indicate that the basic structural integrity of the hydrofoil was not affected by the sheet-cavitation testing. However, the before- and after-cavitation-testing spectra differ at frequencies above 500 Hz with the after-cavitation-testing results showing greater spectral content. Frames (b) through (e) of Fig. 6 further illustrate these results. Figures 6(b) and 6(c) show that the normalized noise cross-correlation function before-and-after wave forms for a 100–250 Hz bandwidth are essentially identical and nearly symmetric in time, while Figs. 6(d) and 6(e) show that the normalized noise cross-correlation function before-and-after wave forms for an 800–1500 Hz bandwidth are different both in character and magnitude. In the higher frequency band, neither noise cross-correlation function is symmetric in time and the wave forms no longer match.

These noise cross-correlation function results point to a few features of hydrofoil vibration in these experiments that could not have been deduced easily from other measurements. First of all, the time symmetry of the noise cross-correlation function from 100 to 250 Hz suggests that the foil's vibration for both mounting conditions was nearly isotropic and does not possess a preferred direction in this frequency range even though the flow excitation clearly has one. In the higher-frequency band, the pre-cavitation-test results shown on Fig. 6(d) place the strongest noise cross-correlation function arrival at $t=+2$ ms. This indicates that there was more vibration excitation near the foil's trailing edge than its leading edge in this frequency range. This observation is consistent with the hydrodynamic character of the flow with stronger boundary layer turbulence being found on the aft half of the foil. Interestingly, the noise cross-correlation function in the higher frequency band after cavitation testing shown on Fig. 6(e) indicates a large amplitude peak near $t=0$. This peak does not result from noise sources that propagate between the receivers along the impulse response paths. Such a peak may indicate an increased level of vibration noise produced at a point approximately equidistant from the two accelerometers. Given the placement of the two accelerometers and the mid-chord mechanical attachment of the foil to the LCC test section, this change in the noise cross-correlation function is consistent with some alteration in the foil's mounting, presumed to have occurred during the sheet cavitation tests. Furthermore, when the foil's mounts were disassembled, minor indentations were found in the structural extensions (tangs) of the foil that were captured by the LCC sidewall clamps, and these indentations were not observed before the sheet cavitation tests. By design each tang was sandwiched between an upper and lower mounting shim that was sized for a slight interference fit within the LCC sidewall clamps. In the initial installation, these shims were hammered into place with significant difficulty. Post-cavitation, the shims were found to be more nearly a size-on-size fit, and required only a light force for removal and re-installation. This change from an interference fit to a size-on-size fit was consistent with the minor plastic deformation observed on the bearing surface of the tangs. When taken together, this circumstantial structural evidence suggests that

the noise cross-correlation function produced by hydrodynamic fluctuations on a high-Reynolds-number hydrofoil can be used to detect changes in the mounting conditions.

V. SUMMARY AND CONCLUSIONS

Two separate ambient vibration experiments were conducted at high Reynolds number at the Naval Surface Warfare Center's William B. Morgan Large Cavitation Channel (LCC) in Memphis Tennessee. Cross correlations of hydrodynamic-flow-induced vibrations from two accelerometers were used to estimate the impulse response(s) between the two accelerometer locations for a steel skin-on-frame flat plate and a solid Ni–Al bronze hydrofoil undergoing hydrodynamic tests. Although the measured vibrations appeared to be random, stable coherent cross-correlation results in a 50 Hz–5 kHz bandwidth emerged after recording times of a few minutes.

The following three conclusions can be drawn from this investigation. First, intentional de-synchronization of the random vibration data destroyed the coherent portion of the cross correlation, thus the final noise cross-correlation function results presented here are unlikely to have been produced fortuitously. Second, the noise-correlation-determined estimates of the impulse responses between accelerometers match active measurements of the impulse response within the experimental limitations of these tests. Here, these limitations arose from imperfect collocation of the active noise source with one of the accelerometers, the unequal spectral excitation of turbulent boundary layers and a pneumatic knocker, and the finite amount of measured data from which the various cross correlations could be computed. And third, the influence of the turbulent flow conditions on the results was found to be minimal compared to the influence of structural changes in the test model and its mounting.

This final point is perhaps the most interesting. It implies that turbulence-induced vibrations can be used for passive real-time monitoring of structures. If the accelerometers (or other sensors) are placed at the time of structural assembly, this technique may be applied to structural monitoring without disassembly or withdrawal of the structure from its intended service. The increasing reliability and cost effectiveness of wireless receivers makes this method particularly advantageous with a dense network of sensors (e.g., microelectromechnical systems) distributed on or within a vibrating structure. In particular, this method can return useful information in the absence of active sources and it can benefit from the potentially high density of cross paths between multiple sensors pairs to increase monitoring sensitivity.

ACKNOWLEDGMENTS

For the hydrofoil experiments, the authors acknowledge the contributions of Shiyao Bian, Joshua Hamel, Carolyn Judge, and Kent Pruss of the University of Michigan; William Blake, Michael Cutbirth, Ken Edens, Robert Etter, Ted Farabee, Jon Gershfeld, Joe Gorski, Tom Mathews, David Schwartzenberg, Jim Valentine, Phil Yarnall, Joel Park, and the LCC technical staff from the Naval Surface Warfare Center—Carderock Division Memphis Detachment; and Pat Purtell and Candace Wark from the Office of Naval Research. The hydrofoil experiments were supported by the Office of Naval Research under Contract Nos. N00014-99-1-0341 and N00014-99-1-0856. For the flat plate experiments, the authors wish to acknowledge the significant contributions of Kent Pruss, of the University of Michigan; Robert Etter, Bruce Hornaday, Michael Cutbirth, and the LCC technical staff from the Naval Surface Warfare Center-Carderock Division Memphis Detachment; and Duncan Brown of the Johns Hopkins University Applied Physics Laboratory. The flat plate experiments were sponsored by the Defense Advance Research Projects Agency (Lisa Porter and Thomas Beutner Program Managers) under Contract No. HR0011-04-01-0001, and the Office of Naval Research (Pat Purtell, Program Manager) under Contract No. N00014-01-1-0880. The content of this document does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

[1] D. Ewins, *Modal Testing: Theory and Practice* (Wiley, New York, 1985).
[2] 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).
[3] E. Larose, A. Derode, M. Campillo, and M. Fink, "Imaging from one-bit correlations of wideband diffuse wavefields," J. Appl. Phys. **95**, 8393–8399 (2004).
[4] A. E. Malcolm, J. A. Scales, and B. van der Tiggelen, "Retrieving the Green function from diffuse equipartitioned waves," Phys. Rev. E **70**, 015601(R) (2004).
[5] C. Farrar and G. James, "System identification from ambient vibration measurements on a bridge," J. Sound Vib. **205**, 1–18 (1997).
[6] J. M. Caicedo, E. Clayton, S. J. Dyke, and M. Abe, "Structural health monitoring for large structures using ambient vibrations," in *Proceedings of the ICANCEER Conference, Hong Kong*, pp. 379–384 (2002).
[7] T. Nagayama, M. Abe, Y. Fujino, and K. Ikeda, "Structural identification of a non-proportionally damped system and its application to a full-scale suspension bridge," J. Struct. Eng. **131**, 1536–154 (2005).
[8] S. Lin, J. Yang, and L. Zhou, "Damage identification of a benchmark building for structural health monitoring," Smart Mater. Struct. **14**, 162–169 (2005).
[9] R. Snieder and E. Şafak, "Extracting the building response using seismic interferometry; theory and application to the Millikan Library in Pasadena, California," Bull. Seismol. Soc. Am. **96**, 586–598 (2006).
[10] R. Snieder, J. Sheiman, and R. Calvert, "Equivalence of the virtual source method and wavefield deconvolution in seismic interferometry," Phys. Rev. E **73**, 066620 (2006).
[11] G. James, T. Carne, and J. P. Lauffer, "The natural excitation technique for modal parameter extraction from operating wind turbines," Report No. SAND92-1666, UC-261, Sandia National Laboratories (unpublished).
[12] P. Roux, W. A. Kuperman, and the NPAL Group, "Extracting coherent wavefronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[13] 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).
[14] K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, W. S. Hodgkiss, and W. A. Kuperman, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[15] K. Wapenaar, "Retrieving the elastodynamic Greens function of an arbitrary inhomogeneous medium by cross correlation," Phys. Rev. Lett. **93**, 254301 (2004).
[16] K. G. Sabra, P. Gerstoft, P. Roux, W. Kuperman, and M. C. Fehler, "Surface wave tomography using microseisms in southern california," Geophys. Res. Lett. **32**, L023155 (2005).
[17] N. M. Shapiro, M. Campillo, L. Stehly, and M. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **29**, 1615–1617 (2005).
[18] J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge

18, 957–960 (1999).
[19]W. Sanders, E. S. Winkel, D. R. Dowling, M. Perlin, and S. Ceccio, "Bubble friction drag reduction in a high-Reynolds-number flat-plate turbulent boundary layer," J. Fluid Mech. **552**, 353–380 (2006).
[20]D. Bourgoyne, J. Hamel, S. Ceccio, and D. Dowling, "Time-averaged flow over a hydrofoil at high Reynolds number," J. Fluid Mech. **496**, 365–404 (2003).
[21]R. Etter, J. Cutbirth, S. Ceccio, D. Dowling, and M. Perlin, "High Reynolds number experimentation in the U.S. navy William B. Morgan Large Cavitation Channel," Meas. Sci. Technol. **16**, 1701–1709 (2005).
[22]H. Wenzel and D. Pichler, *Ambient Vibration Monitoring* (Wiley, New York, 2005).
[23]J. Juang and R. Pappa, "An eigensystem realization algorithm for modal parameter identification and model reduction," J. Guid. Control Dyn. **8**, 620–627 (1985).
[24]P. Roux, K. Sabra, W. Kuperman, and A. Roux, "Ambient noise cross-correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2005).
[25]K. G. Sabra, P. Roux, and W. A. Kuperman, "Emergence rate of the time-domain Greens function from the ambient noise cross-correlation function," J. Acoust. Soc. Am. **118**, 3524–3531 (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]F. Sanchez-Sesma and M. Campillo, "Retrieval of the Green function from cross-correlation: The canonical elastic problem," Bull. Seismol. Soc. Am. **96**, 1182–1191 (2006).
[28]W. A. Kuperman and F. Ingenito, "Spatial correlation of surface noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[29]R. L. Weaver and O. I. Lobkis, "Diffuse fields in open systems and the emergence of the Greens function," J. Acoust. Soc. Am. **116**, 2731–2734 (2004).
[30]P. J. Shorter and R. S. Langley, "On the reciprocity relationship between direct field radiation and diffuse reverberant loading," J. Acoust. Soc. Am. **117**, 85–95 (2005).
[31]E. S. Winkel, B. Elbing, D. Dowling, M. Perlin, and S. Ceccio, "High-Reynolds-number turbulent-boundary-layer surface pressure fluctuations with bubble or polymer additives," in *ASME IMECE Symposium on External Body Flow Noise*, November 5–11, Orlando, FL (2005).
[32]K. G. Sabra, P. Gerstoft, P. Roux, W. Kuperman, and M. C. Fehler, "Extracting time-domain Greens function estimates from ambient seismic noise," Geophys. Res. Lett. **32**, L0331 (2005).
[33]R. L. Weaver and O. I. Lobkis, "Fluctuations in diffuse field-filed correlations and the emergence of the Green's function in open systems," J. Acoust. Soc. Am. **117**, 3432–3439 (2005).
[34]J. D. N. CHeeke, *Fundamentals and Applications of Ultrasonic Waves* (CRC, Boca Raton, FL, 2002), pp. 162–167.
[35]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**, L003521 (2005).

Noise in the adult emergency department of Johns Hopkins Hospital

Douglas Orellana, Ilene J. Busch-Vishniac,[a] and James E. West
Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218

(Received 18 August 2006; revised 8 November 2006; accepted 19 January 2007)

While hospitals are generally noisy environments, nowhere is the pandemonium greater than in an emergency department, where there is constant flow of patients, doctors, nurses, and moving equipment. In this noise study we collected 24 h measurements throughout the adult emergency department of Johns Hopkins Hospital, the top ranked hospital in the U.S. for 16 years running. The equivalent sound pressure level (L_{eq}) throughout the emergency department is about 5 dB(A) higher than that measured previously at a variety of in-patient units of the same hospital. Within the emergency department the triage area at the entrance to the department has the highest L_{eq}, ranging from 65 to 73 dB(A). Sound levels in the emergency department are sufficiently high [on average between 61 and 69 dB(A)] to raise concerns regarding the communication of speech without errors—an important issue everywhere in a hospital and a critical issue in emergency departments because doctors and nurses frequently need to work at an urgent pace and to rely on oral communication. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2642309]

PACS number(s): 43.50.Jh, 43.50.Cb, 43.55.Gx [BSF] Pages: 1996–1999

I. INTRODUCTION

Noise in a hospital Emergency Department (ED) is unavoidable. It has been estimated by Vincent and Wears[1] that emergency medicine staff spend 80% of their time communicating face to face, by phone or by radio. Ten percent of the time, the staff have two or more conversations at once and a third of the time, conversations are interrupted by people. Further, Tijunelis *et al.*[2] have shown that even small events such as a garbage can closing, drawers shutting or phones ringing can cause the sound pressure levels in an ED to rise temporarily to 90–100 dB. (This article and others cited below do not make clear whether their data is A weighted. We assume from the use of dB rather than dB(A) that the results given are unweighted.)

In a study by Buelow[3] four Phoenix EDs, including three large urban hospitals and a medium-sized suburban hospital, were monitored for sound level comparisons. The mean sound pressure levels of the three large urban hospitals were 69, 70, and 73 dB, and of the medium-sized suburban hospital was 67 dB. At Baystate Medical Center in Springfield, MA, Baevsky[4] found the mean sound pressure level to be 57 dB. At an ED located in the University Hospital of Bellvitge (el Hospital Universitario de Bellvitge), Del Campo *et al.*[5] found sound pressure levels to range from 45 to 90 dB.

Zun and Downey[6] studied whether high noise levels affect the ability to hear heart and lung sounds. On average normal body sound levels are 22–30 dB in free space and 60–65 dB through a stethoscope. During their study they found minimum sound levels of 45 dB at the nursing station, trauma room, and private rooms. Mean sound pressure levels were 58, 56, and 46 dB, respectively, while maximum values for these three areas were 70, 81, and 62 dB, respectively. Zun and Downey tested medical staff hearing ability for heart and lung sounds in pink noise and found that 3.8% of the staff tested were unable to hear a heart tone and 8.7% were unable to distinguish lung sounds in the presence of pink noise set to mimic the levels in the ED.

The results cited above suggest that noise in an ED is a serious matter with the potential to compromise the quality of medical care delivery to patients with urgent and sometimes critical problems. However, although the results of specific hospital settings are very interesting, they are provided without the context of noise elsewhere in the same hospital, so it is not possible to compare ED noise to that of other hospital locations.

In this article, we present a study of noise in the Adult ED of Johns Hopkins Hospital (JHH). Results were obtained using the same experimental protocol employed for prior studies of noise in various in-patient units of JHH.[7] This permits us to more fully characterize the sound in the ED and to place it in the context of noise elsewhere in the hospital. Johns Hopkins Hospital is a particularly interesting choice of a facility to study as it is a very large urban hospital which has been ranked as the top hospital in the United States for the last 16 years in a row by *U.S. News and World Reports*.

II. METHODS

The JHH ED is located within three different buildings (Marburg, Park, and CMSC, which stands for the Children's Medical and Surgical Center). Throughout the course of a year they see about 59,000 patients in a space of 18,500 sq. ft. The ED is divided into the main clinical area and a hallway of offices. The main clinical area is further subdivided into a several areas: triage, urgent care, central nursing station, three patient treatment pods, and three critical care rooms. Figure 1 shows the architectural schematic of

[a]Author to whom correspondence should be addressed. Electronic mail: ilenebv@jhu.edu

 0001-4966/2007/121(4)/1996/4/$23.00

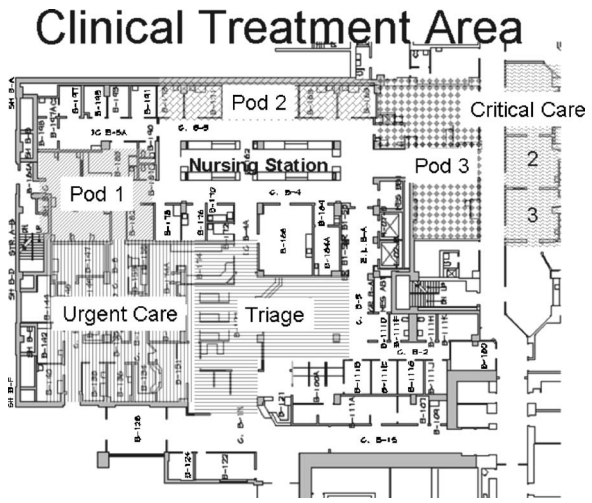


FIG. 1. Johns Hopkins Hospital Emergency Department schematic.

the department. Each pod specializes in a particular type of treatment, although the hectic pace of the ED often results in pods filling with patients based more on availability of space than on the type of treatment required. Pods vary significantly in architecture. The triage and urgent care areas are at the entrances of the ED. When ambulances arrive they must go through the triage area to get to the main clinical treatment area. Some of the pods have patient spaces separated by curtains, while others use separate rooms, Pod 3, for instance, is two separate rooms with patient beds separated by curtains while Pod 2 is a row of private rooms. The critical care rooms are three private rooms at one end of the clinical treatment area. All the pods surround a central nursing station. In the triage, a patient is seen in a cubicle open to the waiting area and a security desk. The urgent care area has its own waiting area and behind the waiting area there are closed examination rooms with a nursing station in the middle of all the rooms.

Sound pressure levels were obtained in the ED during weekdays in seven different areas using a Larson Davis System 824 sound level meter serving as both a precision sound level meter and real-time frequency analyzer. (We were unable to obtain measurements in Pod 2 because it was in constant use and inaccessible to us for installation of the sound level meter.) The sound level meter was placed at chest height and secured in each room while the sound was monitored for a 24 h period. The data were stored on the sound level meter and then downloaded onto a laptop for further analysis. In each measurement the instrument was placed inside a thin plastic bag to prevent contamination and a small hole was made to provide for the ac connection to the device. On the bag we taped a piece of paper explaining that the device was being used for a hospital study and our contact information. The doctors and staff were requested to go about their day as usual without making any changes in the manner in which they worked. All of the measurements were obtained over a two week period.

A-weighted equivalent sound pressure levels L_{eq} and flat frequency spectra in octave bands from 16 Hz to 16 kHz were obtained. For all measurements, the sound level meter slow setting was used.

III. RESULTS

Figure 2 shows the 24 h average A-weighted sound pressure levels at each measurement location obtained with the sound level meter on the slow setting. This figure shows the L_{eq}, the L_{max}, and the L_{min}. The L_{eq} is the continuous level that would produce the same amount of sound energy and thus represents the sound pressure level average. The L_{max} is the maximum level occurring over the measurement interval, and the L_{min} the minimum level observed during the measurement interval.

Figure 2 shows a number of interesting patterns. First, there is substantial variation from location to location within the ED. The L_{eq} values span 8 dB(A), the L_{max} about 7 dB(A), and the L_{min} roughly 11 dB(A). This suggests that a few of the measurement locations had an occasional quiet moment, while other areas likely did not. Second, there is no area which is clearly noisier than all others. Pod 1 displayed the highest L_{max} but Triage the highest L_{eq} and Critical Room 3 the highest L_{min}. In the Triage, the L_{eq} was 69 dB(A). Critical Room 2, which exhibited the lowest L_{eq} produced a level

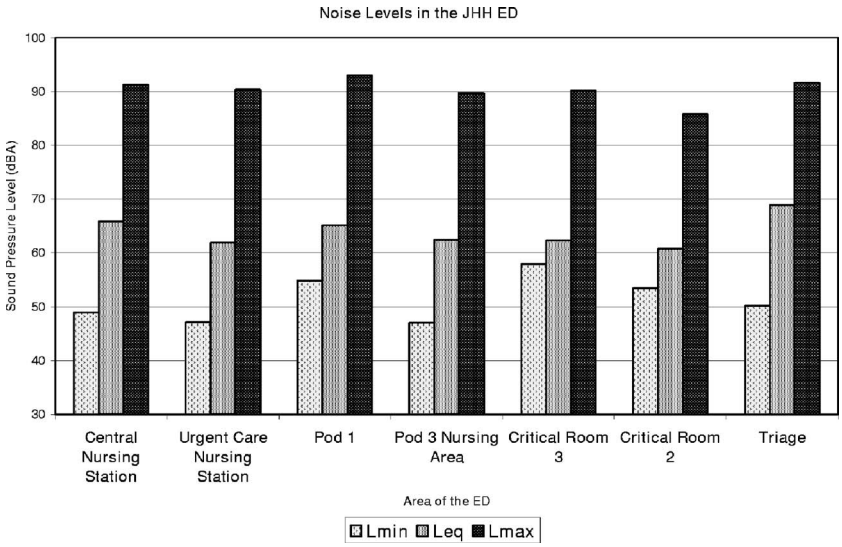


FIG. 2. Time-averaged A-weighted sound pressure levels at various locations in the Johns Hopkins Hospital Emergency Department.

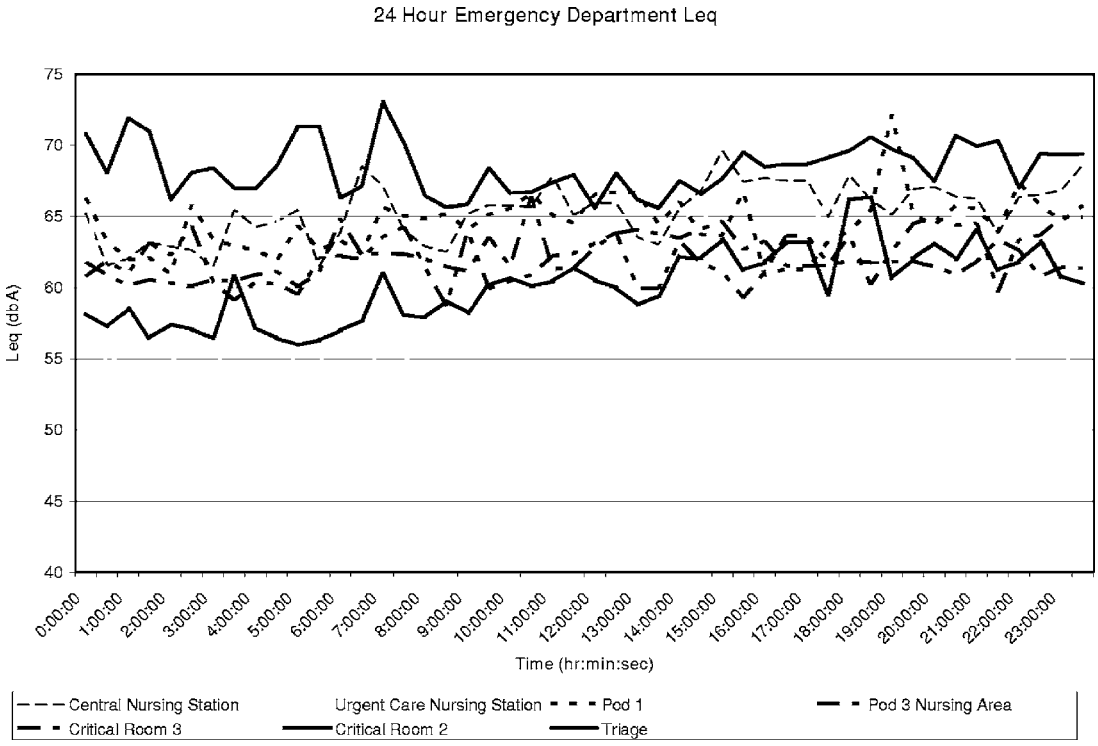


FIG. 3. L_{eq} vs time at each measurement location in the Emergency Department of Johns Hopkins Hospital.

of 61 dB(A). All of the levels measured are well in excess of the World Health Organization guidelines for hospitals[8] which suggest that equivalent sound levels be kept below 40 dB(A). However, as discussed in Busch-Vishniac *et al.*,[7] virtually all reports of noise in hospitals in the literature display sound pressure levels well in excess of the WHO guidelines.

The results shown in Fig. 2 can be compared to those previously measured in Johns Hopkins Hospital using the same equipment. These data have been reported in Busch-Vishniac *et al.*[7] Five hospital in-patient units at JHH were measured. The L_{eq} were found to vary from a low of 51 dB(A) on Weinberg 4C (an oncology unit in a building opened in 1999) to a high of 58 dB(A) in the Pediatric Intensive Care Unit in an older building. We note that all of the in-patient unit levels are below those found in the ED, typically by roughly 10 dB(A).

Figure 3 shows the L_{eq} at each measurement location as a function of time. In general, the levels are fairly constant although there is a small drop in level from midnight until about 7:00 a.m. This is a pattern similar to that observed elsewhere at JHH. The study previously reported by Busch-Vishniac *et al.*[7] found almost no variation of L_{eq} with time of day in the Pediatric Intensive Care Unit, and little variation with time in other monitored in-patient units.

Figure 4 shows the 24 h average L_{eq} as a function of frequency in octave bands from 16 to 16,000 Hz. As has been observed previously in JHH, the spectrum is quite flat in the speech range from roughly 125 to 2000 Hz. Below this range the sound level rises significantly and above this frequency range it tails off. Given the sheer volume of speech utterances in the ED, it comes as no surprise that the spectrum shows this pattern.

IV. DISCUSSION AND CONCLUSIONS

The results presented here reveal that noise is a serious issue in a busy Emergency Department. The sound pressure levels are high day and night and they are particularly high in the speech frequency band due to the need to communicate constantly to perform necessary functions. Compared to the in-patient units monitored in a previous study, the ED tends to exhibit sound levels roughly 5–10 dB(A) higher. The L_{eq} tends to be between 60 and 70 dB(A). While this is certainly not a high enough level to cause any concern regarding hearing damage of patients or staff, it is high enough that it is likely that people are using a somewhat raised voice in order to be understood above the din. The main concern is for patient safety which could be compromised by less than perfect speech communication. Additionally, medical staff fatigue is an issue since speaking in a raised voice is tiring. The present study has not sought to quantify the extent to which either of these issues is a problem in the Johns Hopkins Hospital Emergency Department and further speculation is unwarranted.

A prior study of hospital noise found that sound pressure levels throughout the world are less variable than one might expect.[7] Further, it was found that hospital sound levels were not particularly dependent on the type of hospital (community or major research, for instance). Thus, although the work reported here is for a single ED, there is reason to believe that the results might apply more broadly. Certainly, there is no surprise in finding that the sound pressure levels in an

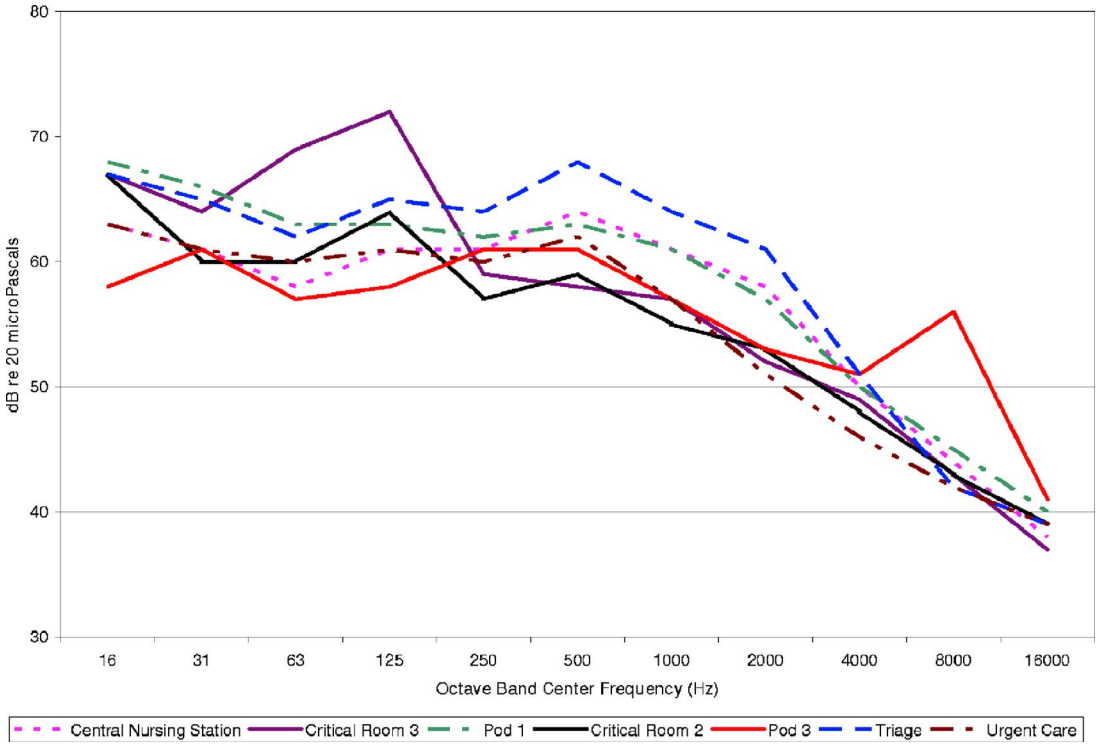


FIG. 4. (Color online) Spectrum in the Emergency Department of Johns Hopkins Hospital.

active Emergency Department are roughly 5–10 dB(A) above those in the in-patient units of the hospital.

ACKNOWLEDGMENTS

Our work at Johns Hopkins Hospital would not have been possible without the support of JHH leaders. We are indebted to Ron Peterson, President of Johns Hopkins Hospital, for his support of our noise study. We are also grateful to James Scheulen, Director of Emergency Medicine, and Kathy DeRuggerio, Assistant Director of Nursing, for their permission to monitor the Adult ED. We also greatly appreciate the active collaboration of all of the Adult ED staff. This work was supported by Johns Hopkins Hospital through the Center for Quality Improvements and Patient Safety, under the direction of Chip Davis.

[1]C. Vincent and R. Wears, "Communication in the emergency department: Separating the signal from the noise," Med. J. Aust. **176**, 409–410 (2002).
[2]M. Tijunelis, E. Fitzsullivan, and S. Henderson, "Noise in the ED," Am. J. Emerg. Med. **23**, 332–335 (2005).
[3]M. Buelow, "Noise level measurements in four Phoenix emergency departments," J. Emerg. Nurs. **27**, 23-26 (2001).
[4]R. Baevsky, "Sound levels in the Emergency Department setting," Acad. Emerg. Med. **13**, 233 (2006).
[5]S. del Campo Rodriguez, O. Gutierrez, and M. Jaramillo, "Ruidos contaminacion acustica en urgencias," Rev. Enferm **28**, 100–104 (2005).
[6]L. Zun and L. Downey, "The effect of noise in the emergency department," Acad. Emerg. Med. **12**, 663–666 (2005).
[7]I. Busch-Vishniac, J. West, C. Barnhill, T. Hunter, D. Orellana, and R. Chivukula, "Noise levels in Johns Hopkins Hospital," J. Acoust. Soc. Am. **118**, 3629–3645 (2005).
[8]B. Berglund, T. Lindvall, and D. S. (ed), "Guidelines for community noise," Technical Report, World Health Organization (1995).

Noise within the social context: Annoyance reduction through fair procedures

Eveline Maris,[a),b)] Pieter J. Stallen,[a),c)] Riel Vermunt, and Herman Steensma
Faculty of Social and Behavioral Sciences, Section of Social and Organizational Psychology, Universiteit Leiden, P.O. Box 9555, 2300 RB Leiden, The Netherlands

(Received 16 February 2006; revised 30 August 2006; accepted 7 January 2007)

The social context of noise exposure is a codeterminant of noise annoyance. The present study shows that fairness of the exposure procedure (sound management) can be used as an instrument to reduce noise annoyance. In a laboratory experiment ($N=117$) participants are exposed to aircraft sound of different sound pressure level (SPL: 50 vs 70 dB A)—which is experienced as noise—while they work on a reading task. The exposure procedure (fair versus neutral) is modeled in line with findings from social justice theory. In the fair condition, participants can voice their preference for a certain sound sample, although they cannot deduce whether their preference is granted. In the neutral condition, participants are not asked to voice their preference. Results show the predicted interaction effect of sound pressure level and procedure on annoyance: Annoyance ratings are significantly lower in the fair condition than in the neutral condition, but this effect is found only in the 70 dB condition. When the SPL is considerably disturbing, fair procedures reduce noise annoyance. Consequences of the reported findings for both theory and practice are discussed.
 [DOI: 10.1121/1.2535507]

PACS number(s): 43.50.Qp [BSF] Pages: 2000–2010

I. INTRODUCTION

It is a common observation that people's evaluations of environmental sound differ widely, given equal acoustics. Where to some the arousing roar of a Concorde supersonic jet was like music to their ears, to others it was truly intimidating noise (Adams, 1981). Noise is unwanted sound, and therewith a subjective description. Beside acoustic variables (e.g., loudness, pitch), nonacoustic variables (like perceived control, noise sensitivity, and attitudes toward the source) explain a substantial proportion of variance in annoyance reactions to noise (e.g., Job, 1988; Fields, 1993). Many studies of community reactions to noise consider nonacoustic variables. They address them mainly as personal sources of variance that blur the dosage-response relationship (e.g., Schultz, 1978; Fidell *et al.*, 1988; Schomer, 1988; Green and Fidell, 1991; Miedema and Vos, 1998). Few studies have addressed nonacoustic variables as a potential instrument to reduce (or increase) noise annoyance (e.g., Cederlöf *et al.*, 1967; Maziul and Vogt, 2002). The present study addresses the social side of sound exposure as a potential instrument for annoyance reduction.

"Increased attention in the late 1960s and early 1970s to noise as a social problem stimulated the initial interest of social psychologists in noise research" (Cohen and Spacapan, 1984, p. 221). This attention has since not faded. But despite this recognition of noise as a *social* problem, the research focus has not been on the social side of the issue, but rather on the acoustic side, specifically the measurement of annoyance, and the predictive relationship between noise metrics and annoyance. Job (1988) has reviewed a body of survey studies on subjective reactions to environmental noise, and concludes that even when data are collected with the most accurate measurement of both the acoustics and the annoyance reaction, noise exposure accounts for only 25%–40% of variation in reaction (Job, 1988; also Guski, 1999). A range of variables other than noise exposure has been shown to correlate significantly with annoyance. Such nonacoustic variables are repeatedly estimated to account for more variation in annoyance scores in survey data than acoustical variables do (e.g., Job, 1988; Fields, 1993; Guski, 1999; Ouis, 2001, 2002). Job (1988) presumes that some nonacoustic variables (i.e., attitudes toward the noise source and noise sensitivity), besides being part of the reaction to noise, may also be codeterminants of annoyance. In theory, ameliorating codeterminants of annoyance will result in annoyance reduction (Guski, 1999). However, this abatement strategy is interesting for policy makers only if such codeterminants are tractable on a large scale. In that respect, *social* nonacoustical variables (e.g., Guski, 2001) may possess the required features. In the present paper it is experimentally tested whether social nonacoustical variables are tractable on a group level, and whether they operate as codeterminants of noise annoyance. It is argued that, in order to study and hopefully profit from, the possibilities for annoyance reduction through social nonacoustic variables, noise as a *social* problem needs to be acknowledged. Sound exposure has a social side, and social processes have the potential of modifying nonacoustic codeterminants of noise annoyance.

In this paper, "management of the sound by the source" (Stallen, 1999) is explored as a nonacoustic instrument for annoyance reduction. To this end, noise annoyance is re-

[a)]Current affiliation: Faculty of Social and Behavioral Sciences, Section of Cognitive Psychology, Universiteit Leiden, P.O. Box 9555, 2300 RB Leiden, The Netherlands.
[b)]Electronic mail: email@evelinemaris.nl
[c)]Electronic mail: stallen@fsw.leidenuniv.nl

garded from a social psychological perspective: The sound source, being either a person or an institution operating the source, allocates a (negative) outcome (i.e., sound) to the exposed. For example, Heathrow airport has decided upon a runway operation regime of "daytime runway alternations" (i.e., using one main runway for departures and another for arrivals, and changing this segregation halfway through the day). "Alternation gives a wider distribution of noise than permanent segregated mode (without alternation), and reduces overall noise exposure for those most heavily exposed while at the same time increasing overall noise exposure for those areas around the airport that would not otherwise have been overflown" (Flindell and Witter, 1999, p. 34). Stallen captures this social relationship between the source and the exposed by the phrase "You expose Me." From social psychology it is known that the evaluation of the outcome of an allocation depends on both the *actual outcome* as well as the *fairness* of the allocation procedure (e.g., Lind and Tyler, 1988). When the allocation procedure is perceived to be fair, the subjective evaluation of the related negative outcome is more positive. For example, Folger (1977) found that boys evaluated a disappointing monetary reward less negatively when this reward was brought about by a fair procedure (when the reward was not disappointing, the fairness of the procedure had no effect on the outcome evaluation). A fair procedure (as an instrument of sound management) can therefore be expected to have a positive influence on the affective evaluation of the sound, and hence be an instrument to reduce noise annoyance. Results from the laboratory experiment described in this paper corroborate this expectation. The social psychological perspective on noise annoyance, used for the design of the experiment, will be outlined in the remainder of this introduction.

A. A social psychological model of noise annoyance

From a psychological perspective, annoyance can be considered as psychological stress (e.g., Glass and Singer, 1972). The assumption that the amount of psychological stress is not simply a reflection of the severity of the stressor is pivotal to the cognitive theory of stress and coping (e.g., Lazarus and Folkman, 1984). In this theory stress is defined as "a relationship between the person and the environment that is appraised by the person as taxing or exceeding his or her resources and as endangering his or her well-being" (Folkman, 1984, p. 840). The appraisal of a situation is theorized to take place in two phases: primary and secondary appraisal. In primary appraisal the person evaluates the significance of a specific situation with respect to well-being. An array of personal and situational factors shapes the primary appraisal.[1] When a situation is appraised as being harmful or threatening to well-being, negative emotions such as anger or fear arise, and a secondary appraisal is triggered. "In secondary appraisal, coping resources, which include physical, social, psychological, and material assets, are evaluated with respect to the demands of the situation" (Folkman, 1984, p. 842). Perceived control (known to be an important modifier of stress responses, e.g., Glass and Singer, 1972; Campbell, 1983) is in the context of this theory

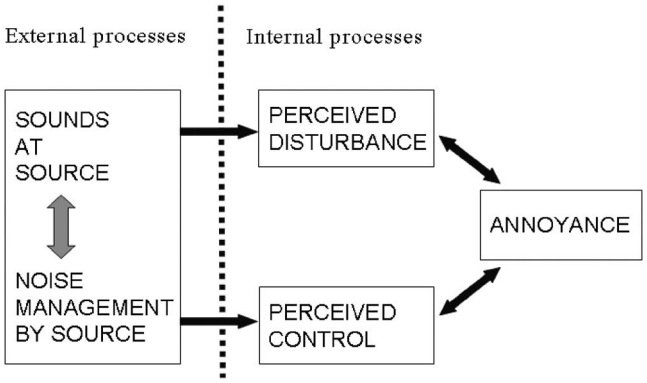


FIG. 1. Social psychological model of noise annoyance. The model, which is a simplification of the model by Stallen (1999), considers both the sound ("sounds at source") and its management ("noise management by source") as determinants of noise annoyance. The perception of these external processes results in perceptions of disturbance and/or control (internal processes). A perceived misbalance between disturbance and control results in annoyance. The model predicts that annoyance can be reduced by improving the acoustics, or by improving the sound management.

considered both as a personality trait (influencing primary appraisal) and as a situational appraisal (secondary appraisal). People continuously appraise their situation; hence the relationship between the environment and the person is dynamic (Folkman, 1984).[2]

When applied to the sound exposure situation, the cognitive theory of stress and coping predicts that a person will experience annoyance when they appraise the sound as threatening or harmful (primary appraisal), and, considering their options to cope with the sound (secondary appraisal), find that their coping resources fall short. Noise annoyance becomes a *social* problem when the sound is man-made and, consequently, a source is held responsible for the sound production. Therewith, the relationship between source and exposed becomes a relevant resource. If the exposed has, for instance, little control over the source, or little trust in the source, the perceived coping resources will be reduced and psychological stress will rise.

Stallen (1999) alludes to the social relationship between the source and the noise exposed person as a resource when he argues that the management of the sound (e.g., activities by the source ranging from keeping the sound volume within limits, to supplying residents with sound insulation or information, to asking the opinion of residents) relates to the degree of perceived control: "to a large extent perceived control is rooted in how noise is managed in practice by the source. Thus, pointing at perceived control implies pointing at another external determinant of annoyance next to sound levels: the management of sound levels. This outside stimulus is as much a stimulus for annoyance causation as the stimulus 'sound' itself" (Stallen, 1999, p. 77). Sound management by the source is, in other words, considered as a potential stressor, like noise (e.g., Evans *et al.*, 1995).

The social psychological model of noise annoyance (Stallen, 1999) is an application of stress theory on the noise exposure situation. It emphasizes the social side of noise annoyance. In short: "You expose Me." A simplification of this model is used for the design of the present study, and is depicted in Fig. 1. The model considers as codeterminants of

noise annoyance both the sound ("sounds at source") and its management by the source ("noise management by source"). The appraisal of these "external processes" results in perceptions of disturbance and control ("internal processes," reminiscent of primary and secondary appraisal). A perceived misbalance between disturbance and control results in annoyance. The model predicts that annoyance can be reduced by improving the acoustics, and/or by improving the sound management.

In the present paper, too, noise annoyance is considered an expression of psychological stress, related to the perceived adverse influence of acoustical variables (i.e., sound) and nonacoustical variables (i.e., sound management), given personal variables (e.g., basic coping capacity, perceived control as a personality trait). This description is quite in line with the World Health Organization's definition of annoyance as "a feeling of discomfort which is related to adverse influencing of an individual or a group by any substances or circumstances" (WHO, 2004, p. 3). Having outlined the social psychological perspective taken in the current study, now the earlier noise research addressing more or less explicitly sound management and the social perspective will be briefly reviewed to relate the current study to the experiments that ultimately inspired it. After that, social justice theory, used for the design of the management procedures used in the current experiment, will be introduced.

B. Preceding studies addressing social nonacoustical variables

To the best of the authors' knowledge, the potential of nonacoustical variables or sound management as an instrument to reduce noise annoyance has rarely been experimentally tested. The present authors know of four studies, which will be described in this paper. A Swedish scenario experiment (Jonsson and Sörensen, 1967) describes how participants' anticipated disturbance with sound is reduced (or aggravated) by giving them a positive (or negative) description of the noise source. A related field study reports a reduction in noise annoyance among residents of an area surrounding a Swedish Air Force base as a result of them reading positive, propaganda-like statements concerning the Air Force in a bogus questionnaire (Cederlöf *et al.*, 1967). No strong conclusions can be drawn from these data, however, as both studies suffer from methodological weaknesses.[3] In a more recent German field experiment, supplying residents with an informative telephone service reduced their annoyance, but the effect has been found only among the few residents who made use of the service (Maziul and Vogt, 2002). A fourth study has investigated the effect of citizen participation in the decision making process (i.e., selection of a sound protection barrier) on their annoyance. Results from this scenario experiment (with realistic sound samples) show no significant effect of participation on annoyance compared to a control group (ZEUS GmbH, 2002). The results of the above-mentioned studies suggest that sound management, like providing people with relevant information, may influence evaluations of noise. However, the results are inconclusive, and the theoretical underpinning of the design of the sound management procedures used is unclear.

The results of three other experiments (Glass and Singer, 1972) are illustrative of the social nature of the sound exposure situation, even though annoyance has not been assessed as a dependent variable. Glass and Singer describe a series of experiments on noise as a stressor in which the moderating effect of several cognitive factors is investigated. Dependent measures are negative after effects (e.g., performance on a proof reading and a Stroop task). One experiment investigates the effect of indirect control over the sound. Participants are, in the company of a confederate, exposed to high-intensity noise (108 dB A). Three conditions are compared: two experimental conditions and a reference condition. In the two experimental conditions the confederate is given a switch to control their noise exposure. In one experimental condition (indirect control) the participants are allowed to communicate to the confederates if they prefer the noise to be switched off. In the other experimental condition (no-indirect control) the participants are not allowed to communicate their preference. In the reference condition, neither the confederate nor the participant is given a switch. The results show that indirect control reduces the negative aftereffects of noise exposure compared to the reference condition. In the no-indirect control condition negative aftereffects increase in comparison to the reference condition, much to the researchers' surprise. Glass and Singer have explained the latter result as a serendipitous effect of *relative deprivation*[4] *of control resources*. This study nicely illustrates how sound management affects perceptions of (indirect) control over a stressor, and thus attenuates the impact of sound. Moreover, it illustrates how the perception of a difference in resource availability (i.e., access to the control switch) between two noise exposed individuals can aggravate the impact of the stressor. A follow-up of the experiment failed to replicate the presumed relative deprivation effect.

The disappearance of the relative deprivation effect may be due to a change in the verbal instructions and interactions between experimenter, confederate, and participant. Where in the initial study the experimenter instructs the confederate about the switch in a one-to-one conversation (the participant is seated behind a wooden partition and overhears them talking), in the follow-up study the experimenter includes both the participant and the confederate in the conversation. (For the original description of the two experiments, see Glass and Singer, 1972, Chap. 5) Possibly, what creates the effect in the initial study is the participant's *social exclusion*, rather than their relative deprivation of control. Another possibility is that the overt advance notice of relative deprivation in the follow-up study causes the situation to be felt as being less unfair, and hence attenuates the relative deprivation effect (Cropanzano and Randall, 1995). Yet another experiment investigates the effect of relative deprivation of exposure. The results indicate that receiving more (or less) intense sound than a comparable other aggravates (respectively, ameliorates) negative aftereffects. These three experiments described by Glass and Singer illustrate the social side of noise annoyance, namely the importance of the relative value of noise as an outcome. That is: people evaluate their outcomes relative to outcomes of comparable others.

C. Fair management procedures

The above-described experiments nicely illustrate the two major stances taken in this paper: (1) Sound management, or allocation procedure, has an influence on sound evaluation, and (2) social processes modify sound effects. This leads us to social justice theory. One of its major contributions is that effects of a negative outcome (e.g., being relatively deprived of something valuable, or receiving a lot of something unpleasant) are ameliorated when the outcome is realized by a fair procedure: *the fair process effect* (e.g., Folger, 1977; Lind and Tyler, 1988; Van den Bos *et al.*, 1997; Van den Bos and Lind, 2002).

People have a strong interest in fairness or justice (in this literature, the terms "justice" and "fairness" are used interchangeably[5]), e.g., Cohen, 1986. Being treated fairly results in a positive reaction, and the opposite situation holds too: An unfair treatment results in negative affect, protest, contraproductive behaviors, and illegal actions (e.g., Tyler, 2000). These so-called fair process effects have been found in both laboratory experiments and in field settings, as well as in a variety of situations like organizations, court trials, police-citizen encounters, and political situations (e.g., Lind and Tyler, 1988). Social justice theory will now be briefly introduced as it is the theoretical underpinning of the design of the management procedures applied in the current experiment.

Social justice theory investigates under which circumstances people consider a procedure to be fair. Studies of procedural justice judgments have identified several primary criteria that people use to evaluate fairness of procedures: (i) whether there are opportunities to participate in the decision making process ("voice"), (ii) whether the opinions of all parties involved are taken into account, (iii) whether authorities are free from bias, and whether people trust their motives, (iv) whether people are treated with dignity and respect, (v) whether the information used to come to the decision is accurate and relevant, (vi) whether the provided information about the process and the decision is clear and appropriate, and (vii) whether procedures are applied consistently across people and across time (e.g., Tyler, 2000; Greenberg, 1993; Steensma and Doreleijers, 2003; Steensma and Otto, 2000; for a concise review and meta analysis of 25 years of justice research, see Colquitt *et al.*, 2001).

The participation criterion (or voice) of procedural justice is the most often studied criterion of the above-presented list, and is also used as the fairness manipulation in the current experiment. Effects of voice on procedural fairness judgments are strongest when the voice is given before the decision is made ("predecision voice"). When the voice is given afterwards, it still enhances fairness judgment ("postdecision voice") (Lind *et al.*, 1990). Thibaut and Walker (1975) refer to this distinction as "instrumental" voice, in which people's comments may influence the decision, and "noninstrumental" voice, in which the comments will have no bearing on the outcome (e.g., comments are only allowed after the decision had been made). "Mediation analyses showed that perceptions of control account for some, but not all, of the voice-based enhancement of procedural justice" (Lind *et al.*, 1990, p. 952). "People have also been found to value the opportunity to express their views to decision-makers in situations in which they believe that what they are saying has little or no influence upon the decisions being made (…) People are primarily interested in sharing the discussion over the issues involved in their problem or conflict, not in controlling decisions about how to handle it." (Tyler, 2000, pp. 121–122). Consequently, giving people voice with regard to their sound exposure situation will increase the perceived fairness of the management procedure (even when this voice is noninstrumental), and may result in a more positive reaction toward the sound and less psychological stress.

The required link between psychological stress and the fairness of outcome distribution and allocation procedures has been explored within an organizational context (Tepper, 2001; Vermunt and Steensma, 2001, 2003, 2005). A framework that integrates the cognitive theory of stress and coping (e.g., Lazarus and Folkman, 1984) with social justice theory (e.g., Lind and Tyler, 1988) has been proposed (Tepper, 2001). The perceived fairness of distributions and procedures is hypothesized to influence the primary and secondary appraisal of the situation, and hence affect psychological stress. In a work environment, managers distribute demands and resources among their subordinates. When a subordinate perceives a discrepancy between these demands and available resources, they may experience stress, in line with the cognitive theory of stress and coping. Tepper (2001) has found that the perceived fairness of outcomes as well as allocation procedures correlates negatively with psychological stress. Vermunt and Steensma (2001, 2003, 2005) have theorized and shown that fair procedures can be used to reduce stress, and conclude that a fair treatment reduces the threat value of an event. An instrumental explanation for the relation between fairness and stress is that fair procedures offer opportunities for process control (i.e., the opportunity to present information or evidence as input into the decision) and decision control (i.e, the opportunity to influence the decision itself) which increases the likelihood of receiving favorable outcomes. A noninstrumental explanation holds that people care about procedural justice because it provides feedback regarding their status in the group or community: A high status provides the group member with two vital coping resources: a social support system and a sense of self-efficacy (Tepper, 2001; Tyler and Lind, 1992).

In social justice literature, generally additive main effects of procedural and distributive justice on outcome satisfaction are reported (Lind and Tyler, 1988, pp. 68–69). In his study on the relationship between procedural fairness and stress, Tepper (2001) has found that the effects of distributive fairness and procedural fairness on stress interact: The effect of procedural fairness is far stronger when the distribution is unfair. Sometimes, procedures only have an effect when outcomes are unfair. Tepper (2001) argues that, since the secondary appraisal (in which the procedure is evaluated) is triggered by perceived harm or threat (i.e., distributive unfairness, primary appraisal), it is the distributive fairness that moderates the effect of procedural fairness. Vermunt and Steensma (2003) have found that the effect of procedural fairness depends on the level of stress a person is actually

experiencing. However, solely based on empirical data the opposing explanation, that procedures moderate the effect of the distribution, is also plausible.

D. Hypotheses

In the present experiment effects of two sound pressure levels (SPL), low and high ("sounds at source"), and two exposure procedures, neutral and fair ("noise management by source"), on noise annoyance are compared. Participants are exposed to interfering sound (low or high SPL) while working on a task. It is assumed that the low SPL will induce lower stress than the high SPL will; hence it is hypothesized that the noise annoyance ratings will be lower in the low SPL condition than in the high SPL condition (Hypothesis 1). Within each SPL condition and before exposure, half of the participants are given noninstrumental voice (fair procedure). The other half do not receive voice (neutral procedure). Based on fairness research (e.g. Tepper, 2001; Folger, 1977; Lind and Tyler, 1988, pp. 68–69, 72), it is predicted that a *fair process effect* will be present only when the sound is appraised as harmful or threatening (i.e., disturbing sound). For most people, this will be the case in the high SPL condition, and not or to a lesser extent in the low SPL condition. Specifically, we predict that, within the high SPL condition, participants will report lower annoyance in the fair procedure condition than in the neutral procedure condition, while this effect of procedure is absent or less strong in the low SPL condition (Hypothesis 2).

II. METHOD

A. Participants

One hundred and seventeen students (75% female; mean age 22 years) are paid 4 Euro each to participate in the experiment, which lasts approximately 45 min. Participants are randomly and evenly spread over the four cells of the experimental design.

B. Experimental design

The experimental design is a 2 (procedure: fair versus neutral procedure) $\times 2$ (sound pressure level (SPL): low (50 dB) versus high (70 dB) complete factorial design.

C. Laboratory layout and stimulus material

The laboratory consists of four separate cubicles, each of which contains a desk and chair, and a complete PC set with two loudspeakers plus one subwoofer.

The two sound samples are composed of self-recorded audio material of aircraft passages of various loudness and duration.[6] The "hearing test" sample is a 1 min sound sample of a single aircraft passage, and is played at 60 dB A (1 min Leq). The experimental sample lasts 15 min, during which, at random intervals, 11 aircraft passages are audible. The experimental sample is played at either 50 dB A (15 min LAeq) (low SPL condition) or 70 dB A (15 min LAeq) (high SPL condition), which implies a sound level of quiet background noise in the low condition, or of speech interfering loudness in the high condition. The maximal sound pressure level of the experimental sample is 68 or 88 dB A Lmax, respectively. All sound pressure levels are measured in the cubicle at the position of the listener.

The reading task (an English text with multiple choice questions, taken from a Dutch exam from preuniversity education) is selected to match the cover story, and because it assures participants' motivation to perform well, and to closely match their capacities. With too easy a task, the experimental noise may not cause any disturbance and hence not induce any annoyance, whereas too difficult a task may give rise to performance effects, which may cloud the effects of the procedure and/or the SPL manipulation (Smith, 1989).

D. Experimental procedure and manipulations

Upon their arrival at the laboratory, participants check with the experimenter who obtains their informed consent and guides them to their cubicle. After being seated, participants are left to themselves. All further communication is through the computer, which is used for the presentation of the stimulus information and for the recording of the dependent variables. Participants are told (on screen) that they are engaged in a study on effects of sound on people's performance during high school exams. As part of the experiment, they will take an exam while being exposed to possibly interfering sound. Then, a bogus hearing test is administered: The "hearing test" sample is played and participants are asked to judge how loud and how annoying this sample is to them. In fact, this test gives them a frame of reference for loudness, assuring that they will later on experience the experimental sound sample either as softer or as louder than the reference sample. The test provides a baseline measure for annoyance as well, to be used as a dummy for basic coping capacity (covariate in the analyses). Next, participants are led to believe that the experimenter intends to compare three different sound samples which are equal in length, but differing in number and duration of the aircraft passages. The three concocted samples are described as: (A) many, but short lasting aircraft passages; (B) few, but longer lasting aircraft passages; or (C) aircraft passages of intermediate number and duration. Participants are led to believe they are to listen to one of these three samples during the task. It is suggested that they may have a personal preference for one sound sample over another. Participants in the *fair procedure condition* are asked to express (voice) their personal preference for one of the three samples. A confirmation of their expressed preference is given, and the experimenter states that they will take this preference into account as much as possible (literal text, on screen, in a small font: "You have indicated that you expect sample X to cause you the least annoyance", and then in a larger, bold font: "We will take this into account as much as possible"). Participants in the fair procedure condition then are asked to confirm their preference. The experimenter states once more that they will take this preference into account as much as possible. Participants in the *neutral procedure condition* are not asked to voice any preference. They are informed that the experimenter will select one of the three samples for them. After the procedure manipulation, all participants start with the exam (reading

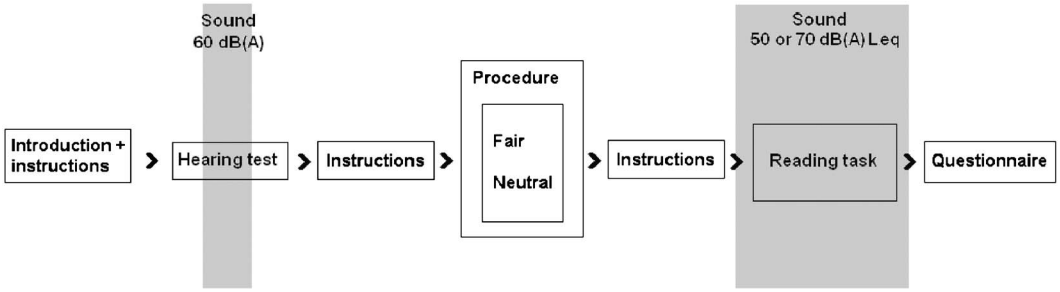


FIG. 2. Visual representation of the flow of the experiment.

task) while being exposed to the experimental sound sample. The sample is identical for all subjects and ambiguous with respect to the relative number and duration of the passages. It is played at either 50 or 70 dB, depending on the participant's sound pressure level condition. After 15 min of exposure, the experimental sound and the task are terminated (none of the participants have by then finished the task). Participants are then presented with the questionnaire that assesses the dependent measures and the manipulation checks. (See Fig. 2 for a visual representation of the flow of the experiment).

E. Measures

The baseline measure for annoyance (covariate) is a single-item question: "How annoying was the aircraft sound you have just listened to, to you?" Answers are given on a 7-point Likert scale with verbal markers at each end point: 1="not annoying at all," 7="highly annoying" [$M(\text{s.d.})=4.38(1.40)$].

In the questionnaire, three questions assess annoyance with the experimental sound: (i) "To what extent did the sound annoy you while you were working at the task?" (ii) "How did you experience the aircraft sound while answering the exam questions?" (iii) "How pleasant did you feel the aircraft sound was while working on the exam?" Answers are given on a 7-point Likert scale with verbal markers at each end point: (i) 1="not at all annoying," 7="highly annoying," (ii) 1="very positive," 7="very negative," (iii) 1="very pleasant," 7="very unpleasant." An annoyance scale is constructed from the three items (Cronbach's $\alpha=0.76$). The mean annoyance score of the scale [$M(\text{s.d.})=5.21$ (1.03)] is significantly different from 4 (neutral score) $t(116)=12.71, p<0.001$], indicating that, on average, participants consider the sound to be annoying.

One question checks the effectiveness of the sound pressure level manipulation: "If you were to give a grade for the average loudness of the aircraft sound, what grade would you give?" In the instructions, verbal labels are given to the end points of the scale: 1="very soft," 10="very loud." Participants respond by clicking on the virtual button numbered with the grade of their choice [$M(\text{s.d.})=6.38(1.91)$]. (The measure for the hearing test is identical to this manipulation check for sound pressure level.) A 10-point scale is used (deviant from the annoyance measure, which uses a 7-point scale) to prevent participants from ticking the exact same number on the annoyance measure and the loudness measure, aiming to give a consistent (socially desirable) rather than a faithful answer. The Pearson's correlation between perceived loudness and annoyance ($r=0.36$, $p<0.001$, $N=117$) indicates that participants regard loudness and annoyance as two related but conceptually different concepts.

Five items check the effectiveness and fairness of the procedure manipulation: (i) "The experimenters sought to take my preference for a certain combination of sound characteristics into account," (ii) "In my opinion, the procedure that was applied to select my sample is…," (iii) "In your opinion, how fair is it that the participants were not all given the same sample?," (iv) "I was given the sample of my preference," and (v) "The experimenters have made an effort to tax the participants as little as possible." Answers are given on a 7-point Likert scale with verbal markers at each end point: 1="completely disagree" and 7="completely agree" for items i, iv and v, and 1="very unfair" and 7="very fair" for items ii and iii. A "don't know" option is included and scored as a missing datum. A perceived procedural fairness scale is constructed from the five items, excluding the missing values [$N=83$, $M(\text{s.d.})=4.73$ (1.11), Cronbach's $\alpha=0.64$], and including the missing values replaced by the series' mean [$N=117$, $M(\text{s.d.})=4.62$ (1.01), Cronbach's $\alpha=0.60$].

Finally, some general questions [e.g., gender, self-reported hearing impairments ("Do you have any hearing impairments?," response categories: (i) "yes" (ii) "somewhat" and (iii) "no"] are included.

Besides, some explorative measures of task performance, to be used as a check for unintended performance effects, are automatically registered by the computer ("time:" time taken to read the first text and answer question one; "correct:" total number of correct answers; "false:" total number of false answers).

III. RESULTS

A. Manipulation checks

1. Perceived loudness

Analysis of variance (ANOVA) with perceived loudness as the dependent variable and sound pressure level (SPL) and procedure as the independent variables indicates that the sound pressure level manipulation was successful: Participants in the low conditions experience the sound to be significantly less loud than participants in the high conditions [$M_{\text{low}}(\text{s.d.})=5.24$ (1.59) vs $M_{\text{high}}(\text{s.d.})=7.49$ (1.50), $F(1,113)=61.11$, $p<0.001$; an alpha level of 0.05 is used for

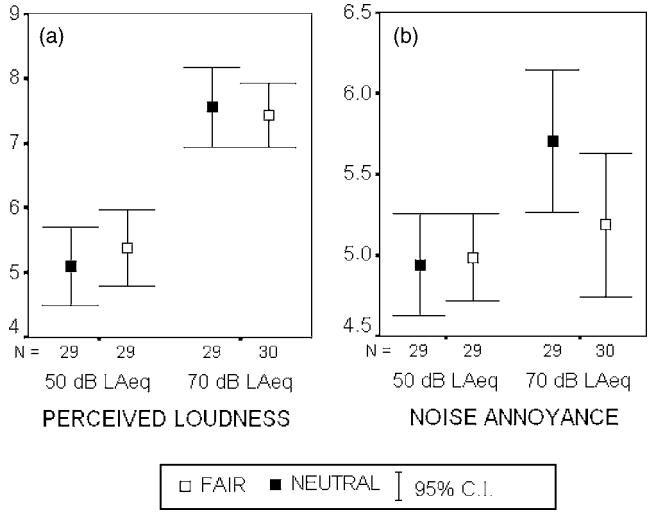


FIG. 3. Perceptions of loudness (left) of and annoyance (right) with the experimental sound. Means (dots) and 95%-confidence intervals (whiskers) are shown. Whereas the perceived loudness of the experimental sound is not influenced by procedure, the evaluation of the sound in terms of annoyance is: Within the high sound condition, a fair procedure results in a reduction of reported noise annoyance.

all statistical analyses]. The SPL manipulation is independent of procedure: Perceived loudness is not influenced by procedure ($M_{\text{fair}}(\text{s.d.})=6.42$ (1.78) vs $M_{\text{neutral}}(\text{s.d.})=6.33$ (2.05), $F(1,113)=0.08, p=0.79$, *n.s.*), nor is there an interaction effect of SPL and procedure on perceived loudness ($F(1,113)=0.47, p=0.50$ *n.s.*) (See Fig. 3).

2. Perceived fairness

ANOVA with perceived fairness (excluding missing values, $N=83$) as the dependent variable and SPL and procedure as the independent variables indicates that the fairness manipulation was successful: Participants in the fair condition experience the procedure to be significantly more fair than participants in the neutral condition [$M_{\text{fair}}(\text{s.d.})=5.09$ (1.11) versus $M_{\text{neutral}}(\text{s.d.})=4.30$ (0.96), $F(1,79)=10.93$, $p<0.002$]. The procedure manipulation is independent of SPL: Perceived fairness is not influenced by SPL ($M_{\text{low}}(\text{s.d.})=4.85$ (1.14) versus $M_{\text{high}}(\text{s.d.})=4.58$ (1.07), $F(1,79)=0.94, p=0.34$, *n.s.*, nor is there an interaction effect of SPL and procedure on perceived fairness ($F(1,79)=0.94, p=0.34$, *n.s.*).

ANOVA with perceived fairness (missing values replaced with the series mean, $N=117$) as the dependent variable and SPL and procedure as the independent variables gives similar results [Effect of procedure: $M_{\text{fair}}(\text{s.d.})=4.89$ (1.07) vs $M_{\text{neutral}}(\text{s.d.})=4.34$ (0.86), $F(1,113)=9.51$, $p<0.01$; effect of SPL: $M_{\text{low}}(\text{s.d.})=4.75$ (1.07) vs $M_{\text{high}}(\text{s.d.})=4.49$ (0.94), $F(1,113)=2.04, p=0.16$, *n.s.*; interaction effect of SPL and Procedure: $F(1,113)=1.94, p=0.17$, *n.s.*].

3. Performance measures

The three checks for unintended performance effects indicate no systematic differences between the four conditions in performance on the reading task. The mean time spent on task item 1 is about 90 s, and no differences between conditions are found [$M_{\text{time(s)}}(\text{s.d.})=96.62$ (37.49), $F(3.113)=1.62$, $p=0.19$, *n.s.*]. On average, participants answer ten task items correctly and four task items falsely, and no differences between conditions are found $M_{\text{correct}}(\text{s.d.})=10.05$ (4.03), $F(3.113)=1.80$, $p=0.15$, *n.s.*; $M_{\text{false}}(\text{s.d.})=4.37$ (2.77), $F(3.113)=0.76$, $p=0.52$, *n.s.*].

TABLE I. Annoyance scores (1="not annoyed at all," 7="highly annoyed") arranged by conditions of sound pressure level (low of high) and procedure (neutral or fair). Estimated marginal means and cell means (M), standard deviations (s.d.), and number of cases per cell (N).

Sound pressure level	Procedure	M	s.d.	N
Low	Neutral	4.94	0.83	29
50 dB	Fair	4.99	0.70	29
	Total low	4.97	0.76	58
High	Neutral	5.70	1.16	29
70 dB	Fair	5.19	1.19	30
	Total high	5.44	1.19	59
Total	Neutral	5.32	1.07	58
	Fair	5.09	0.98	59
	Total	5.21	1.03	117

B. Dependent variables

Analysis of Covariance (ANCOVA) with annoyance as the dependent variable, SPL and procedure as the independent variables, and baseline annoyance as the covariate shows a significant main effect of SPL, $F(1,112)=11.25$, $p<0.005$, $\eta^2=0.09$ (see Table I for marginal means). Participants who are exposed to low sound express less annoyance than those receiving high sound (this finding confirms Hypothesis 1). No significant main effect of procedure is found [$F(1,112)=2.39$, $p=0.13$, *n.s.*]. The interaction effect of SPL by procedure is significant at the $p<0.05$ level [$F(1,112)=3.95$, $\eta^2=0.03$, see Table I for cell means, and Fig. 3 for a visual representation of these results] and indicates a *fair process effect* in the high sound condition (confirmation of Hypothesis 2). If the SPL is high, the fair procedure condition yields lower levels of annoyance than the neutral procedure condition does. Within the High SPL condition, the fair process effect explains 9% of the variance in annoyance scores. Within the low SPL condition, variance in annoyance scores cannot be attributed to procedure conditions.

Hearing impairments. Several participants indicate having hearing impairments ("yes:" $N=2$, "somewhat:" $N=12$). When these cases are left out of the analysis, the pattern and significance level of the effects remains largely the same [main effect of SPL: $F(1,98)=7.79$, $p<0.01$, $\eta^2=0.07$; main effect of procedure: $F(1,98)=1.53$, $p=0.22$, *n.s.*; interaction effect of SPL by Procedure: $F(1.98)=6.65$, $p<0.05$, $\eta^2=0.06$].

IV. CONCLUSIONS

Despite the definition of noise as a *social* problem (Cohen and Spacapan, 1984), noise annoyance has rarely been

studied from a social perspective, that is: addressing the social relationship between the source and the noise exposed person(s). In the present laboratory experiment, the source relates to the exposed persons by allocating a negative outcome (i.e., disturbing sound) to them. The exposed persons are dependent on the source, as the latter has control over the stressor ("You expose Me," Stallen, 1999). The fairness of the allocation procedure applied by the source is manipulated. The results confirm that characteristics of the social relationship (or allocation procedure) codetermine noise annoyance: when sound levels are high, a fair procedure can reduce annoyance.

The model used for the design of the experiment (see Fig. 1) incorporates "sound" in combination with "sound management by the source" as substantive determinants of annoyance. Both the sound and its management influence perceptions of disturbance (threat or harm) and opportunities for control (resources), which can result in annoyance (psychological stress). Bad management can be a nuisance in itself; good management can be a coping resource (e.g., by providing a feeling of trust). In the experiment, the annoying effects of two sound pressure levels (50 and 70 dB aircraft sound) and two management alternatives (a fair and a neutral procedure) are compared. The reported findings corroborate the argument that the perception of both the sound and its management influence noise annoyance. It has been found that participants who are exposed to rather disturbing sound (i.e., 15 min of 70 dB aircraft noise) report significantly less noise annoyance when they have been given voice over their noise exposure than participants who are exposed to the same sound but who have not been given voice (see Table I and Fig. 3). This *fair process effect* is of considerable strength: Within the 70 dB SPL condition, the fair procedure reduces the mean annoyance level to approximate the mean level in the 50 dB SPL condition. Up to 9% of variance in annoyance scores can be explained by the procedure manipulation (see Sec. III B). When participants are exposed to less disturbing sound (i.e., 50 dB aircraft noise) no fair process effect is found (see Table I and Fig. 3). The present study demonstrates that, under laboratory conditions and with rather disturbing sound, a fair sound management can be used as a nonacoustic instrument to reduce annoyance.

The current study adds on the numerous survey studies (e.g., Fields, 1993; Job, 1988) that address nonacoustical variables and noise annoyance in four ways. First, the current study offers a theoretical framework that describes how an external nonacoustical variable (i.e., the allocation procedure) can be a codeterminant of noise annoyance. Second, social justice theory identifies characteristics of the allocation procedure that may ameliorate sound evaluations and are tractable for policymakers. Nonacoustical sources of variance are thus redefined into a potentially useful instrument for influencing annoyance. Third, being an experiment, the study allows for more firm conclusions regarding the direction of causality between social nonacoustical variables and annoyance than a survey study would. Fourth, the study shows that the shape and/or the position of the dosage-response curve (for sounds from the same source) can vary depending on identifiable aspects of the social context of exposure. Specifically, the dosage-response curve of the current data shows a far steeper slope in the neutral situation than in the fair situation, while standard deviations are found to be equal in both situations (see Fig. 3). This suggests that the substantial variance in annoyance scores, typical of any dosage-response plot, can be attributed not only to (internal) personal predispositions, but also to systematic external factors of social nature, which are tractable on a collective level.

The current results corroborate the argument made by Stallen (1999) that in order to fully understand noise annoyance, the sound management has to be considered as a stimulus beside the sound. An important question that remains to be resolved, is whether sound management operates as a determinant of annoyance, or as a moderator of sound effects. Based on the justice literature, it can be argued that the sound level moderates the effects of sound management (e.g., Folger, 1977; Tepper, 2001). In the noise annoyance literature, several studies have found an interaction effect of SPL and nonacoustic variables (e.g., Maris *et al.*, 2004a, b; Schümer, 1974; Schümer-Kohrs and Schümer, 1974; Fields and Walker, 1982; Fidell *et al.*, 2002; Miedema and Vos, 2003). However, a large study on noise sensitivity has not found an interaction effect (Van Kamp *et al.*, 2004). The larger part of this data suggests a multiplicative model. Some authors assume a variable threshold (dependent on nonacoustic moderator variables) above which people start to consider sound as noise (e.g., Schümer, 1974; Fidell *et al.*, 2002; see also Dubois (2000) for a semantics perspective on the distinction between sound and noise). In this regard, it is interesting to recollect that in the current experiment the procedure affects the evaluation of the sound, but not the perception of its loudness [see Sec. III A and Fig. 3; see also Janssen *et al.* (2004) for similar results in pain research where reduced control over the pain affected the willingness to endure the pain, but not the sensory experience of it].

In a preceding study with the same paradigm (Maris *et al.*, 2004a, 2004b), annoyance was assessed twice: after 1 min and after 15 min of exposure. The results indicated that the effect of SPL on noise annoyance was immediate and significantly lost strength over time, whereas the effect of procedure grew over time to a significant interaction after 15 min. It is interpreted that the consideration of social factors, like the procedure, (1) is more likely when considerably disturbing sound levels are perceived, and (2) takes place somewhat later in the psychological process than does the perception of the sound itself. This alludes to a model of noise annoyance that includes attentional processes (e.g., secondary appraisal) as a moderator of procedure effects (e.g., Botteldooren *et al.*, 2004; Ulrich, 1983).

Some issues of validity have to be addressed. The data underlying the current findings have been gathered among students only. This may restrict generalizations of the current findings to the general public. Students are, on average, younger and more highly educated than the general public. In combination, these characteristics may cause students to have a higher need for autonomy, which, in turn, may make them more sensitive to (not) having voice, compared to the

general public (Avery and Quiñones, 2004). Notwithstanding, in the general public substantial variation in need for autonomy will be present, and hence the current findings will likely apply to a significant proportion of the general public.

Generalizations from the current findings to situations outside the lab may be restricted in other respects. First situational and individual differences may influence the value of fair procedures. It is known that people value voice more when the situation is uncertain, or when trust in authorities is low. Personality differences [e.g., Big Five, belief in a just world (Vermunt and Steensma, 2005), noise sensitivity] and attitudes (e.g., attitude toward the source) influence whether or not a situation is perceived as disturbing, and hence whether coping resources are appraised.

Second, in the lab, participants are well aware that their exposure will not last longer than the course of the experiment (i.e., 15 min). They are participating voluntarily and can terminate their participation at any desired moment, plus they are financially compensated for their discomfort. These aspects may make the participants care less about the sound. However, the mean reported annoyance level indicates that the participants have not been indifferent to the sound, and the fact that, in the lab, an effect of procedure is found, bolsters rather than weakens the importance of fair procedures. The aim of the current study is to show, theoretically and experimentally, that social nonacoustical variables play a crucial role in the psychology of noise annoyance. Although it cannot simply be assumed that the psychology of annoyance will be identical in the field, results from survey studies confirm that social variables like trust and attitudes toward the source play a significant role (Guski, 1999).

Third, it may be objected that an instance of fair noise management is capable of drawing people's attention in the confined reality of an experiment, but may easily go unnoticed in real life where an excessive number of stimuli and cognitions compete for attention. Consequently, outside of the lab, the effect of a single instance of fair sound management may be attenuated, should these not be reinforced by the continuous public debates, protests, media attention, and policy processes regarding aircraft noise and its management. Together, these sociological processes may exert a significant influence on people's ideas about the fairness of sound management (Bröer, 2002; Wirth and Bröer, 2004). In sum, it is important to consider differences between lab and field, but it seems warranted to make careful generalizations. Finally, experiences with, e.g., community consultation and transparent communication around Heathrow airport (Flindell and Witter, 1999) and Sydney airport [D. Southgate, 2002], illustrate the practical value of fair noise management.

With regard to the present sound manipulation, some remarks need to be made. The recording and play back of the sound may not have created a state-of-the-art soundscape due to the unpretentious techniques used. However, it is not likely that a sound quality issue will endanger the conclusions drawn from the data. First, research has indicated that the cognitive responses to source events (other than to background sound where no source is easily identifiable) are rather robust to charges in sound reproduction method (Guastavino *et al.*, 2005). Second, the sound quality has been identical for all participants, ruling out the possibility that the procedure effects found are due to artifacts related to sound quality differences. Indeed, with regard to sound quality and exposure duration, a sound experience in the lab is different from outdoor situations. But even though this may influence the relative strength of the various processes within the psychological model of noise annoyance, the authors have no reason to expect that the psychological model of noise annoyance itself will be essentially different inside or outside of the lab.

Knowledge of the social determinants of noise annoyance will be of growing importance. On the one hand, decibel levels are expected to increase due to the increasing mobility of increasing numbers of people. On the other hand, a changing noise situation implies a lot of negotiation and allocation decisions, and is usually associated with increased annoyance levels (Fidell *et al.*, 2002). Thorough knowledge of the social processes that codetermine noise annoyance is needed to keep annoyance from nonacoustic sources as low as possible. Application of fair procedures in sound management is a promising instrument for annoyance reduction (e.g., Vermunt and Steensma, 2001, 2003, 2005), but some caution should be taken to prevent a reversal of the fair process effect (Van den Bos *et al.*, 1999). The influence of a variety of criteria of fair procedures has to be studied within a noise context, and their application and effectiveness in the field have to be explored.

Given the general WHO definition of annoyance, the framework suggested here for noise annoyance may also be applicable for annoyance with other man-made substances. For instance, studies on odor annoyance (Matthies *et al.*, 2000) and urban nuisances like incivilities (Robin *et al.*, 2004) signal the importance of a social perspective on annoyance.

Some caution is warranted, however. A reduction of self-reported annoyance does not necessarily indicate less bother (or increased well being). Several studies point out that a reduction of reported annoyance can also indicate that people suppress their annoyance (Fields and Walker, 1982), or compensate by adjusting their aims (Staples, 1997; Tafalla *et al.*, 1988). One study reports a negative correlation between expressed annoyance and physiological stress levels, suggesting that a suppressed expression of annoyance results in an increase of physiological stress (Miyakawa *et al.*, 2004).

If future noise annoyance levels are to be kept to a minimum, it is needed that, in addition to the important and impressive developments that are being made in the field of noise reduction engineering, both noise researchers and policy makers address social nonacoustic codeterminants of noise annoyance.

ACKNOWLEDGMENTS

This research was financed by the Platform Nederlandse Luchtvaart (PNL, Platform Dutch Aviation), The Netherlands. The authors would like to thank four people in particular for their useful comments on earlier drafts of this paper: Wokje Abrahamse, Geertje Schuitema, and two anonymous reviewers.

[1]Among the most important personal factors thought to shape primary appraisal are *beliefs* [ranging from generalized (e.g., religious) beliefs to specific beliefs (e.g., personal control over important outcomes)] and *commitments* [ranging from values and ideals (e.g., care for the environment) to specific goals (e.g, living in a quiet neighborhood)]. Situational factors include the nature of the threat or harm, the familiarity, novelty, and likelihood of occurrence of the event, and the ambiguity of the expected outcome (Folkman, 1984, pp. 841–842). The authors regard perceived loudness, disturbance and pleasantness of the sound as personal or situational factors.

[2]The effect of the continuous appraisal of situations is that, in practice, primary and secondary appraisal are parallel processes which are difficult to discriminate. Experiences in previous encounters may alter people's beliefs and goals, and consequently have an impact on the appraisal of new encounters. Nonacoustical variables like attitudes may be a result of one encounter, and the modifier of another. A fair sound management procedure may be evaluated as a coping resource in one situation (and influence the secondary appraisal of that situation), and at the same time changes people's attitudes toward the source (influencing the primary appraisal of a subsequent encounter with the sound).

[3]Caution in the interpretation of these data is warranted. In the first study (Jonsson and Sörensen, 1967) anticipated disturbance is measured with a 4-point verbal scale consisting of items that refer more to attitude toward the source than to disturbance with the sound. It is questionable which concept was measured. In the second study (Cederlöf *et al.*, 1967) the population studied showed unusually high prelevels of annoyance (due to exceptionally high exposure levels). It is therefore unclear whether any effect would have been found under normal starting conditions. Moreover, the design of the study may have made individuals in the experimental condition sensitive to give socially desirable answers.

[4]Relative deprivation: A situation in which someone compares their situation to that of a relevant reference peer, and feels discontented with it, not because of the situation itself, but because the peer seems better off.

[5]It should be noted that the semantics of the concept "justice" or "fairness" are context dependent. The concept has been studied by philosophers for centuries, if not millennia, and in a whole range of scientific disciplines (e.g., political sciences, anthropology, sociology, computer sciences) research on fairness (or justice) is conducted. "In contrast to other disciplines, social psychology does not take a normative approach [to justice]. It deals with justice in a descriptive rather than a prescriptive way. The aim is not to define what is just and unjust, and how justice can be achieved. The focus on the contrary is on the subjective sense of justice and injustice and its impact on human action and judgment. Social psychologists study what people regard as just and unjust under given circumstances, how people deal with the concept of justice, how they react to situations that they regard as unjust, and under which circumstances, and why, people care about justice" (Mikula, 2001, pp. 8063–8064). In the current paper, the following definition of fair (or just) procedures is operated: procedures that people judge to be fair. Although it is very likely that substantial cultural differences exist with regard to which procedures people regard as just, the wish to be treated in a just way appears to be an anthropological universal (Montada, 2001).

[6]The audio material was recorded outdoors, with clear weather conditions, on one location in the vicinity of a runway in use for landings only. Ambient sounds were removed from the recordings by a professional company. The 15 min experimental sample was made up of 11 noise events of aircraft passages of various loudness, duration, and aircraft type. The quiet time dispersing two passages (1 min, on average) was shorter than in real life and of variable length.

Adams, J. (**1981**). "What noise annoys?," in *Transport Planning. Vision and Practice*, edited by J. Adams (Routledge, London), pp. 254–258.

Avery, D. R., and Quiñones, M. A. (**2004**). "Individual differences and the voice effect," Group Org Management, **29**, 106–124.

Botteldooren, D., Lercher, P., and De Muer, T. (**2004**). "The influence of active coping on the adverse effects of noise," Internoise 2004, Prague, Czech Republic, 22–25 August.

Bröer, C. (**2002**). "Sound, meaning and politics. The social construction of noise annoyance," Forum Acusticum, Sevilla, Spain, 16–20 September.

Campbell, J. M. (**1983**). "Ambient stressors," Environ. Behav. **15**, 355–378.

Cederlöf, R., Jonsson, E., and Sörenson, S. (**1967**). "On the influence of attitudes to the source on annoyance reactions to noise. A field experiment," Scand. J. Work Environ. Health **48**, 46–59.

Cohen, R. L. (**1986**). *Justice: Views from Social Sciences* (Plenum, New York).

Cohen, S., and Spacapan, S. (**1984**). "The social psychology of noise," in *Noise and Society*, edited by D. M. Jones and A. J. Chapman (Wiley, Chichester), pp. 221–245.

Colquitt, J. A., Conlon, D. E., Wesson, W. J., Porter, C. O. L. H., and Ng, K. Y. (**2001**). "Justice at the millennium: A meta-analytic review of 25 years of organizational justice research," J. Appl. Psychol. **86**, 425–445.

Cropanzano, R., and Randall, M. L. (**1995**). "Advance notice as a means of reducing relative deprivation," Soc. Justice Res. **8**, 217–238.

Dubois, D. (**2000**). "Categories as acts of meaning: The case of categories in olfaction and audition," Cognit. Sci. Q. **1**, 35–68.

Evans, G. W., Hygge, S., and Bullinger, M. (**1995**). "Chronic noise and psychological stress," Psychol. Sci. **6**, 333–338.

Fidell, S., Schultz, T., and Green, D. M. (**1988**). "A theoretical interpretation of the prevalence of noise-induced annoyance in residential populations," J. Acoust. Soc. Am. **84**, 2109–2113.

Fidell, S., Silvati, L., and Haboly, E. (**2002**). "Social survey of community response to a step change in aircraft noise exposure," J. Acoust. Soc. Am. **111**, 200–209.

Fields, J. M. (**1993**). "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763.

Fields, J. M., and Walker, J. G. (**1982**). "The response to railway noise in residential areas in Great Britain," J. Sound Vib. **85**, 177–255.

Flindell, I. H., and Witter, I. J. (**1999**). "Non-acoustical factors in noise management at Heathrow Airport," Noise Health **3**, 27–44.

Folger, R. (**1977**). "Distributive and procedural justice: Combined impact of 'voice' and improvement of experienced inequity," J. Pers Soc. Psychol. **35**, 108–119.

Folkman, S. (**1984**). "Personal control and stress and coping processes: A theoretical analysis," J. Pers Soc. Psychol. **46**, 839–852.

Glass, D. C., and Singer, J. E. (**1972**). *Urban Stress. Experiments on Noise and Social Stressors* (Academic, New York).

Green, D. M., and Fidell, S. (**1991**). "Variability in the criterion for reporting annoyance in community noise surveys," J. Acoust. Soc. Am. **89**, 234–243.

Greenberg, J. (**1993**). "The social side of fairness. Interpersonal and informational classes of organizational justice," in *Justice in the Workplace. Approaching Fairness in Human Resource Management*, edited by R. Cropanzano (Erlbaum, Hillsdale, NJ), pp. 79–103.

Guski, R. (**1999**). "Personal and social variables as co-determinants of noise annoyance," Noise Health **3**, 45–56.

Guski, R. (**2001**). "Environmental stress and health," in *International Encyclopedia of the Social and Behavioural Sciences*, edited by N. J. Smelser and P. B. Baltes (Elsevier, Oxford, UK), Vol. **7**, pp. 4667–4671.

Guastavino, C., Katz, B. F. G., Polack, J.-D., Levitin, D. J., and Dubois, D. (**2005**). "Ecological validity of soundscape reproduction," Acta Acust. **91**, 333–341.

Janssen, S. A., Spinhoven, P., and Arntz, A. (**2004**). "The effect of failing to control pain: An experimental investigation," Pain **107**, 227–233.

Job, R. F. S. (**1988**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.

Jonsson, E., and Sörensen, S. (**1967**). "On the influence of attitudes to the source on annoyance reactions to noise. An experimental study," Nord. Hyg. Tidskr **48**, 35–45.

Lazarus, R. S., and Folkman, S. (**1984**). *Stress, Appraisal and Coping* (Springer, New York).

Lind, E. A., Kanfer, R., and Earley, P. C. (**1990**). "Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments," J. Pers Soc. Psychol. **59**, 952–959.

Lind, E. A., and Tyler, T. R. (**1988**). *The Social Psychology of Procedural Justice* (Plenum, New York).

Maris, E., Stallen, P. J. M., Steensma, H., and Vermunt, R. (**2004a**). "The influence of procedural fairness on evaluations of noise," IAPS 2004, Vienna, Austria, 7–10 July.

Maris, E., Stallen, P. J. M., Steensma, H., and Vermunt, R. (**2004b**). "The influence of procedural fairness on evaluations of noise," Internoise 2004, Prague, Czech Republic, 22–25 August.

Matthies, E., Höger, R., and Guski, R. (**2000**). "Living on polluted soil. Determinants of stress symptoms," Environ. Behav. **32**, 207–286.

Maziul, M., and Vogt, J. (**2002**). "Can a telephone service reduce annoyance?," 43rd Conference of the Deutsche Gesellschaft für Psychologie, Berlin, Germany, 22–26 September.

Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.
Miedema, H. M. E., and Vos, H. (**2003**). "Noise sensitivity and reaction to noise and other environmental conditions," J. Acoust. Soc. Am. **113**, 1492–1504.
Mikula, G. (**2001**). "Justice: Social psychological perspectives," in *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. J. Smelser and P. B. Baltes (Elsevier, Amsterdam), pp. 8063–8067.
Miyakawa, M., Matsui, T., Matsui, Y., Murayama, R., Uchiyama, I., Itoh, T., and Yoshida, T. (**2004**). "Physiological effects of noise on salivary chromogranin A (CgA) as a measure of stress response," Internoise 2004, Prague, Czech Republic, 22–25 August.
Montada, L. (**2001**). "Justice and its many faces: Cultural concerns," in *International Encyclopedia of the Social and Behavioral Sciences*, edited by N. J. Smelser and P. B. Baltes (Elsevier, Amsterdam), pp. 8037–8042.
Ouis, D. (**2001**). "Annoyance from road traffic noise: A review," J. Environ. Psychol. **21**, 101–120.
Ouis, D. (**2002**). "Annoyance caused by exposure to road traffic noise: An update," Noise Health **4**, 69–79.
Robin, M., Ratiu, E., Matheau-Police, A., and Lavarde, A. M. (**2004**). "The evaluation of urban stressors. A preliminary report," IAPS, Vienna, Austria, 7–9 July.
Schomer, P. D. (**1988**). "On a theoretical interpretation of the prevalence rate of noise-induced annoyance in residential populations—High-amplitude impulse-noise environments," J. Acoust. Soc. Am. **86**, 835–836.
Schultz, T. J. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.
Schümer, R. (**1974**). "DFG-forschungsbericht fluglärmwirkungen. Sozialwissenschaftlicher ergänzungsbericht ("The impact of aircraft noise. An Interdisciplinary study into the consequences of aircraft noise for people. Additional social scientific report")," Bolt, Boppard, Germany.
Schümer-Kohrs, A., and Schümer, R. (**1974**). "DFG-forschungsbericht fluglärmwirkungen. Hauptbericht. Der sozialwissenschaftliche untersuchungsteil ("The impact of aircraft noise. An interdisciplinary study into the consequences of aircraft noise for people. Main report. Chapter 4. The social scientific part")," (Bolt, Boppard, Germany), pp. 150–246.
Smith, A. (**1989**). "A review of the effects of noise on human performance," Scand. J. Psychol. **30**, 185–206.
Southgate, D. (**2002**). "Expanding ways to describe and assess aircraft noise," EC Conference on Good Practice in Integration of Environment into Transportation Policy, Bruxelles, Belgium, 10–11 October. Retrieved 21 March 2007 from http://ec.europa.eu/environment/gpc/pdf/ws2d_southgate.pdf
Stallen, P. J. M. (**1999**). "A theoretical framework for environmental noise annoyance," Noise Health **3**, 69–79.
Staples, S. L. (**1997**). "Public policy and environmental noise: Modelling exposure or understanding effects," Am. J. Public Health **87**, 2063–2067.
Steensma, H., and Doreleijers, C. (**2003**). "Personnel selection: Situational test or employment interview? The validity versus justice dilemma," J. Indiv. Employment Rights **10**, 215–232.
Steensma, H., and Otto, L. (**2000**). "Perception of performance appraisal by employees and supervisors: Self-serving bias and procedural justice," J. Collective Negotiations Public Sector **29**, 307–319.
Tafalla, R. J., Evans, G. W., and Chen, A. (**1988**). "Noise and human performance: The potential role of effort," in *Noise as a Public Problem: Proceedings of the Fifth International Congress on Noise as a Public Health Problem*, edited by B. Berglund, U. Berglund, J. Karlsson, and T. Lindvall (Swedish Council for Building Research, Stockholm, Sweden), Vol. **3**, pp. 95–100.
Tepper, B. J. (**2001**). "Health consequences of organizational injustice: Tests of main and interactive effects," Org. Behav. Hum. Decis. Process **86**, 197–215.
Thibaut, J., and Walker, L. (**1975**). *Procedural Justice* (Erlbaum, Hillsdale, NJ).
Tyler, T. R. (**2000**). "Social justice: Outcome and procedure," Int. J. Psychol. **35**, 117–125.
Tyler, T. R., and Lind, E. A. (**1992**). "*Advances in Experimental Social Psychology*, edited by M. Zanna (Academic Press, NY), Vol. **25**, pp. 115-192.
Ulrich, R. S. (**1983**). "Aesthetic and affective response to natural environment," in *Human Behavior and Environment: Advances in Theory and Research*, Behavior and the Natural Environment, Vol. **6**, edited by I. Altman and J. F. Wohlwill (Plenum, New York), pp. 85–125.
Van den Bos, K., Bruins, J., Wilke, H. A. M., and Dronkert, E. (**1999**). "Sometimes unfair procedures have nice aspects: On the psychology of the fair process effect," J. Pers Soc. Psychol. **77**, 324–336.
Van den Bos, K., and Lind, E. A. (**2002**). "Uncertainty management by means of fairness judgments," in *Advances in Experimental Social Psychology*, edited by M. P. Zanna (Academic, San Diego), Vol. **34**, pp. 1–60.
Van den Bos, K., Lind, A. E., Vermunt, R., and Wilke, A. M. (**1997**). "How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect," J. Pers Soc. Psychol. **72**, 1034–1046.
Van Kamp, I., Job, R. F., Hatfield, J., Haines, M., Stellato, R. K., and Stansfeld, S. A. (**2004**). "The role of noise sensitivity in the noise-response relation: A comparison of three international airport studies," J. Acoust. Soc. Am. **116**, 3471–3479.
Vermunt, R., and Steensma, H. (**2001**). "Stress and justice in organizations: An exploration into justice processes with the aim to find mechanisms to reduce stress," in *Justice in the Workplace. From Theory to Practice*, edited by R. Cropanzano (Erlbaum, London), Vol. **2**, pp. 27–48.
Vermunt, R., and Steensma, H. (**2003**). "Physiological relaxation: Stress reduction through fair treatment," Soc. Justice Res. **16**, 135–150.
Vermunt, R., and Steensma, H. (**2005**). "How can justice be used to manage stress in organizations?," in *Handbook of Organizational Justice*, edited by J. G. Greenberg and J. A. Colquitt (Erlbaum, Mahwah, NJ), pp. 383–410.
Wirth, K., and Bröer, C. (**2004**). "More annoyed by aircraft noise than 30 years ago? Some figures and interpretations," Internoise 2004, Prague, Czech Republic, 22–25 August.
World Health Organization. (**2004**). "WHO LARES. Final report. Noise effects and morbidity," by H. Niemann and C. Maschke, Berlin Center of Public Health. Retrieved 21 June 2006 from http://www.euro.who.int/Document/NOH/WHO_Lares.pdf
ZEUS GmbH. (**2002**). "Wirkbezogenes lärmbeurteilungsverfahren ("Effect-related noise assessment")," Report No. F&E-Vorhaben 298 532 65, edited by U. Felscher-Suhr, R. Höger, and D. Schreckenberg, Bochum, Germany.

Acoustic diffraction effects at the Hellenistic amphitheater of Epidaurus: Seat rows responsible for the marvelous acoustics

Nico F. Declercq[a] and Cindy S. A. Dekeyser
Georgia Institute of Technology, George W. Woodruff School of Mechanical Engineering, 801 Ferst Drive, Atlanta, Georgia 30332-0405
and Georgia Tech Lorraine, 2 rue Marconi, 57070 Metz, France

(Received 13 November 2006; revised 25 January 2007; accepted 25 January 2007)

The Hellenistic theater of Epidaurus, on the Peloponnese in Greece, attracts thousands of visitors every year who are all amazed by the fact that sound coming from the middle of the theater reaches the outer seats, apparently without too much loss of intensity. The theater, renowned for its extraordinary acoustics, is one of the best conserved of its kind in the world. It was used for musical and poetical contests and theatrical performances. The presented numerical study reveals that the seat rows of the theater, unexpectedly play an essential role in the acoustics—at least when the theater is not fully filled with spectators. The seats, which constitute a corrugated surface, serve as an acoustic filter that passes sound coming from the stage at the expense of surrounding acoustic noise. Whether a coincidence or not, the theater of Epidaurus was built with optimized shape and dimensions. Understanding and application of corrugated surfaces as filters rather than merely as diffuse scatterers of sound, may become imperative in the future design of modern theaters. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2709842]

PACS number(s): 43.55.Gx, 43.20.El, 43.20.Fn [NX] Pages: 2011–2022

I. INTRODUCTION

In the classical world, the "asclepieion" at Epidaurus was the most celebrated and prosperous healing center;[1] in its vicinity there was the amphitheater, designed by Polycleitus the Younger in the fourth century B.C. and famous for its beauty and symmetry. The original 34 seat rows were extended in Roman times by another 21 rows. The theater is well preserved because it has been covered for centuries by thick layers of earth. A recent picture of the theater is presented in Fig. 1.

Marcus Vitruvius Pollio (first century B.C.) describes in his famous books "De Architectura"[2] the state of the art in architecture and shows evidence that man was aware of the physical existence of sound waves. He writes, "Therefore the ancient architects following nature's footsteps, traced the voice as it rose, and carried out the ascent of the theater seats. By the rules of mathematics and the method of music, they sought to make the voices from the stage rise more clearly and sweetly to the spectators' ears. For just as organs which have bronze plates or horn sounding boards are brought to the clear sound of string instruments, so by the arrangement of theaters in accordance with the science of harmony, the ancients increased the power of the voice."

This indicates that the construction of theaters was performed according to experimental knowledge and experience and that it was done such as to improve the transmission of sound from the center of the theater (the orchestra) toward the outer seats of the "cavea." It has however always been believed, even in the same chapter written by Vitruvius,[2] or the work by Izenour,[3] that it was mainly the aspect of the slope of the theater, as a result of the constructed seats, rather than the seats themselves, that have been a key factor in the resulting acoustic properties.

The current study was triggered by the marvels of Epidaurus and by recent advances in the explanation of a variety of diffraction effects on corrugated surfaces.[4–9]

The theory of diffraction of sound is based on the concepts of the Rayleigh decomposition of the reflected and transmitted sound fields.[10–12] The theory, earlier applied to describe a number of diffraction effects for normal incident ultrasound on corrugated surfaces,[13] has been used successfully to understand the generation of ultrasonic surface waves in the framework of nondestructive testing. The theory was later expanded to include inhomogeneous waves and enabled a description and understanding of the backward displacement of bounded ultrasonic beams obliquely incident on corrugated surfaces, a phenomenon which had been obscure for 3 decades.[9,14] Even more, it was later exposed that predictions resulting from that theory were in perfect agreement with new experiments.[15]

An expansion of the theory to pulsed spherical acoustic waves revealed special acoustic effects at Chichen Itza in Mexico.[7,8] The advantage of the theory is its ability to make quantitative simulations as they appear in reality. From those simulations, it is then possible to detect and characterize patterns and characteristics of the diffracted sound field such as in the case of a short sound pulse incident on the staircase of the El Castillo pyramid in Chichen Itza. The study indicated that the effects were slightly more complicated than the earlier considered principle of Bragg scattering. In the meantime, the fact that the Quetzal echo at Chichen Itza is influ-

[a]Author to whom correspondence should be addressed; electronic mail: nico.declercq@me.gatech.edu

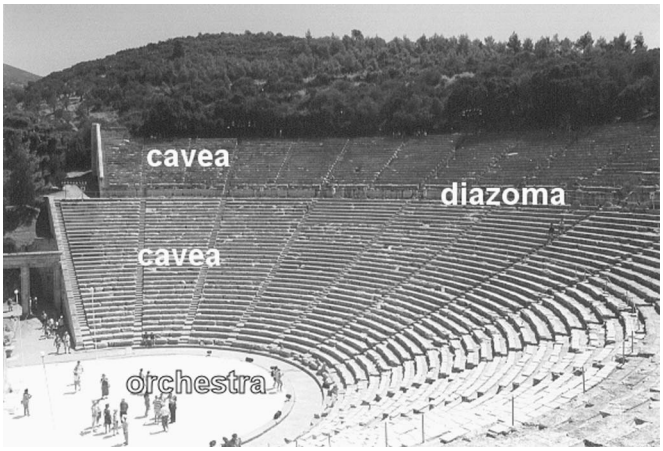


FIG. 1. Picture of the theater of Epidaurus (picture taken by the authors).

enced by the properties of the sound source as well as the existence of the "raindrop effect," have been experimentally verified by Cruz *et al.*[16] Bilsen[17] later showed that if one is only interested in the position of time delay lines on a sonogram and not in the entire amplitude pattern, that it is possible to apply a simpler model based on the gliding pitch theory.

For a study of acoustic effects at Epidaurus however, we are not interested in the response to a pulse. We are merely interested in how, for each frequency, sound behaves after interaction with the seats of the theater. Therefore the extensive diffraction theory, as used earlier,[7] is the pre-eminent tool.

Until now, there have appeared a number of "explanations" for the excellent acoustics of Epidaurus, such as that sound is driven by the wind because the wind is mostly directed from the orchestra toward the cavea. The wind direction has indeed some influence, but it is also known that the acoustics of Epidaurus is very good when there is no wind or when wind comes from other directions; wind even has a general negative effect because it produces undesirable noise. Another theory is the importance of the rhythm of speech but there are also modern performances taking place at Epidaurus where the typical rhythm of Hellenistic poems and performances composed by Homerus, Aeschylus, Sophocles, or Euripides is not there; still the acoustics seems perfect.

The last theory is that special masks, worn by performers, may have had a focusing effect on the generated sound, but that does not explain why speakers with weak voices are also heard throughout the theater.

Izenour[3] points out that the acoustics is so good because of the clear path between the speaker and the audience. The current work proves numerically that the effect of diffraction on the seat rows is probably an even more important effect than the "clear path effect."

In what follows, we describe the geometry of the theater. Consequently we explain briefly how the numerical simulations are performed. Then we present and explain the numerical results. We end our paper with the most important conclusions.

The material parameters at Epidaurus have been taken as: 2000 kg/m^3 for the density of the theater's limestone, and a shear wave velocity of 2300 m/s and longitudinal wave velocity of 4100 m/s.

For the air at Epidaurus, we have taken two cases: "summer," corresponding to an air density of 1.172 kg/m^3 and a (longitudinal) wave velocity of 348.04 m/s; and "winter," corresponding to an air density of 1.247 kg/m^3 and a (longitudinal) wave velocity of 337.50 m/s.

II. GEOMETRY: MILLER PROJECTION

In this paper, we only focus on the geometrical properties of the theater that are important for the acoustics. The theater is almost semicircular. This means that the acoustics, for a sound source situated at the center of the theater, will have a circular symmetry similar to the theater itself. A Miller projection (as in cartography), mathematically transforming the semicircular theater into a rectangular theater resulting in seat rows in the cavea becoming straight rows having the same length as the outer seat row; and transforming or "stretching" the central spot (at the center of the orchestra) into a straight line parallel with the transformed theater and having the same length as the seat rows; makes the sound source at the center of the orchestra become a line source that generates cylindrical waves. For simplicity, we do not take into account edge effects at the edges of the seat rows. We may therefore disregard one Cartesian coordinate and study the entire problem in a two-dimensional space.

All this is physically correct if we also perform a Miller projection of the entire sound field. In other words the sound amplitude must be multiplied by a function describing the sound density variation along the theater slope due to the Miller projection. An inverse function must then be applied to the results if we want to transform the results back to the circular theater. Conveniently it is therefore unnecessary to consider this function because we do not want to show results that are valid for the transformed theater, but in the real circular theater. Furthermore a source not exactly situated at the center of the orchestra will deliver exact results along the theater radius passing through the source but will yield slightly deviating results for other positions in the theater's cavea.

III. GEOMETRY: SHAPE, SIZE, AND DISTANCES

A relict of the extension from 34 seat row to 55 in Roman times is the presence of a "diazoma" in between both constructions as can be seen in Fig. 1. This acoustic discontinuity is neglected in our study because the upper seat rows are built along the same straight line as the inner seat rows and this neglect mathematically corresponds to adding a few seat rows and makes the two-piece theater a single-piece theater having 60 seat rows instead of 55.

It is necessary to define a number of vectors and angles of importance. They are depicted in Figs. 2 and 3.

Tables I–III explain the variables depicted in Figs. 2 and 3 and indicate the numerical values for the theater at Epidaurus.

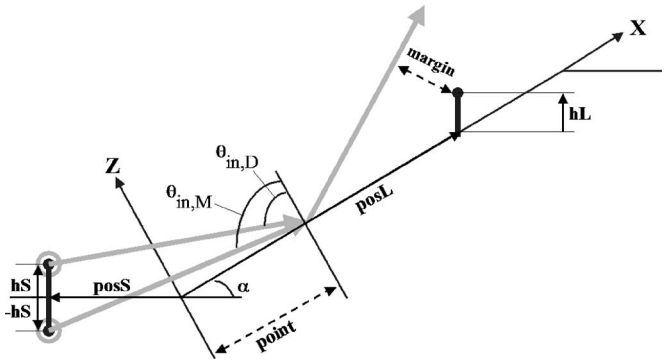


FIG. 2. Vectors and angles used to describe the theater and the acoustics.

Straightforward geometrical considerations yield for the direct distance between source and receiver:

$$\mathrm{RD} = \sqrt{(\mathrm{posL}\cos(\alpha) + \mathrm{posS})^2 + (\mathrm{hL} + \mathrm{posL}\sin(\alpha) - \mathrm{hS})^2}. \tag{1}$$

Analogously we obtain for the distance between source and receiver, taking into account the mirror effect caused by the foreground:

$$\mathrm{RM} = \sqrt{(\mathrm{posS} + \mathrm{posL}\cos(\alpha))^2 + (\mathrm{hL} + \mathrm{posL}\sin(\alpha) + \mathrm{hS})^2}. \tag{2}$$

The numerical procedure developed for this paper is based on consecutive consideration of diffraction in subsequent spots of diffraction. With respect to these spots "P" of diffraction, we define a number of valuable distances:

$$\mathrm{RSP} = \sqrt{(P - \mathrm{hS}x - \mathrm{posS}x)^2 + (\mathrm{posS}z + \mathrm{hS}z)^2} \tag{3}$$

which is the distance between the actual sound source and the spot of diffraction;

$$\mathrm{RMP} = \sqrt{(P + \mathrm{hS}x - \mathrm{posS}x)^2 + (\mathrm{posS}z - \mathrm{hS}z)^2} \tag{4}$$

is the distance between the mirror source and the spot of diffraction and

$$\mathrm{RPL} = \sqrt{(xL - P)^2 + zL^2}, \tag{5}$$

being the distance between the diffraction spot and the listener where

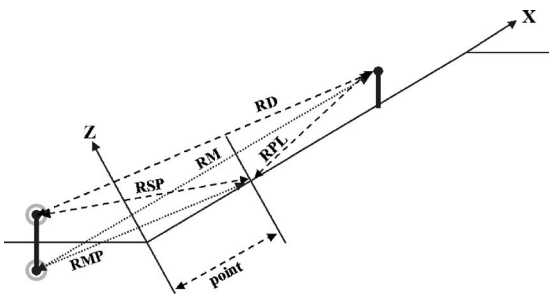


FIG. 3. Vectors and angles used to describe the theater and the acoustics.

TABLE I. Measured values describing the theater.[a]

Quantity	Value	Meaning
b	0.746 m	Width of the seats
α	26.6°	Slope of the theater
$\mathrm{SD}_{\max}$	22.63 m	Distance between center of orchestra and lower seat row
l_{theater}	49.88 m	Length of the seats

[a]See Ref. 1.

$$xL = \mathrm{posL} + \mathrm{hL}\sin(\alpha) \quad \text{and} \quad zL = \mathrm{hL}\cos(\alpha). \tag{6}$$

A considered ray of sound is incident at the diffraction spot at the angle of incidence θ_{in} (cf. Fig. 2). If we do not consider any reflections on the foreground, then this angle is equal to

$$\theta_{\mathrm{in},D} = \arccos\left(\frac{\mathrm{posS}z + \mathrm{hS}z}{\mathrm{RSP}}\right). \tag{7}$$

If we also consider a reflection on the foreground then the angle of incidence is

$$\theta_{\mathrm{in},M} = \arccos\left(\frac{\mathrm{posS}z - \mathrm{hS}z}{\mathrm{RMP}}\right). \tag{8}$$

IV. ACOUSTIC SIMULATIONS

A. Sound field description

As noted earlier, the sound source is considered cylindrical. The generated sound field is thought of as a bunch of rays spread over all directions and widening with increased distance (involving an amplitude inversely proportional to the square root of the traveled distance) from the source just as a real cylindrical sound field. The phase of the sound within each considered beam also behaves as the actual cylindrical sound field. Normalization yields the summation of the amplitudes of all beams to be equal to unity. These rays interact with the theater. Because at considerable distances from the sound source the sound field pattern in each of the "rays" approximates a plane wave, the interaction of the rays with the theater is modeled as a plane wave interaction, allowing the use of earlier developed techniques based upon Rayleigh's theory of diffraction.[9,10,13] The diffracted sound fields are also thought of as a widening sound ray continuing the same widening pattern and pace as the incident sound ray. We apply Rayleigh's decomposition, therefore the incident sound field (displacement field) is given by

$$\mathbf{N}^{\mathrm{inc}} = A^{\mathrm{inc}}\varphi^{\mathrm{inc}}(ik_x^{\mathrm{inc}}\mathbf{e}_x + ik_z^{\mathrm{inc}}\mathbf{e}_z). \tag{9}$$

The reflected ($\zeta = r$) and transmitted longitudinal ($\zeta = d$) sound fields are given by

TABLE II. Calculated values describing the theater.

Quantity	Equal to	Value	Meaning
h	$b\tan(\alpha)$	0.367 m	Height of the seats
Λ	$\sqrt{b^2+h^2}$	0.831 m	Periodicity of the seat rows
n	$\frac{l_{\mathrm{theater}}}{\Lambda}$	60	Number of seat rows

TABLE III. Explanation of abbreviations used throughout the paper.

Abbreviation	Explanation
hS	height of the source
posS	position of the source
hL	height of the listener
posL	position of the listener
RD	direct distance
RSP	distance between source and point of diffraction
RPL	distance between point of diffraction and listener
RM	direct distance for mirror source
RMP	distance between mirror source and diffraction point

$$\mathrm{N}^{\varsigma} = \sum_m A_m^{\varsigma} \varphi^{m,\varsigma} (i k_x^{m,\varsigma} e_x + i k_z^{m,\varsigma} e_z), \quad \varsigma = r,d. \tag{10}$$

Finally, the transmitted shear sound field is written as

$$\mathbf{N}^s = \sum_m A_m^s \mathbf{P}^{m,s} \varphi^{m,s} \tag{11}$$

with

$$\varphi^{\tau} = \exp i(k_x^{\tau} x + k_z^{\tau} z) \tag{12}$$

and

$$k_x^{m,s} P_x^{m,s} + k_z^{m,s} P_z^{m,s} = 0. \tag{13}$$

B. Mechanical continuity conditions

The sound fields described in Eqs. (9) and (10) must correspond to incident and diffracted sound on the air-solid interface formed by the seat rows.

In order to determine the unknown coefficients A_m^r, A_m^d, $A_m^s P_x^{m,s}$, and $A_m^s P_z^{m,s}$ we impose continuity of normal stress and normal displacement everywhere along the interface between air and solid. The corrugated surface is given by a function $z=f(x)$. Periodicity of the corrugation yields

$$f(x+\Lambda) = f(x) \tag{14}$$

with Λ the corrugation period. For further use, we define the function $g(x,z)$ as follows:

$$g(x,z) = f(x) - z. \tag{15}$$

Along the interface we have $g(x,z)=0$.

We do not consider viscous damping effects. The stress tensor T^{τ} ($\tau=1$ in air, $\tau=2$ the solid), is calculated as

$$T_{ij}^{\zeta} = \sum_{\eta} \lambda^{\tau} \varepsilon_{\eta\eta}^{\tau} \delta_{i,j} + 2\mu^{\tau} \varepsilon_{i,j}^{\tau} \tag{16}$$

in which λ^{τ} and μ^{τ} are Lamé's constants.

The strain tensor ε^{τ} is calculated as

$$\varepsilon_{ij}^{\tau} = \tfrac{1}{2}(\partial_i N_j^{\tau} + \partial_j N_i^{\tau}). \tag{17}$$

We also incorporate the dispersion relations for longitudinal waves

$$k^{\zeta} = \sqrt{\frac{\rho\omega^2}{\lambda^{\tau} + 2\mu^{\tau}}} \tag{18}$$

with ζ="inc" or "m,r" and for shear waves

$$k^{\zeta} = \sqrt{\frac{\rho\omega^2}{\mu^{\tau}}} \tag{19}$$

with $\zeta=s,2$ for shear waves in the solid.

The dispersion relations (18) and (19) reveal the value of k_z corresponding to each of the values for k_x for the different diffraction orders. The sign of k_z is chosen according to the well-known "Sommerfeld conditions" stating that each of the generated waves must propagate away from the interface and demanding that whenever k_z is purely imaginary (evanescent waves), its sign must be chosen such that the amplitude of the wave under consideration diminishes away from the interface.

Continuity of normal stress and normal displacement everywhere along the interface between air and solid yield

$$(\mathbf{N}^{\mathrm{inc}} + \mathbf{N}^r) \cdot \boldsymbol{\nabla} g = (\mathbf{N}^d + \mathbf{N}^s) \cdot \boldsymbol{\nabla} g \quad \text{along } g = 0, \tag{20}$$

$$\sum_j T_{ij}^1 (\nabla g)_j = \sum_j T_{ij}^2 (\boldsymbol{\nabla} g)_j \quad \text{along } g = 0. \tag{21}$$

Relations (13), (20), and (21) result in four equations that are periodical along the x axis. A discrete Fourier transform with repetition period Λ is eminent and each of the Fourier components on both sides of the equations are then equal to one another.

Straightforward calculations ultimately result in four continuity equations

Equation 1:

$$\begin{aligned} & A^{\mathrm{inc}} I^{\mathrm{inc},p} i(-(k^1)^2 + k_x^{\mathrm{inc}} k_x^p) + \sum_m A_m^r I^{m,r,p} i(-(k^1)^2 + k_x^m k_x^p) \\ & \quad + \sum_m A_m^d I^{m,d,p} i(-(k^{d,2})^2 + k_x^m k_x^p) - \sum_m A_m^s P_x^{m,s} I^{m,s,p} \\ & \quad \times (k_x^p - k_x^m) + \sum_m A_m^s P_z^{m,s} I^{m,s,p} (k_z^{m,s}) = 0. \end{aligned} \tag{22}$$

Equation 2:

$$\begin{aligned} & - A^{\mathrm{inc}} I^{\mathrm{inc},p} \rho_1 (k_x^p - k_x^{\mathrm{inc}}) - \sum_m A_m^r I^{m,r,p} \rho_1 (k_x^p - k_x^m) \\ & \quad + \sum_m A_m^d I^{m,d,p} \rho_2 \left(- k_x^m + \left(1 + 2 \frac{(k_x^m)^2 - (k^{d,2})^2}{(k^{s,2})^2} \right) k_x^p \right) \\ & \quad + \sum_m A_m^s P_x^{m,s} I^{m,s,p} i\rho_2 \left(1 - \frac{k_x^m k_x^p}{(k^{d,2})^2} + \left(\frac{1}{(k^{d,2})^2} \right. \right. \\ & \quad \left. \left. - \frac{1}{(k^{s,2})^2} \right) (k_x^m)^2 \right) + \sum_m A_m^s P_z^{m,s} I^{m,s,p} i\rho_2 (k_z^{m,s}) \left(\left(\frac{1}{(k^{d,2})^2} \right. \right. \\ & \quad \left. \left. - \frac{1}{(k^{s,2})^2} \right) k_x^m - \left(\frac{1}{(k^{d,2})^2} - \frac{2}{(k^{s,2})^2} \right) k_x^p \right) = 0. \end{aligned} \tag{23}$$

Equation 3:

$$A^{\mathrm{inc}} I^{\mathrm{inc},p} \rho_1(k_z^{\mathrm{inc}}) + \sum_m A_m^r I^{m,r,p} \rho_1(k_z^{m,r}) + \sum_m A_m^d I^{m,d,p}(k_z^{m,d}) \rho_2\left(-1 + \frac{2}{(k^{s,2})^2}(k_x^m k_x^p)\right) + \sum_m A_m^s P_x^{m,s} I^{m,s,p} i(k_z^{m,s}) \rho_2\left(\left(\frac{1}{(k^{d,2})^2} - \frac{1}{(k^{s,2})^2}\right) k_x^m - \frac{k_x^p}{(k^{s,2})^2}\right) + \sum_m A_m^s P_z^{m,s} I^{m,s,p} i \rho_2\left(\left(\frac{1}{(k^{d,2})^2} - \frac{1}{(k^{s,2})^2}\right) \times (k_z^{m,s})^2 + 1 - \frac{k_x^m k_x^p}{(k^{s,2})^2}\right) = 0. \quad (24)$$

Equation 4:

$$(A_m^s P_x^{m,s} k_x^{m,s} + A_m^s P_z^{m,s} k_z^{m,s}) \delta_{m,p} = 0. \quad (25)$$

$\delta_{m,p}$ in Eq. (25) is Kronecker's delta.

The grating equation (similar to the one in optics) takes care of k_x^m and k_x^p as follows:

$$k_x^\beta = k_x^{\mathrm{inc}} + \beta \frac{2\pi}{\Lambda}, \quad \beta = m,p \in \mathbb{Z}. \quad (26)$$

The Fourier transformation also leaves integrals within Eqs. (22)–(24):

$$I^{\mathrm{inc},\eta} = \frac{1}{k_z^{\mathrm{inc}}} \int_\Lambda \exp i[(k_x^{\mathrm{inc}} - k_x^\eta)x + k_z^{\mathrm{inc}} f(x)] dx, \quad (27)$$

$$I^{m,\xi,\eta} = \frac{1}{k_z^{m,\xi}} \int_\Lambda \exp i[(k_x^m - k_x^\eta)x + k_z^{m,\xi} f(x)] dx. \quad (28)$$

The integrals (27) and (28) can be solved numerically or analytically.[9] They contain information about the surface and are therefore called "surface integrals."

C. The number of diffraction orders

Equations (22)–(25) actually represent an infinite number of equations and unknown variables because the orders m and p constitute a discrete infinite interval of integer numbers $\mathbb{Z}$. As discussed in earlier papers,[5,7,9,13] finiteness of energy makes a limitation of the number of diffraction orders authorized because only a few orders are really propagating; the others are evanescent and with increasing value of m or p, play a less important role in the energy transformation upon diffraction.[13] Therefore we only take into account two forward and two backward "propagating" evanescent waves for each of the considered frequencies. In other words, for each frequency we consider the propagating bulk waves (their number depends on the frequency) and add two more evanescent waves in each direction. The developed procedure therefore automatically determines the number of waves involved.

D. The formation of observed diffracted sound per generated ray

By "observed diffracted sound" we mean sound that reaches a given observer in the cavea of the theater. Because we model the acoustics by means of cylindrically expanding rays, it is clear that not all of these diffracted rays will ultimately reach the observer.

It is known from textbooks on geometry that the distance from the diffracted ray to the observer is given by

$$\left| \frac{(xL + \mathrm{pos}L - p)\frac{\mathrm{Re}(k_z^m)}{\mathrm{Re}(k_x^m)} - zL}{\sqrt{\left(\frac{\mathrm{Re}(k_z^m)}{\mathrm{Re}(k_x^m)}\right)^2 - 1}} \right|. \quad (29)$$

If this distance is smaller than a predetermined limit, the ray is considered to reach the observer. If the distance is larger, we further discard that ray. The limit is determined by the width of the ray at the point of observance. We approximate this width by its slightly larger value

$$\mathrm{limit} = 2 \max(\theta_{\mathrm{in}}^{+1}, \theta_{\mathrm{in}}^{-1})(\mathrm{RSP} + \mathrm{RPL}) \quad (30)$$

with $\theta_{\mathrm{in}}^{+1}$ and $\theta_{\mathrm{in}}^{-1}$ the angle (in rad) between the considered ray and the consecutive ray, respectively, the angle between the considered ray and the preceding ray.

Whenever we consider sound that is reflected in the orchestra on the foreground of the theater, we replace "RSP" by "RMP" in Eq. (30).

E. The integrated effect for all considered rays

The previous paragraph describes the interaction of one ray at one single spot of the theater and it is determined whether or not an observer will "detect" or "hear" the diffracted rays. Calculation of the integrated effect consists of a repetition of the previous procedures for each considered generated ray and adding up all rays that are detected by the observer. To approach physical reality, we model the generated cylindrical sound field by a bunch of rays that fulfill specific incorporated requirements. The distribution of rays is made such that the rays would be incident at spots P on the theater at equally spaced positions and such that there would be three spots of incidence per wavelength, therefore producing realistic simulations. Furthermore, we can apply the above-mentioned procedure for each position of the listener "posL" and then plot the result as a function of the position of the listeners on the theater.

F. The Lipmann and Wirgin criteria

The considered model, based on Rayleigh's decomposition, is a simplified approach of more complicated models such as the differential[18,19] and integral equation approach[20–23] and Waterman's theory;[24–28] it is not valid for any situation. There are certain requirements that need to be fulfilled as studied by Lipmann[29] and later also by Wirgin.[12]

Wirgin has shown[12] that "contrary to prevailing opinion, the Rayleigh theory is fully capable of describing the scattering phenomena produced by a wide class of corrugated

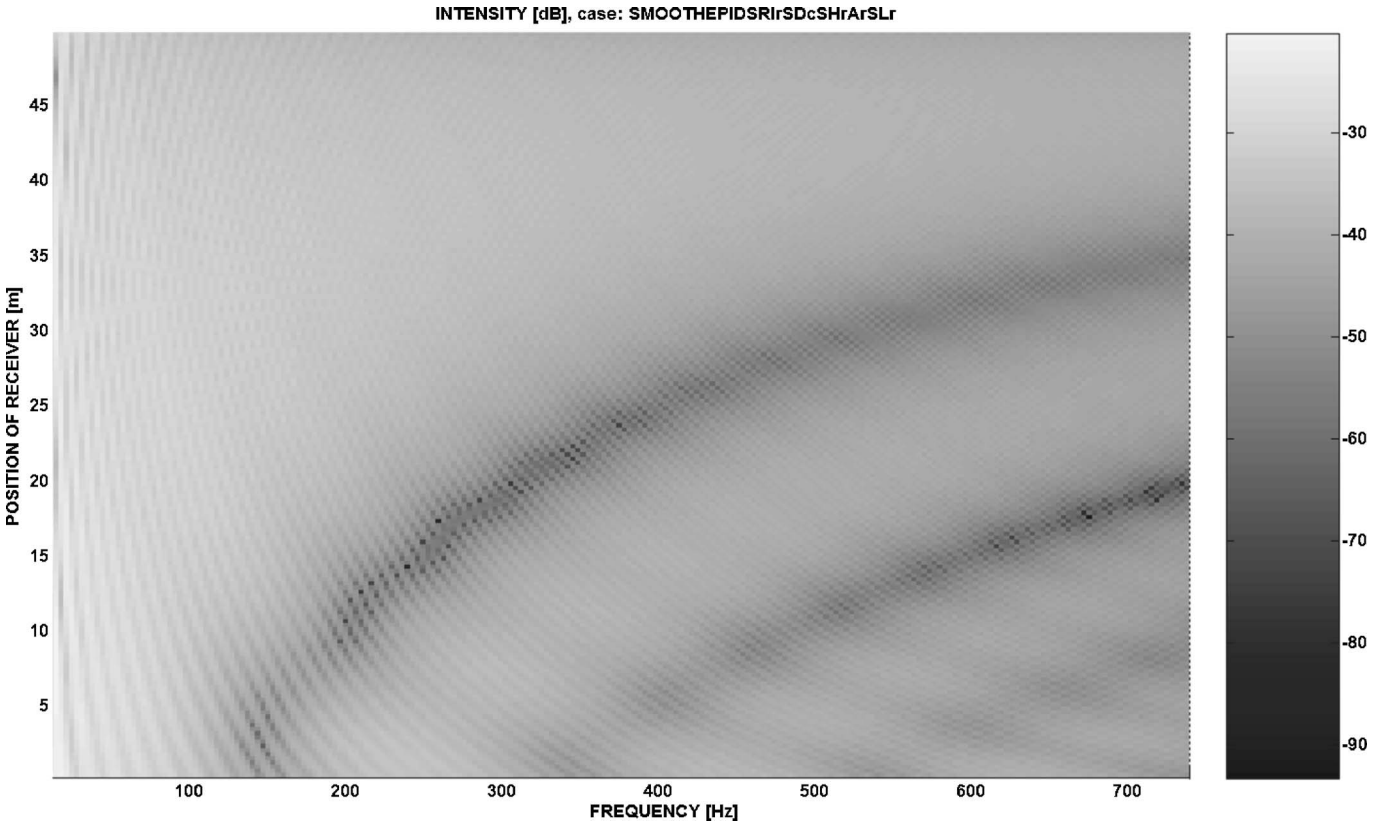


FIG. 4. The received intensity in decibels for listeners situated at heights along the slope of a smooth theater, i.e., without seat rows, given along the vertical axis and for frequencies given along the horizontal axis. The geometry corresponds to the geometry of Epidaurus and the sound source is situated at 22.63 m from the first row of seats, i.e., at the center of the theater. Reflections on the foreground are neglected.

surfaces, including those whose roughness is rather large." Furthermore, Wirgin[12] proves that the Rayleigh theory is valid for λ the largest wavelength involved in the diffraction phenomenon, for Λ the corrugation period and for h the corrugation height, whenever

$$h < 0.34\Lambda \tag{31}$$

and

$$\lambda > 1.53348h. \tag{32}$$

The Wirgin criteria are somewhat tighter than the older Lipmann criteria.[22,29] Nevertheless we may expect that the Rayleigh theory for our purpose is reliable for frequencies below 750 Hz (in summer and in winter). Actually the theory may even be valid to a large extent above 800 Hz. A limitation to 750 Hz means that for a piano with 88 keys, our model would simulate the acoustics at Epidaurus for the first 58 keys, this is almost 70% and is not too bad.

V. NUMERICAL RESULTS

In all our calculations we have considered an observer whose ears are 80 cm above his seat. First consider a smooth Epidaurus theater, i.e., a theater that consists of a smooth slope without seat rows, making the sound rays undergo no diffraction but simple reflections on the slope transmitting a part of their energy into the limestone slope and reflecting most of their energy. Calculations then reveal the observed frequency spectrum for all positions posL on the slope. Consider a sound source placed in the center at 22.63 m from the first seat row and having a height of 2 m. This height is reasonable since in the Hellenistic era the performers, who were not very tall, wore "cothurns" or high theater sandals.

For simplicity, reflections on the foreground are not considered for the moment. Figure 4 shows the calculated results. The grayscale indicates the received sound intensity whereas the horizontal axis gives the frequency and the vertical axis corresponds to the height along the slope of the theater (posL).

Notice the appearance of distinct patterns due to the interference between sound reaching the listener uninterrupted and sound reaching the listener after being reflected upon the slope of the theater. The "bands" of diminished intensity are actually due to a phase canceling effect and are positions where the audience will receive a much lower intensity than at other positions in the cavea.

Figure 5 is similar to Fig. 4, except that here reflections on the foreground are also taken into account, resulting in a more complicated pattern, but with less distinct regions of diminished intensity. Reflections on the foreground are therefore responsible for a better distribution of sound throughout the theater.

Consider the situation at Epidaurus. Figure 6 shows the results for Epidaurus with the seat rows installed (at a periodicity of 0,831 m) and with reflections on the foreground. The sound source is again situated as in Fig. 5.

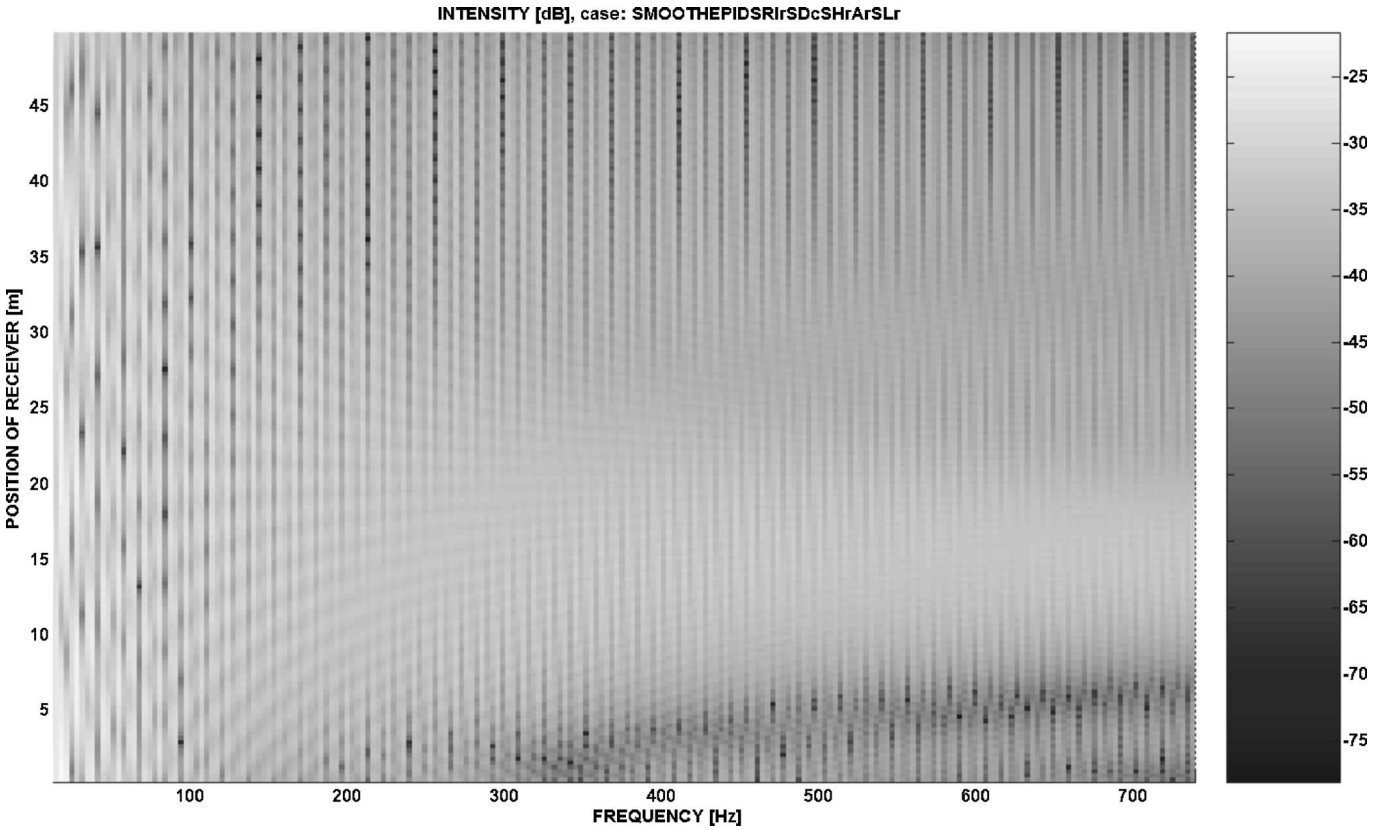


FIG. 5. Similar to Fig. 4, but with incorporation of reflections on the foreground. Reflections on the foreground are responsible for a better distribution of sound throughout the theatre.

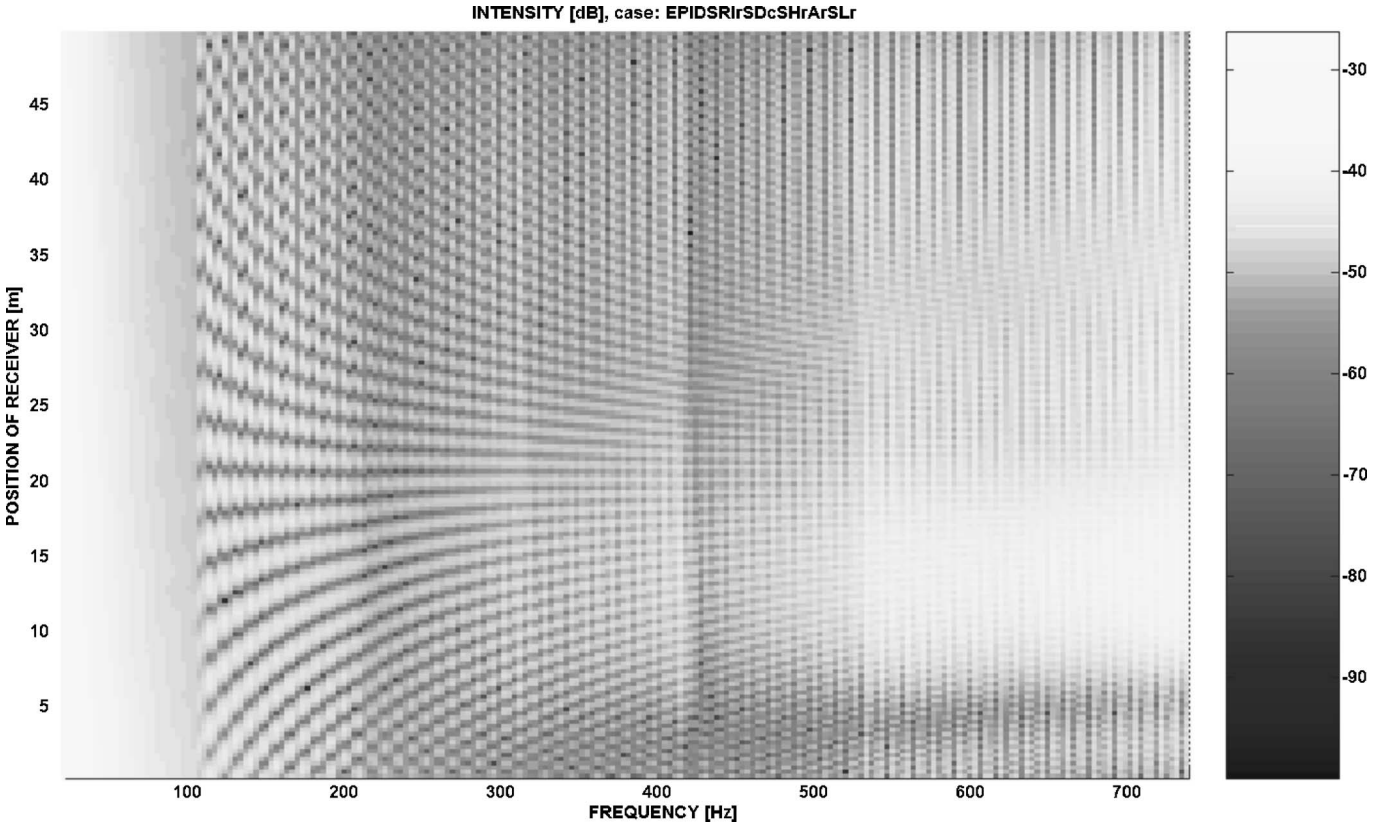


FIG. 6. Calculated intensities, comparable with Fig. 5, but with the seat rows installed. The sound patterns are now influenced by the effect of diffraction. Note that there is a relatively increased amplitude noticeable for high frequencies, whereas the overall sound intensity is lower than in the case without seat rows.

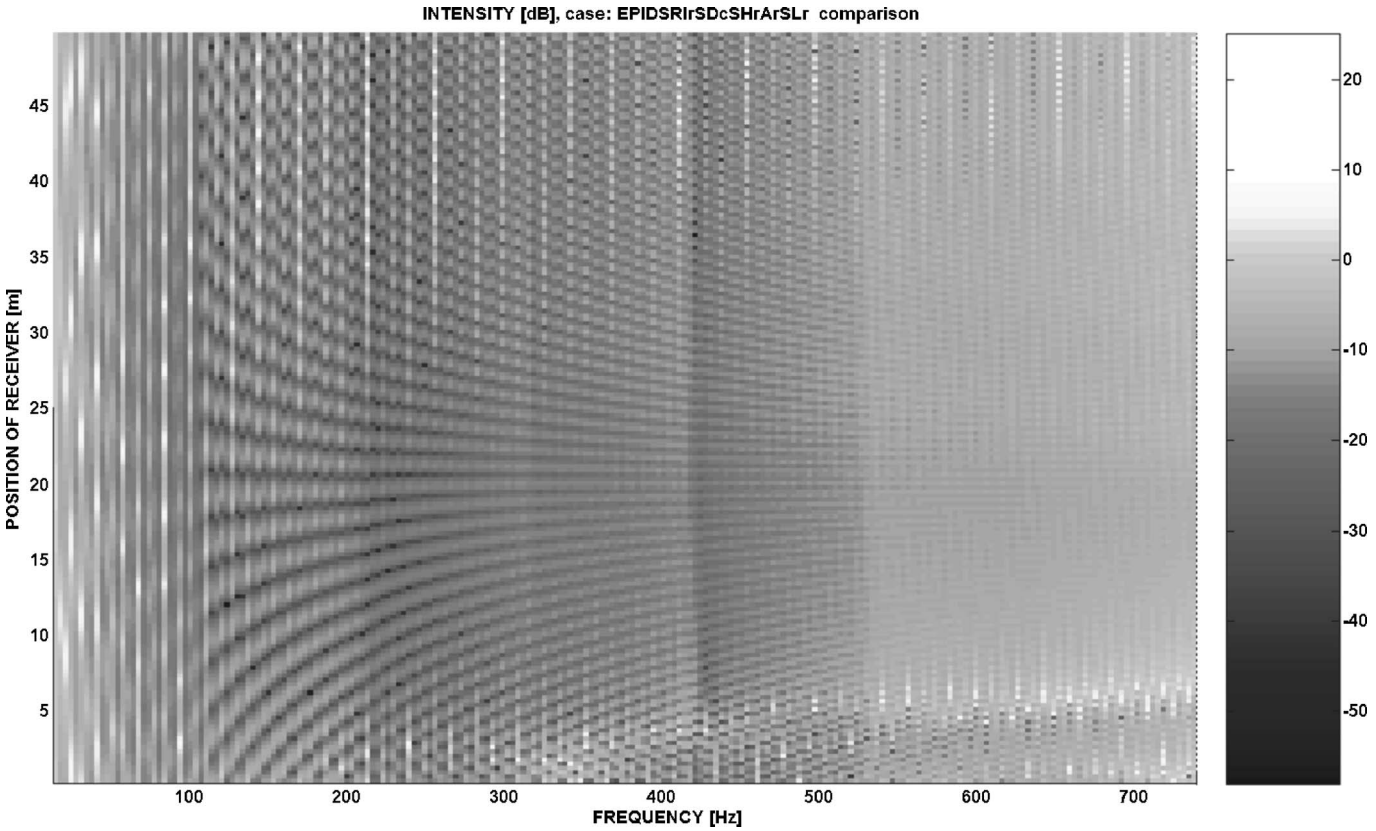


FIG. 7. Comparison of Fig. 6 with Fig. 5, highlighting the effect of diffraction due to the installation of seat rows. At most positions and for most frequencies, the intensity is diminished. However for frequencies beyond 530 Hz, one can see a relatively increased intensity. This is due to the filter effect caused by the seat rows.

Note that the intensities are slightly lower than for the case without seat rows. In other words, the installation of seat rows has a negative effect on the overall intensity of sound throughout the theater.

As a matter of fact, the results are not really simple to interpret because they show the cumulative effect caused by the installed seat rows, caused by reflections on the foreground and caused by the effect of the slope of the theater.

In order to highlight the particular effect of the seat rows, which is the main purpose of this paper, it is necessary to subtract Fig. 6 from Fig. 5. The result is shown in Fig. 7.

Because of the complexity of the diffraction phenomenon, the results are not really "smooth." Still there are certain tendencies visible. First, the relative intensities are almost everywhere negative. This means that the presence of stairs has a "damping effect" due to scattering in multiple directions. Nevertheless, an overall drop of intensity is not dramatic as the human ear is capable of adjusting its sensitivity. What is more important is the fact that frequencies beyond 530 Hz are less damped than frequencies between 50 and 530 Hz. Therefore there is a relative amplification of high frequencies. There is also a dependency of the position in the theater on the observable intensity, but this is mainly caused by the slope and not really by the seat rows, as can be seen in Fig. 5.

Note that reflections on the foreground are also very important for the real theater of Epidaurus. This can be clearly seen in Fig. 8, which is comparable to Fig. 7, except that reflections on the foreground are neglected.

There are high intensity bands appearing from down under to right up, which are merely due to interference effects due to straight sound and zero order diffracted sound that reaches the listener after diffraction. These bands correspond to low physical intensities and are very distinct when the theater contains no seats. In other words the existence of a reflective foreground results in a better distribution of sound throughout the theater and this redistribution is further improved, in addition to the filtering effect in favor of high frequencies, by the presence of seat rows.

Further results (left out of the paper) have revealed the influence of the seat row periodicity on the acoustics. For Aphrodisias, with a periodicity of 0.736 m, the relatively amplified frequencies are higher than 600 Hz. For Pergamon, with a periodicity of 1.657 m the relatively amplified frequencies begin around 300 Hz. In other words, the periodicity of the seat rows influences the band of amplified frequencies: the smaller the periodicity, the higher the amplified frequency band.

Additional results (equally left out of the paper) show that patterns appearing at a certain height on the slope of the theater shift to higher positions if the source is placed higher.

We have also studied the effect on the acoustics of the distance between the source and the first seat row (the "prohedriai"). Apart from an increased overall intensity when the source is positioned closer to the seat rows, we did not detect any spectacular effects except that reflections on the foreground become less important for a source closer to the seat rows, therefore destroying the positive effect of a better dis-

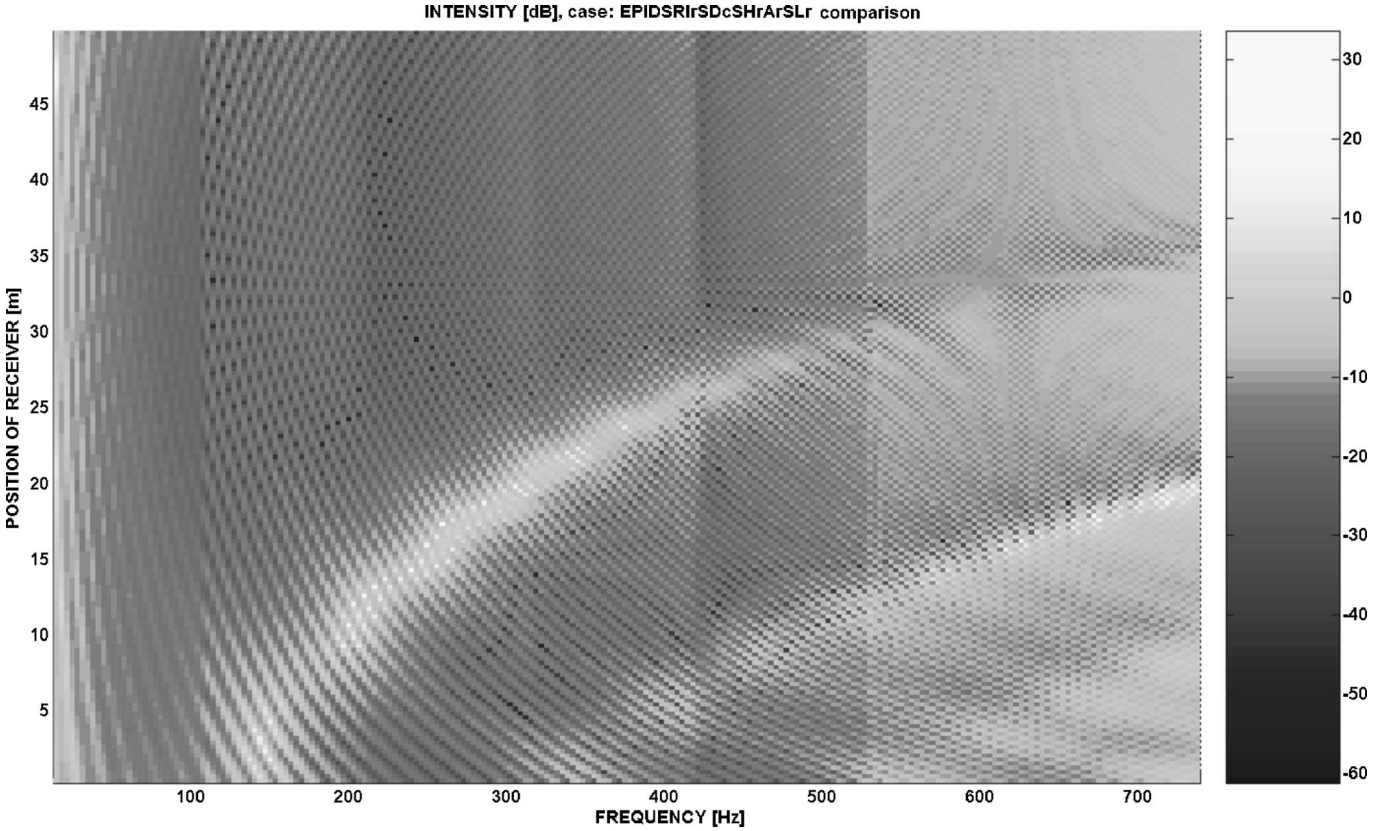


FIG. 8. Comparable to Fig. 7, but in the case of the absence of a reflective foreground. Frequencies above 530 Hz are still favored by the filtering effect, but there appears a position dependent intensity which is caused by the slope, just as in Fig. 4, and is only partly annihilated by the seat rows. This is mainly due to the interaction of uninterrupted sound beams and zero-order diffracted (i.e., undiffracted) sound beams reflected from the cavea.

tribution of sound throughout the theater. Still diffraction of sound on the seat rows makes the effect less dramatic.

We have also studied the influence of the slope of the theater on the acoustics. This effect is very important for a smooth theater without reflections on the foreground. The effects are still noticeable in the case of installed seat rows and reflections on the foreground, but it is less outspoken. The slope does not really influence the frequency values where the amplified frequency band appears.

The previous results were all for summer. Another aspect that we have studied is the influence of the season on the acoustics of Epidaurus. The season has an influence on both the sound velocity in air and the density of air. The differences in the limestone are negligible. We found that there was no significant difference between the acoustics in summer and the acoustics in winter.

VI. THE PHYSICAL ORIGIN OF THE HIGH PASS FILTER EFFECT

For low frequencies, the seat rows do not really diffract sound, which means that there is no big difference compared with a smooth slope. For higher frequencies, diffraction plays a role and higher order reflected sound is generated, causing sound to be distributed in different directions upon reflection into the air and upon transmission into the bulk of the theater's slope. Figures 9 and 10 show the intensity of the isolated reflected diffraction orders "-1" and "-2," respectively.

There is a "vertical raster" added to the figure that indicates the transition from evanescent waves to propagating bulk waves; for frequencies below the raster, sound is evanescent and is stuck to the slope of the theater, for frequencies passing the raster, sound is really propagating in space and is observable. Note that there are negative first-order diffracted waves observable at frequencies above 200 Hz, but that their intensity is really small (−15 dB and much less). The second negative order diffracted waves appear beyond around 450 Hz and their intensity is higher (−10 dB and higher). These facts result in the following analysis: For low frequencies there is no significant effect caused by the seat rows. For higher frequencies the reflected sound is distorted by the diffraction effect resulting in a distribution of the sound energy in many directions and actually causing a drop in the measured sound intensities for the audience. For frequencies in between 100 and 500 Hz, there is a physical influence of the negative first-order diffracted waves on the acoustics of the theater. The amplitude of these first-order waves is very small and therefore it mainly causes a distortion of the sound field and diminishes the observed intensities. For frequencies beyond 500 Hz, the negative second-order waves become important and they do have a significant intensity. These negative second-order waves actually consist of backscattered sound; for a given listener somewhere on the cavea of the theater, they consist of sound that has passed the listener and is backreflected toward this person. Because the accompanied intensity is considerable, it results in an

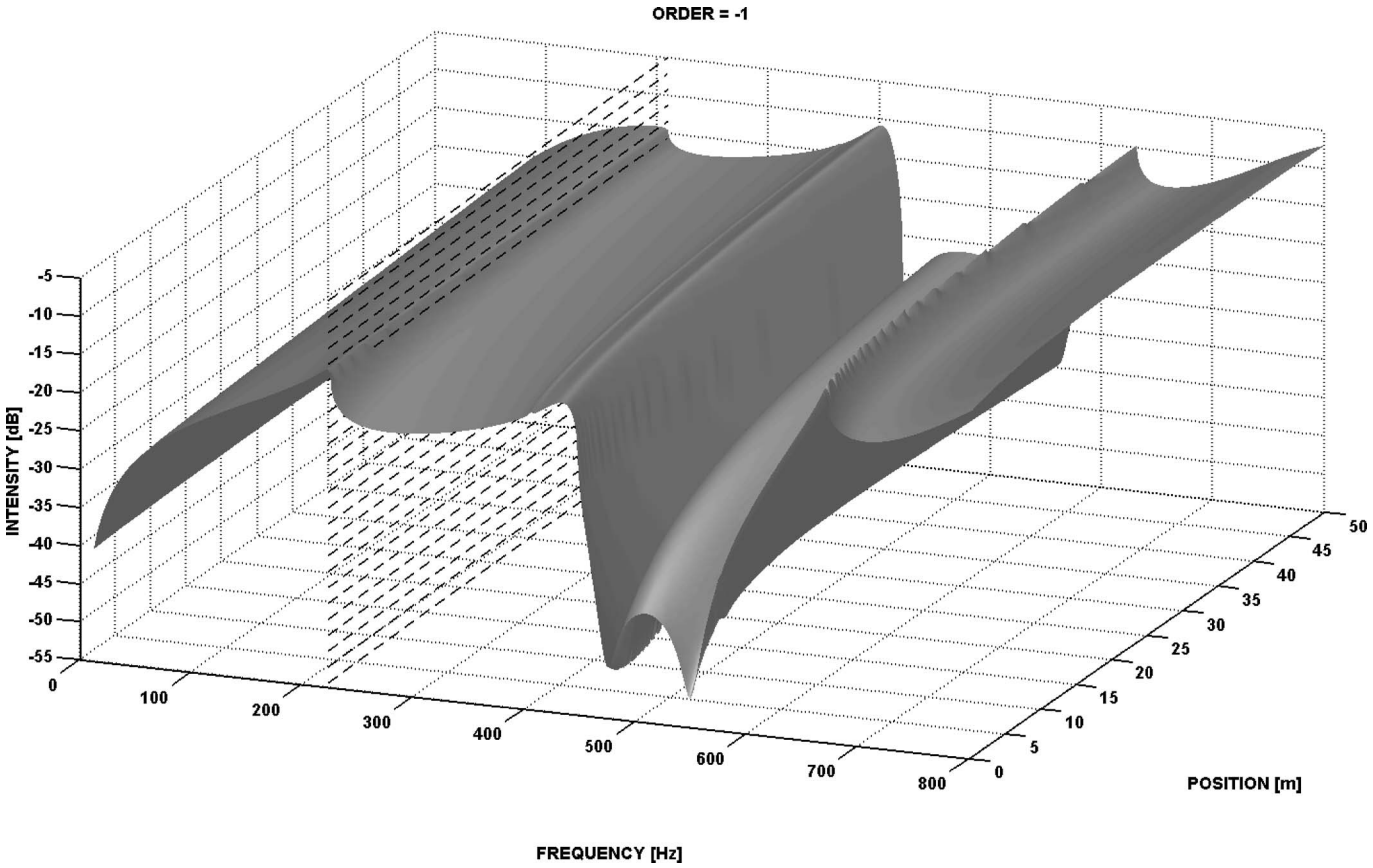


FIG. 9. The diffraction spectrum as a function of frequency and position along the "slope" of the cavea, of the -1 order diffracted sound waves. The raster at approximately 200 Hz indicates the transition between evanescent sound and propagating sound. Note that the amplitudes at frequencies beyond 200 Hz are very small: −15 dB and much less.

increased observed intensity. Contrary to lower frequencies, here the diffraction effect plays a constructive effect.

Besides negative diffraction orders, there are of course also positive diffraction orders involved at Epidaurus. These orders, however, are evanescent throughout the entire considered frequency interval and are therefore never observable by the audience.

VII. CONSEQUENCES OF NUMERICAL RESULTS FOR EPIDAURUS

We have shown in Sec. VI that the most important effect caused by the seat rows at Epidaurus is the effect of relative amplification of a frequency band above 530 Hz. In this section, we discuss the consequences of this effect for the acoustics of the theater.

Izenour[3] already pointed out that background noise is very important for the acoustics of a theater. Background noise is extremely important in a modern motorized society, still at the old Epidaurus there were many visitors that must have caused noise too. Furthermore there is also wind, typically[30,31] up to 500 Hz, rustling trees, etc. Most of the noise produced in and around the theater was probably low frequency noise and even if high frequency noise was produced to some extent, it would have been filtered out by the fact that low frequency noise always spans much further in open air than high frequency noise. The presented calculations indicate that a high frequency band is favored at the expense of lower frequencies. This is true for sound produced at the location of the speaker. Sound coming from other directions will be influenced differently. Nevertheless, we have shown that the position of the sound source and its height has no significant influence on the properties of the amplified frequency band. This means that the conclusions hold for noise coming from any direction.

Still, a reduction of the lower frequencies does not only filter out low frequency noise, but it also filters out the fundamental tones of the human voice (85–155 Hz for men, 165–255 Hz for women). This is not dramatic as the human nerve system and brain are able to reconstruct this fundamental tone, by means of the available high frequency information; this is the phenomenon of virtual pitch in the case of a missing fundamental tone.[32–36] As a matter of fact, virtual pitch is the basic effect behind the creation of the illusion of bass in small radios, miniature woofers, and in telephones.[37] In other words the seat rows of the theater filter out low frequency noise which has a positive influence on the clarity of a speaker throughout the theater, despite the fact that the lower tones of the human voice are filtered out as well.

VIII. COMPARISON WITH OTHER CLASSICAL THEATERS

Table IV shows the physical parameters of different ancient theaters.

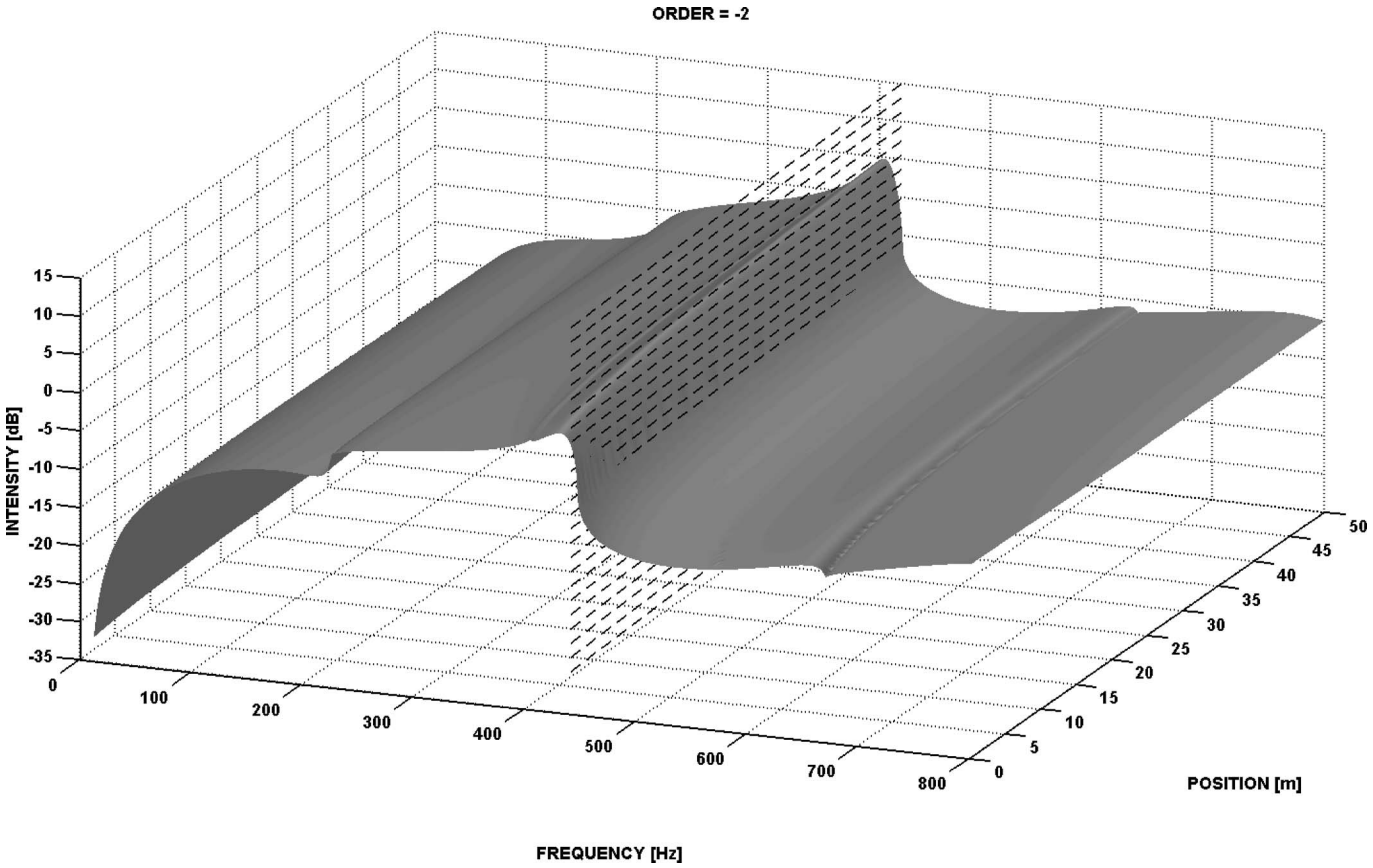


FIG. 10. Comparable to Fig. 9, but for the -2 order diffracted sound. The raster is now situated at approximately 450 Hz. The amplitude for frequencies beyond 450 Hz are −10 dB or higher. It is this -2 order diffracted sound that is responsible for the filter effect and for favoring frequencies beyond 500 Hz for the audience.

Note that most theaters, apart from Pergamon, have a seat row periodicity that is comparable to Epidaurus. The slope values are more scattered. The discussion of our obtained numerical results shows that the periodicity is the key factor for the filtering effect of the stairs. Within that scope, it is not surprising that most theaters copy Epidaurus' seat rows.

TABLE IV. Properties of classical theaters.

Theater	Dated	Location	Style	Λ (m)	α (deg)
Epidaurus[a]	300 B.C.	Greece	Hellenistic	0.831	26.6
Aphrodisias[b]	300 B.C.	Turkey	Hellenistic	0.736	31.1
Aspendos[b]	161–180 A.D.	Turkey	Roman	0.788	33.1
Dionysus[c] (Athens)	400–300 B.C.	Greece	Greek	0.829	23.5
Ostia Antica[b]	19–12 B.C.	Italy	Roman	0.762	22.1
Pergamon[b]	197–159 B.C.	Turkey	Roman	1.657	62.7
Pompeii[d] (Odium)	80 B.C.	Italy	Roman	0.805	32.3
Priene[b]	330 B.C.	Turkey	Hellenistic	0.772	31.2
Side[b]	?	Turkey	Greek	0.749	34.1

[a]See Ref. 1.
[b]See Ref. 38.
[c]See Ref. 39.
[d]See Ref. 40.

Still, the fact that the acoustics of Epidaurus is much more renowned than the acoustics of the other theaters is probably because of the fact that Epidaurus has been renowned from the very beginning (historical reason) and that it has been preserved so well (conservational reason).

Dionysus is the theater whose dimensions best resemble the dimensions of Epidaurus, but is in a much worse condition and therefore it will never be really possible to compare the acoustics of both theaters experimentally.

IX. CONCLUDING REMARKS

It is shown that reflections on the foreground of the theater result in a better distribution of sound throughout the cavea so that all positions become acoustically similar to one another. The installation of seat rows on a smooth cavea generates diffraction effects that change the acoustic properties of the theater.

The intensity observed by the audience will be lower than in the case of a smooth cavea. This is not dramatic because the human ear is capable of adapting its sensitivity. It is more important that the damping effect is frequency dependent: the seat rows act like a filter. For frequencies beyond a certain threshold, second-order diffracted sound plays an important role and causes sound to be backscattered from the cavea to the audience making the audience receive sound from the front, but also backscattered sound from behind. This has a positive outcome on the reception of sound.

For frequencies below the threshold (mostly noise), the effect of backscattering is less important and is to a great extent filtered out of the observed sound. The threshold frequency of the filtering effect is mainly determined by the periodicity of the seat rows in the cavea of the theater. For Epidaurus this threshold is around 500 Hz, which is usually the upper limit for wind noise.[30,31]

The slope of the cavea does not really influence the frequency values where the amplified frequency band appears and there is no significant difference between the acoustics in summer and the acoustics in winter.

[16]Private communication between N. F. Declercq and Jorge Antonio Cruz Calleja (Department of Acoustics, Escuela Superior de Ingeniería Mecánica y Eléctrica UC, Avenida Santa Ana No. 1000 México D. F. Del. Coyoacan. C. P. 04430. San Francisco Culhuacan, e-mail: jorgeacruzc@hotmail.com).

[1]A. Von Gerkan and W. Muller-Wiener, *Das Theater von Epidaurus (The Theater of Epidaurus)* (Kohlhammer, Stuttgart, 1961) (in German).
[2]M. Vitruvii Pollionis, *De Architectura (On architecture)*, book V, Public places, Chap. 3 (1st Century BC).
[3]G. C. Izenour, *Theater Design*, 2nd ed. (Yale University Press, New Haven, CT, 1996).
[4]N. F. Declercq, J. Degrieck, and O. Leroy, "Diffraction of complex harmonic plane waves and the stimulation of transient leaky Rayleigh waves," J. Appl. Phys. **98**, 113521 (2005).
[5]N. F. Declercq, J. Degrieck, R. Briers, and O. Leroy, "Diffraction of homogeneous and inhomogeneous plane waves on a doubly corrugated liquid/solid interface," Ultrasonics **43**, 605–618 (2005).
[6]N. F. Declercq, J. Degrieck, R. Briers, and O. Leroy, "Theory of the backward beam displacement on periodically corrugated surfaces and its relation to leaky Scholte-Stoneley waves," J. Appl. Phys. **96**, 6869-6877 (2004).
[7]N. F. Declercq, J. Degrieck, R. Briers, and O. Leroy, "A theoretical study of special acoustic effects caused by the staircase of the El Castillo pyramid at the Maya ruins of Chichen-Itza in Mexico," J. Acoust. Soc. Am. **116**, 3328–3335 (2004).
[8]P. Ball, "Mystery of 'chirping' pyramid decoded," News@nature.com, 14 December 2004; doi:10.1038/news041213-5.
[9]N. F. Declercq, J. Degrieck, R. Briers, and O. Leroy, "Theory of the backward beam displacement on periodically corrugated surfaces and its relation to leaky Scholte-Stoneley waves," J. Appl. Phys. **96**, 6869–6877 (2004).
[10]Lord Rayleigh, *Theory of Sound* (Dover, New York, 1945).
[11]H. W. Marsh, "In defense of Rayleigh's scattering from corrugated surfaces," J. Acoust. Soc. Am. **35**, 1835–1836 (1963).
[12]A. Wirgin, "Reflection from a corrugated surface," J. Acoust. Soc. Am. **68**, 692–699 (1980).
[13]K. Mampaert, P. B. Nagy, O. Leroy, L. Adler, A. Jungman, and G. Quentin, "On the origin of the anomalies in the reflected ultrasonic spectra from periodic surfaces," J. Acoust. Soc. Am. **86**, 429–431 (1989).
[14]M. A. Breazeale and M. A. Torbett, "Backward displacement of waves reflected from an interface having superimposed periodicity," Appl. Phys. Lett. **29**, 456–458 (1976).
[15]A. Teklu, M. A. Breazeale, N. F. Declercq, R. D. Hasse, and M. S. McPherson, "Backward displacement of ultrasonic waves reflected from a periodically corrugated interface," J. Appl. Phys. **97**, 1–4 (2005).
[17]F. A. Bilsen, "Repetition pitch glide from the step pyramid at Chichen Itza," J. Acoust. Soc. Am. **120**, 594–596 (2006).
[18]G. Wolken, "Theoretical studies of atom-solid elastic scattering—He +LiF," J. Chem. Phys. **58**, 3047–3064 (1973).
[19]M. Neviere, M. Cadilhac, and R. Petit, "Applications of conformal mappings to diffraction of electromagnetic waves by a grating," IEEE Trans. Antennas Propag. **AP21**, 37–46 (1973).
[20]J. L. Uretsky, "The scattering of plane waves from periodic surfaces," Ann. Phys. (N.Y.) **33**, 400–427 (1965).
[21]N. Garcia and N. Cabrera, "New method for solving scattering of waves from a periodic hard surface—Solutions and numerical comparisons with various formalisms," Phys. Rev. B **18**, 576–589 (1978).
[22]W. C. Meecham, "Variational method for the calculation of the distribution of energy reflected from a periodic surface," J. Appl. Phys. **27**, 361–367 (1956).
[23]R. Petit, "Electromagnetic grating theories—Limitations and successes," Nouv. Rev. Opt. **6**, 129–135 (1975).
[24]P. C. Waterman, "Scattering by periodic surfaces," J. Acoust. Soc. Am. **57**, 791–802 (1975).
[25]P. C. Waterman, "Comparison of the T-matrix and Helmholtz integral equation methods for wave scattering calculations—Comments," J. Acoust. Soc. Am. **78**, 804 (1985).
[26]G. Whitman and F. Schwering, "Scattering by periodic metal-surfaces with sinusoidal height profiles—Theoretical approach," IEEE Trans. Antennas Propag. **25**, 869–876 (1977).
[27]A. Wirgin, "New theoretical approach to scattering from a periodic interface," Opt. Commun. **27**, 189–194 (1978).
[28]A. K. Jordan and R. H. Lang, "Electromagnetic scattering patterns from sinusoidal surfaces," Radio Sci. **14**, 1077–1088 (1979).
[29]B. A. Lippmann, "Note on the theory of gratings," J. Opt. Soc. Am. **43**, 408 (1953).
[30]M. R. Shust and James C. Rogers, "Electronic removal of outdoor microphone wind noise," http://www.acoustics.org/press/136th/mshust.htm (1998) (Accessed 11/10/06).
[31]S. Morgan and R. Raspet, "Investigation of the mechanisms of low-frequency wind noise generation outdoors," J. Acoust. Soc. Am. **92**, 1180–1183 (1992).
[32]E. Terhardt, "Zur tonhöhenwahrnehmung von klängen. I. Psychoakustische grundlagen ("On the pitch perception of sounds. I. Psychoacoustic study")," Acustica **26**, 173–186 (1972).
[33]E. Terhardt, "Zur tonhöhenwahrnehmung von klängen. II. Ein funktionsschema ("On the pitch perception of sounds. II. A functional pattern")," Acustica **26**, 187–199 (1972).
[34]J. F. Schouten, "The perception of pitch," Philips Tech. Rev. **5**, 286–294 (1940).
[35]J. F. Schouten, "The residue revisited," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg, (Sijthoff, Leiden, 1970), pp. 41–84.
[36]A. Seebeck, "Beobachtungen über einige bedingungen der entstehung von tönen ("Observations of some conditions related to the origin of tones")," Ann. Phys. Chem. **53**, 417–436 (1841).
[37]E. Larsen and R. M. Aarts, "Reproducing low-pitched signals through small loudspeakers," J. Audio Eng. Soc. **50** 147-164 (2002).
[38]"The Ancient Theater Archive," http://www.whitman.edu/theatre/theatretour/home.htm(link visited October 2006 and January 2007).
[39]Perseus Digital Library Project, http://www.perseus.tufts.edu(link visited October 2006).
[40]C. Campbell, "The uncompleted theaters of Rome," Theater Journal **55**, 67–79 (2003) (The John Hopkins University Press, Baltimore, MD).

Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments

Murray Hodgson, Gavin Steininger, and Zohreh Razavi
[1]*Acoustics & Noise Research Group, SOEH-MECH, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, BC, Canada V6T 1Z3*

(Received 14 September 2006; revised 8 January 2007; accepted 9 January 2007)

Measurements made of the acoustical characteristics of, and occupied noise levels in, ten eating establishments are described. Levels to which diners and employees were exposed varied from 45 to 82 dB(A). From these levels and diner questionnaire responses, the number of customers present and average noise levels to which individual diners were exposed during their visits were estimated. These data, assumptions about the number of talkers per customer, and classical room-acoustical theory were used to deduce talker voice output levels. These varied from slightly above "casual" to "loud." An iterative model for predicting speech and noise levels in eating establishments, including the Lombard effect as described by a new, proposed model, was developed. With the measured noise levels as the target for prediction, optimization techniques were used to find best estimates of unknown prediction parameters—such as those defining the Lombard effect, the number of talkers per customer, and the average absorption per customer—with highly credible results. The prediction algorithm and optimal parameters constitute a novel model for predicting speech and noise levels—and thus speech intelligibility—in eating establishments, as a function of the number of customers, including a proven, realistic model of the Lombard effect. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2535571]

PACS number(s): 43.55.Hy, 43.55.Ka, 43.70.Bk [NX] Pages: 2023–2033

I. INTRODUCTION

Eating establishments (restaurants, bistros, cafeterias, dining rooms, etc.) represent acoustical environments in which verbal communication can be difficult because of excessive noise and reverberation. Excessive noise results from music, equipment noise, human activity, and talking. As noise levels increase, talkers increase their voice levels to compensate and to maintain adequate conditions for verbal communication (e.g., speech-to-noise level difference)—this phenomenon is known as the Lombard effect.

Concern about controlling noise for optimal speech communication in eating establishments has led to a number of initiatives around the world. Moulder[1] conducted a program to develop guidelines for restaurants and cafeterias, to provide quiet areas for hearing-impaired individuals. Measured noise levels in the restaurants surveyed ranged from 55 to 68 dB(A). White[2] investigated the appropriate acoustic environment for an enjoyable meal. She performed acoustical measurements in five eating establishments, finding unoccupied background-noise levels that varied between 41 and 66 dB(A) and occupied levels from 66 to 83 dB(A) with between 10 and 94 customers. Astolfi and Filippi[3] studied the optimal acoustical conditions in pizzerias. They used a head-torso simulator as a speech source at a typical seating position 1 m in front of a receiver to study speech intelligibility and privacy. Acceptable values could be achieved only by decreasing the seating density (the number of seats divided by the floor area) from 1 to 0.2 seat/m. With "normal" voice level and a noise level of 72.6 dB(A), speech intelligibility was poor. A "raised" voice level gave improved intelligibility, but inadequate speech privacy, with a distance of 1.5 m between the tables. They used Gardner's work (reviewed below) to take the Lombard effect into consideration. Kang[4] studied the basic characteristics of speech intelligibility in dining spaces arid how these can be optimized by architectural design. He used a radiosity model and talkers with constant output levels to model eating establishments (EEs). Christie[5] studied objective measures, and their ability to predict a subjectively acceptable acoustical environment in bars, cafes, and restaurants. She found that eating establishments were too loud and "undesirable," due to excessive background-noise levels which were, on average, 57 dB(A) in bars, 65 dB(A) in restaurants, and 58 dB(A) in cafes. She used the reference speech levels from ANSI S3.5-1997[6] for STI calculation. The results suggested that the ANSI speech levels are not representative of "normal," "raised," "loud," and "shouting" levels of speech in eating establishments, and should be higher.

In the above studies, the Lombard effect was only considered by Astolfi and Filippi; White noted that the effect is important, but did not include it in her predictions. Independent of eating establishments, the Lombard effect has been studied by a number of researchers. It was first described in 1911 by Lombard[7] as "the adaptation of speech to overcome the deleterious effects of noise ⋯ a nonlinear distortion which depends on the speaker voice level, the background-noise level and the type of noise." In 1954, Korn[8] found that noise levels below 45 dB do not seriously influence speech power. He suggested that higher noise levels—over 55 dB—influence speech output power and result in a 0.38-dB increase in speech level for every 1-dB increase in the noise level (i.e., a Lombard slope of 0.38 dB/dB). In 1958, Pickett[9] used an anechoic chamber to study the Lom-

TABLE I. Main physical and acoustical characteristics of the ten EEs.

	C1	C2	B1	B2	B3	R1	R2	R3	S1	S2
Length/width/height (m)	26/8.5/3	21/7/3	16/10/4.5	15/8.5/4	9/9/3	9/5/4	7/7.5/3	30/6/4	12/9/3.5	17.5/10.5/4
Volume (m^3)	619	412	692	384	333	176	180	960	297	1176
Surface area (m^2)	584	424	599	314	393	202	215	812	315	876
Floor area (m^2)	221	147	69.5	47	85	30	65	240	99	294
Volume/surface area (m)	1.06	0.97	1.16	1.22	0.85	0.87	0.84	1.18	0.94	1.34
No. of seats	120	100	72	46	70	40	54	126	56	106
Seating density ($1/m^2$)	0.5	0.7	1.0	1.0	0.8	1.3	0.8	0.5	0.6	0.4
$RT_{mid,unocc}$ (s)	0.5	1.0	1.5	1.2	0.9	0.9	0.5	0.8	0.5	0.8
$RT_{mid,occ}$ (s)	0.45	0.74	1.41	1.13	0.75	0.82	0.45	0.76	0.47	0.77
α	0.34	0.19	0.23	0.22	0.16	0.18	0.35	0.24	0.30	0.29
$L_{eq,occ}$ (range) [dB(A)]	69.8 (45–75)	70.4 (58–76)	67.1 (60–77)	69.0 (62–76)	74.5 (62–82)	70.3 (52–79)	70.4 (53–76)	69.2 (47–75)	55.3 (48–67)	59.4 (45–66)

bard effect in free-field acoustical conditions. The vocal efforts of talkers in noises of different levels were monitored. He found an increase of 8 dB in mean vocal effort with an increase in noise level of about 8 dB (i.e., a Lombard slope of 1 dB/dB). Webster and Clumpp[10] investigated how talkers raised their voices in the presence of "ambient thermal noise" of varying level, and varying numbers of other voices. The Lombard slope was 0.5 dB/dB with thermal noise and 1 dB/dB with other voices. Gardner[11] studied the magnitude and rate of vocal output changes in group situations. He found, in auditoria, that for each doubling of the number of people, the total vocal output increased by 6 dB once the audience exceeded 12–15 people, on the assumption that one-third of them talked at the same time. Tang *et al.*[12] studied the Lombard effect by measuring the variation of noise level with the number of occupants in a university staff canteen. From their results, they deduced that occupants began to raise their voices when the background noise level exceeded 69 dB(A). They noted that the effect might not be seen when the number of occupants is less than 50, due to the large difference in individual voice levels, and the "granularity" (i.e., nonuniformity) of where people sit. They also proposed a prediction model based on the assumption that, in the presence of noise, talkers raise their voices to maintain a constant speech-noise level difference, and found good agreement with measurement. Dodd and Whitlock[13] asked 18 children to read from a book in an anechoic chamber while different masking noises were fed to their ears. They found that the overall response was nonlinear, but could be approximated as linear above the base voice level of 53.4 dB, with a slope of 0.22 dB per dB of masking noise. Sato and Bradley[14] found this slope to be 0.82 dB/dB by way of noise measurements in 41 classrooms. Sutherland *et al.*[15] proposed a model for the Lombard effect based on published data. These included data for teachers' voice levels in 18 classrooms, for which the Lombard slope was 1 dB/dB. The model assumed that voice levels begin to rise above a noise level of 35 dB(A). They assumed that the free-field voice level at 1 m would be equal to the energy sum of the teacher's voice level in quiet at 1 m—L_o [about 57 dB(A)]—and the sum of the background-noise level, L_n, and a Lombard constant, K_L. They suggested that values of K_L between 22 and 26 dB accommodate a range of Lombard slopes.

A recent study by Razavi and Hodgson[16] evaluated acoustical environments in typical eating establishments by way of physical-acoustical measurements and customer and employee questionnaires. Occupied noise levels in the EEs during normal operation were monitored throughout the day (further details are given below). Of interest here are typical speech and noise levels, and the Lombard effect, in EEs. Information about these has been deduced by further analysis of the measured occupied noise levels, the application of a novel proposed prediction model, and optimization techniques.

II. EATING ESTABLISHMENTS

Ten typical eating establishments of four different types (restaurants, bistros, cafeterias, and seniors' residence dining rooms), on and off of the University of British Columbia campus, were studied. They were chosen on the basis of convenient location, access through existing contacts, their physical characteristics, and customer demographics. Table I shows the main characteristics of the ten EEs. Their volumes varied from 176 to 1176 m^3, with surface areas varying from 202 to 876 m^2, and floor areas from 30 to 294 m^2. The volume-to-surface-area ratios varied from 0.84 to 1.34 m. The number of seats varied from 40 to 126. The seating density varied from 0.4 to 1.3 seat/m^2.

The eating establishments included two different areas in the University student cafeteria (C1 and C2). In both, the floor was carpeted. In C1, the ceiling was covered with acoustic tiles, whereas C2 had an unfinished ceiling with

exposed wooden beams. Bistros B1 and B2, which serve University student and faculty customers, had large windows, and hard floors and ceilings without any acoustical treatment. The tables were metal with thick tablecloths, the chairs were metal with wooden seats. Bistro B3 had indoor and outdoor areas with very different acoustical environments. The indoor area was surrounded by large windows and contained wooden chairs and tables and an open kitchen area, which had a noisy ventilation fan and loud music. The outdoor area is partially enclosed by glass and concrete walls, has no ceiling, and has loud music. Three restaurants on and off campus were enlisted to include a different clientele. The majority of the wall area of restaurant R1 consisted of windows; the ceiling and floor were hard. The furnishings were also hard, with marble-topped tables and wooden chairs. Restaurant R2 had more acoustical treatment, with carpeted floor and some suspended drapes on the ceiling. The furnishings were wooden tables with padded chairs. Restaurant R3, which served University faculty customers, had a ceiling covered with areas of acoustic tiles concealed behind large white drapes. The walls comprised large windows and painted drywall. The floor was wooden in the main area; however, there were some areas with carpeted floor. The tables were wooden with thick tablecloths, the chairs wooden with padded seats. Finally, the dining rooms in two seniors' residences, S1 and S2, were enlisted to involve more elderly, hearing-impaired customers. Both dining rooms had carpeted floors. The ceiling in S1 was covered with acoustic tiles; in S2 it was of painted drywall. In S1, tables were wooden with upholstered chairs. The furnishings in S2 were wooden chairs and tables.

III. PHYSICAL MEASUREMENTS

A. Procedure

In each EE, reverberation-time and noise-level measurement were made to characterize the acoustical environment, as follows:

(i) Reverberation time (RT): RTs were measured in the unoccupied EE. Measurements of the impulse responses between various source and receiver positions were made using the MLSSA system, involving generating noise bursts fed to an amplifier and omnidirectional-loudspeaker system. Octave-band RTs were recorded from 125 to 8000 Hz at six different locations in each EE. $RT_{mid,unocc}$ was calculated by averaging the 500-, 1000-, and 2000-Hz octave-band frequencies most relevant to verbal communication. Of course, the absorption of the occupants decreases the RT. Diffuse-field theory was therefore used to account for this absorption, and to derive $RT_{mid,occ}$ values, considering the average occupancy (number of customers) in each EE. The average absorption per person was assumed to be 0.5 m^2 (this choice is explained below).

(ii) Occupied-noise levels (L_{occ}): Larson-Davis 700 noise dosimeters, usually hung from the ceiling in the centre of the EE, were used to monitor total, A-weighted noise levels in the occupied EE throughout one day of normal operation (the "time history"). They measured noise due to customer and staff activity and 'physical' noise sources, and indicate typical noise levels to which customers and employees were exposed. Equivalent continuous levels $L_{eq,occ}$ were calculated from the time histories.

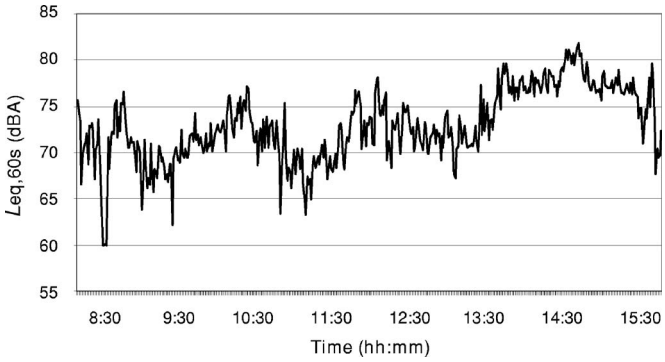


FIG. 1. Variation of $L_{eq,60s}$ with time over a day of operation in EE B3.

B. Results

Table I summarizes the main measurement results. $RT_{mid,unocc}$ varied from 0.5 to 1.5 s. $RT_{mid,occ}$ varied from 0.45 to 1.41 s. Also shown are the average surface-absorption coefficients α calculated from $RT_{mid,unocc}$ and the EE surface areas using diffuse-field theory; values varied from 0.16 to 0.35. $L_{eq,occ}$ varied from 55.3 to 74.5 dB(A); L_{occ} varied by up to 30 dB during the day. Figure 1 shows the time history for the noisiest EE (B3).

IV. FURTHER NOISE ANALYSIS

Measured occupied-EE noise levels, and customer questionnaire responses, were analyzed further, with the following objectives:

- to identify the relationship between the noise level and the occupancy at a given time;
- to investigate the typical voice levels of the talkers;
- to evaluate the typical acoustical conditions for speech intelligibility in EEs by way of signal-to-noise level difference; and
- to investigate how much customers raise their voices with increasing noise level in EEs (the Lombard effect).

In order to identify the relationship between the noise level and the number of customers at a given time in EEs, customer questionnaire responses and noise measurements were subjected to further analysis. The questionnaires asked customers about their visit arrival and departure times, and the number of people in the establishment during the visit. The number of customers present during a given dining period was calculated as the average response to this question of customers who dined during the period. As discussed above, equivalent-continuous sound-pressure levels ($L_{eq,60s}$) had been recorded using noise dosimeters every 60 s during the hours of operation. Equivalent sound-pressure levels $L_{eq,per}$ for each customer dining period were calculated as:

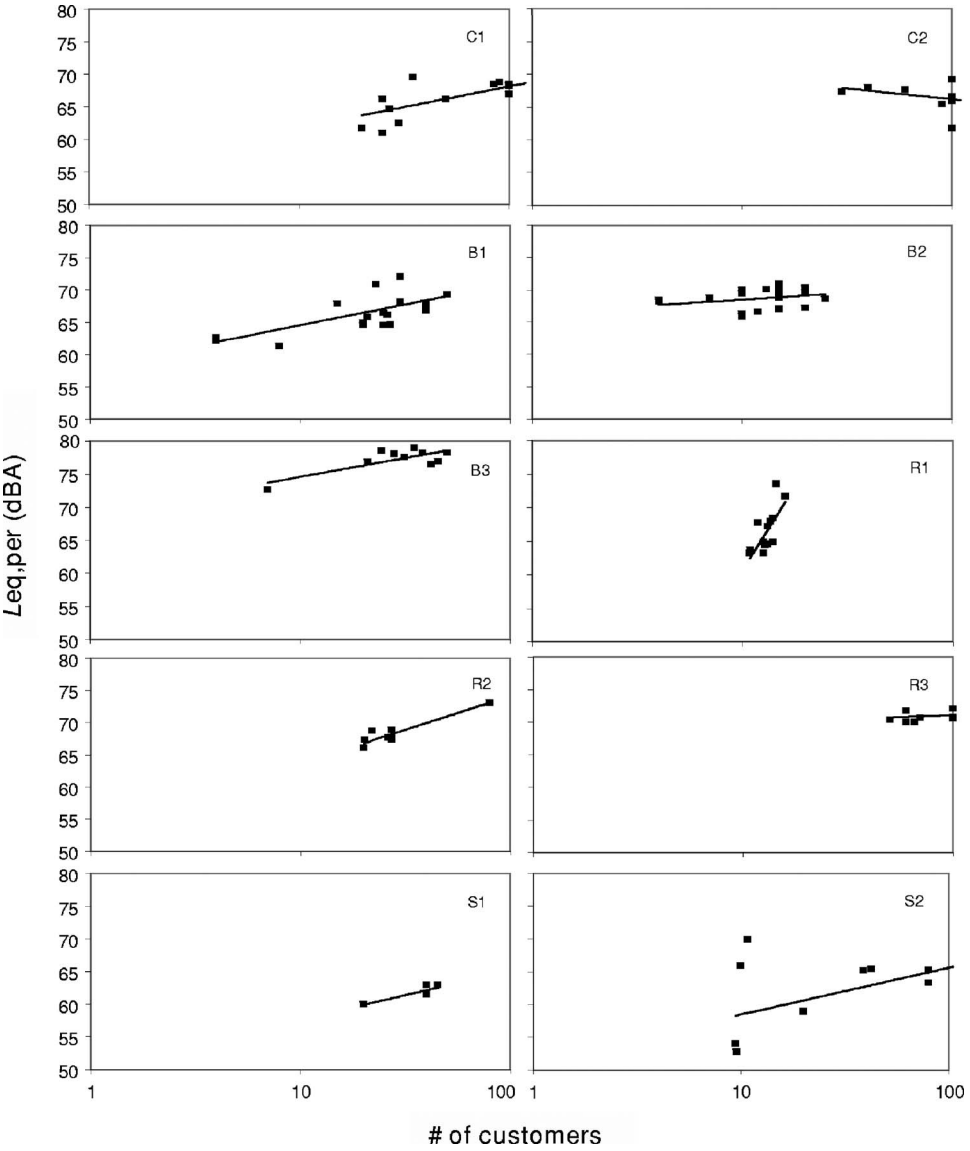


FIG. 2. Variation of $L_{eq,per}$ with the number of customers in individual EEs, and the logarithmic regression lines through the data points.

$$L_{eq,per} = 10 \log \left[\sum_{i=1}^{n} \frac{10^{L_{eq,60s,i}/10}}{T} \right] \text{dB(A)}, \quad (1)$$

where n is the number of $L_{eq,60s,i}$ values measured during the dining period and T is the duration of the dining period in s. For each EE, $L_{eq,per}$ was plotted against the number of customers, as shown in Fig. 2; the pooled data for all EEs are shown in Fig. 3. Also shown in each case is the logarithmic regression line through the data. R^2 values for the individual EEs varied from low (0.05) to quite high (0.85). Noise levels tend to change with an increasing number of customers in all EEs, though with different slopes. The change is positive—levels increase with the number of customers—with the exception of C2. The decrease in noise level with increasing number of customers in C2 could be due to the nature of the EE and its customers. All of the respondents in C2 visited it with the objective of relaxing. The objective of relaxing could cause customers not to communicate with others as much as in the other EEs, and to concentrate mainly on individual work. Other reasons could be an atypical location

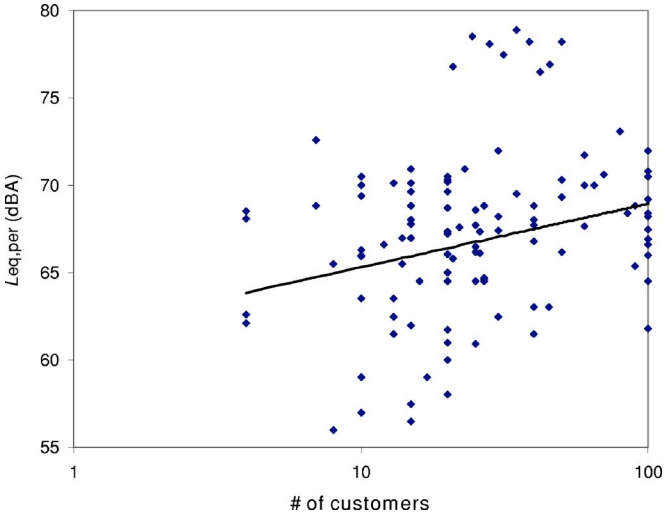


FIG. 3. Variation of pooled $L_{eq,per}$ with the number of customers with all EE data pooled, with the logarithmic regression line.

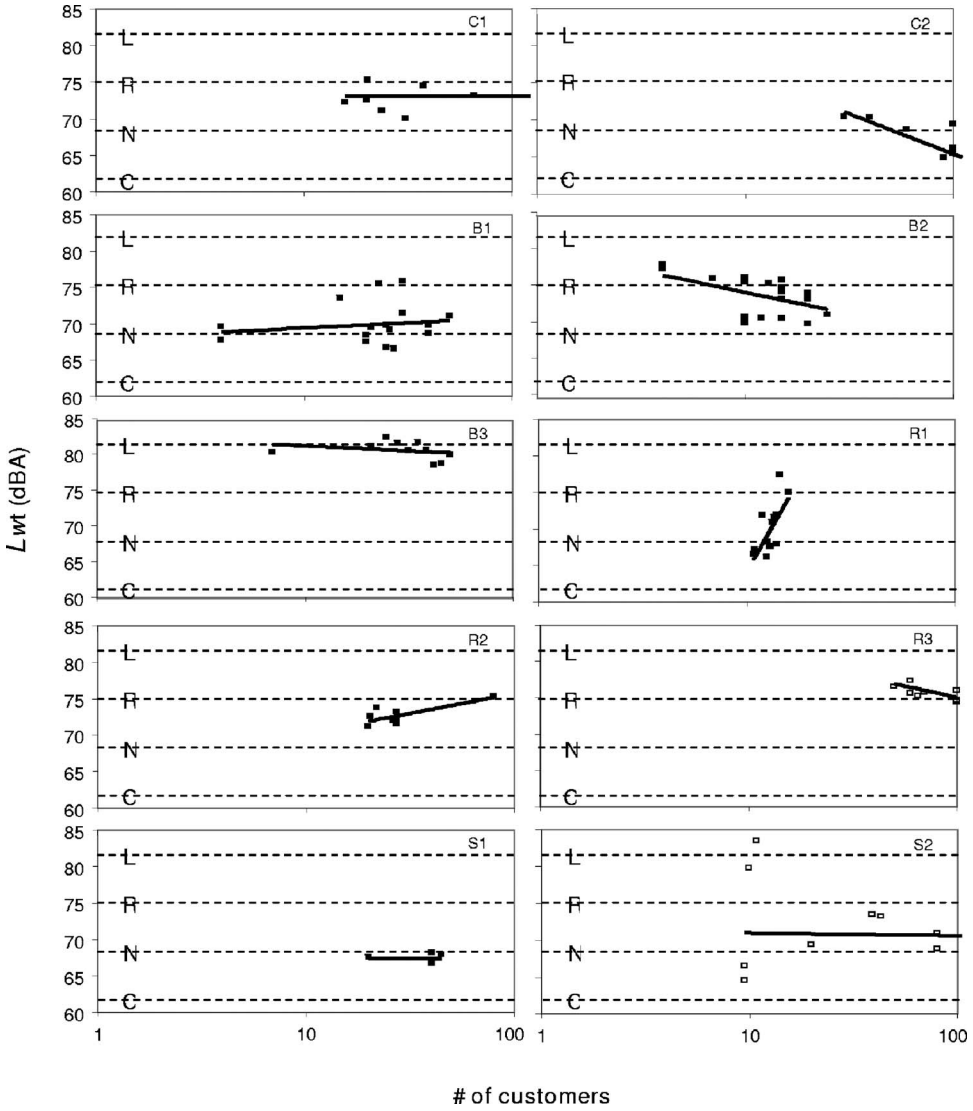


FIG. 4. Variation of voice power level per talker L_{wt} with the number of customers in individual EEs. Also shown are logarithmic regression lines and reference voice levels (C=casual, N=normal, R=raised, L=loud).

of the noise dosimeter, or that not all of the customers were talking at the same time. In the pooled data in Fig. 3, $L_{\mathrm{eq,per}}$ apparently increased slightly with an increasing number of customers, on average. However, due to the different acoustical environments, different occupancies, and different talker voice levels in the different EEs, a large scatter in the data can be observed ($R^2=0.09$).

In order to investigate the typical voice levels of the customers in the EEs, noise levels associated with the customers were calculated as the total noise minus the "physical" noise. That is, the noise level in the unoccupied EE, estimated from minimum levels in the measured time history, was subtracted energetically from $L_{\mathrm{eq,per}}$ to give $L_{\mathrm{eq,corr}}$. The "physical" noise in the unoccupied EE comprised noise from music, EE equipment, its HVAC system, clinking dishes, the movement of chairs, external noise sources, etc. Next, the total sound-power level of the customers $L_{w,\mathrm{tot}}$ was calculated using diffuse-field theory from $L_{\mathrm{eq,corr}}$ on the assumption that the noise-measurement position was far enough from all talkers that only the reverberant field need be considered, as follows:

$$L_{w,\mathrm{tot}} = L_{\mathrm{eq,corr}} - 10\log\left(\frac{4}{R}\right) \quad \mathrm{dB(A)}, \tag{2}$$

in which $R=\alpha S/(1-\alpha)$ is the room constant, with α and S the average absorption coefficient and the total area of the room surfaces, respectively. Under the assumption that all of the customer noise is from talkers' voices, and that the number of talkers is one-third of the number of customers (this choice is explained below), the voice output level—in terms of the sound-power level—of each talker L_{wt} in each dining period in each EE was calculated, as follows:

$$L_{wt} = L_{w,\mathrm{tot}} - 10\log(\text{no. of customers}/3) \quad \mathrm{dB(A)}. \tag{3}$$

The variation of L_{wt} with the number of customers is plotted for each EE in Fig. 4, along with the logarithmic regression lines fit to the data. The associated R^2 values are low, varying from 0.0 to 0.56, generally indicating the lack of a significant trend. Also shown in Fig. 4 are the power levels associated with the reference voice levels (normal, raised, loud, and shouting) from ANSI S3.5-1979.[6] Casual levels, estimated by extrapolation of the slope between the raised and normal levels, are also included. In general, customers

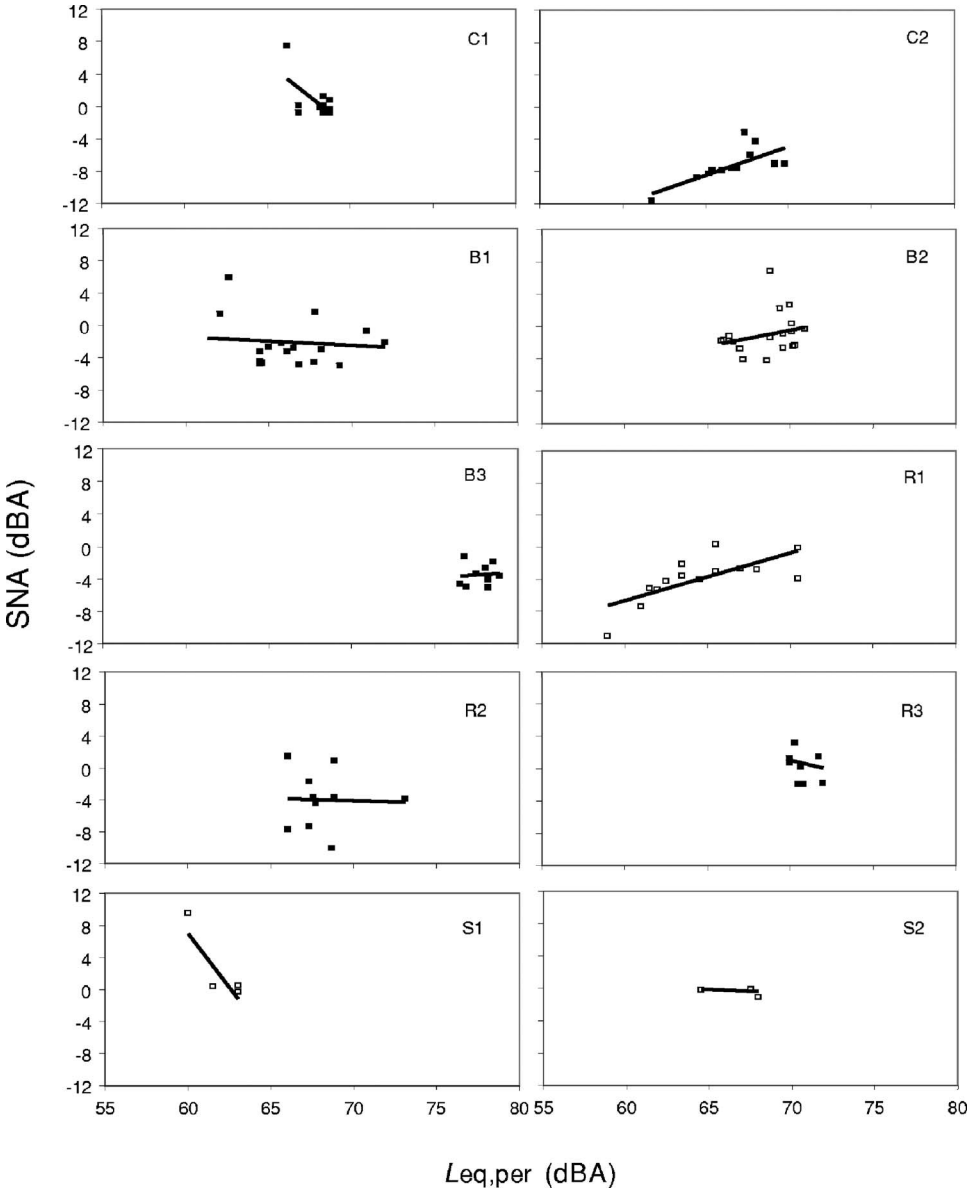


FIG. 5. Variation of *SNA* with $L_{eq,per}$ in individual EEs, with linear regression lines.

used casual to raised voice levels for conversation in EEs. A raised to loud voice level was used in Bistro B3 due to the high background-noise level (due to the loud music). Levels in the S2 seniors' home varied from casual to loud. The pooled data showed no trend and are not presented.

To evaluate the acoustical conditions for speech intelligibility in each EE, speech levels L_s, equal to the equivalent sound-pressure levels at 1 m L_{pff1} were calculated from L_{wt}, as follows:

$$L_s = L_{pff1} = L_{wt} - 20\log(r) = 10\log(Q) - 11 = L_{wt} - 8 \quad \text{dB(A)}. \tag{4}$$

In this equation r, the distance between the talker and the listener, was assumed to be 1 m (a typical table dimension) and the directivity of the talker Q was assumed to be equal to 2.[17] Considering these inputs for Q and r, L_s values are 8 dB less than the corresponding L_{wt} values. A-weighted signal-to-noise level differences SNA at the listener positions were then calculated as the arithmetic difference between the speech-signal level L_s and the noise level L_n, with L_n calculated by energetic subtraction of the L_s from the total level $L_{eq,per}$. SNA values were plotted versus $L_{eq,per}$, and are shown in Fig. 5 for each individual EE, along with the linear regression lines (which, with R^2 varying from 0 to 0.67, show few trends). SNA varied from −12 to +10 dB(A). EE-average values varied between −2.7 in R3 and +2.1 in S2. This range of signal-to-noise level differences is lower than the minimum of 5 to 6 dB(A) that is required for face-to-face talking when facial expression and gestures contribute to intelligibility.[11] Figure 6 shows the variation of SNA with $L_{eq,per}$ with all data pooled, and the linear-regression line through the data. Note that, in Figs. 5 and 6, calculated SNAs occasionally had unrealistic values [generally outside the range −15 to +15 dB(A)], which were omitted. Though there is considerable scatter in the data (R^2=0.08, indicating no trend), on average (see Fig. 6), SNA apparently increases slightly (at 0.13 dB per dB) with increasing noise level. That is, on average, the increase in talker voice output level apparently very slightly over-compensates for the increase in noise level.

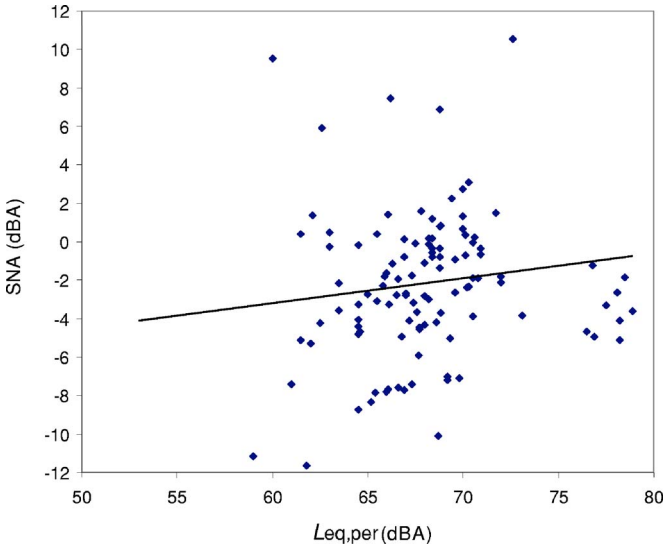


FIG. 6. Variation of *SNA* with $L_{eq,per}$ in ten EEs, with linear regression line.

To investigate the Lombard effect in the EEs, L_{wt} was plotted against $L_{eq,per}$ for the pooled data, as shown by the best-fit linear-regression line in Fig. 7. The associated R^2 is 0.56, indicating a significant trend. Clearly, customers needed to raise their voices to overcome noise in the EEs. The corresponding Lombard slope is 0.69 dB/dB. This compares well with the Lombard slopes of 0.2 to 1 dB/dB reported in the literature.[7–15]

V. PREDICTION MODEL

A model has been developed for predicting speech and noise levels, and the acoustical conditions for verbal communication, in EEs including the Lombard effect. Consider an EE with dimensions L, W, and H, surface area S, average surface-absorption coefficient α, room constant $R=\alpha S/(1-\alpha)$, and uniform physical background-noise level BNL, as illustrated in Fig. 8. A number n_t of pairs of talkers and listeners visits the EE. They are in each others' reverberant fields (the contributions of the direct fields are assumed to be

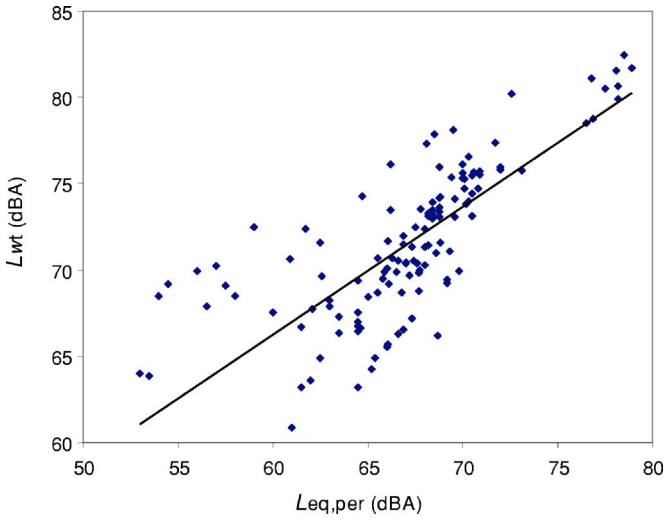


FIG. 7. Variation of talker vocal power level L_{wt} with $L_{eq,per}$ in all EEs, and linear regression line.

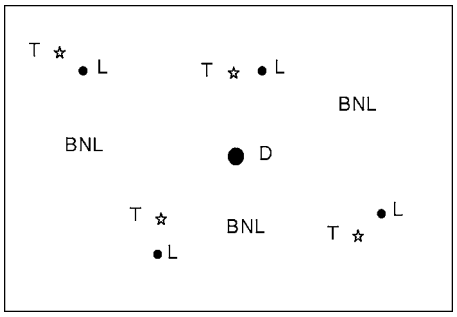


FIG. 8. Sketch of the floorplan of an eating establishment with a uniform background-noise level BNL, showing four talker (T)–listener (L) pairs and the noise-dosimeter position (D).

negligible). The talker and listener in each pair are 1 m apart, and face one another. The listener is only in the direct field of the talker (i.e., the contribution of the reverberant field is assumed to be negligible). The EE may also contain other people, such that the total of customers, $n_c=\mathrm{CPT}n_t$, where CPT is the number of customers per talker. The EE also contains a noise dosimeter (D in Fig. 8) that monitors noise levels; it is in the reverberant fields of all talkers.

In the absence of background noise, each talker would talk with a voice output level—here defined by the total, A-weighted free-field sound-pressure level at 1 m—equal to $L_{pff1,q}$. However, each talker in fact experiences background-noise level $L_n=\mathrm{BNL}$ and, therefore, talks in a louder voice output level $L_{pff1,n} \geqslant L_{pff1,q}$ due to the Lombard effect. It is hypothesized here that the Lombard effect occurs such that the voice output level $L_{pff1,n}$ varies with the background-noise level L_n as follows:

$$L_{pff1,n} = L_{pff1,q} + \frac{asym}{\{1+\exp[(xmid - L_n)/scale]\}} \quad \mathrm{dB(A)}, \tag{5}$$

in which *asym*, *xmid*, and *scale* are Lombard-effect parameters, assumed unknown *a priori*, as is $L_{pff1,q}$. Figure 9 illustrates the behavior of this proposed model. The model assumes that voice output levels vary between a minimum of $L_{pff1,q}$ and a maximum of $L_{pff1,max}=L_{pff1,q}+asym$. Furthermore, it assumes that voice output levels vary between these two limits with a slope of $asym/(4scale)$ dB/dB (in fact, this is the slope at $L_n=xmid$). As explained in Sec. IV, under the assumptions made the corresponding voice power level $L_{wt}=L_{pff1,n}+8$. The speech level at each lis-

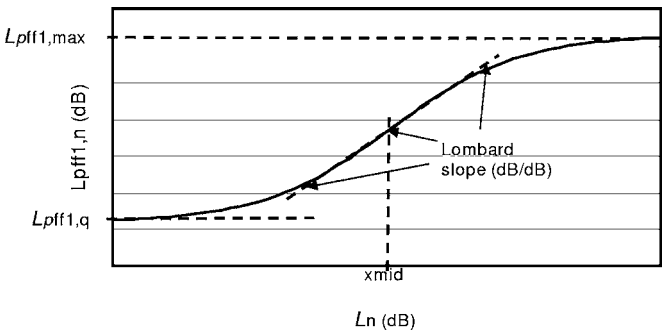


FIG. 9. Proposed Lombard-effect model used in the study.

TABLE II. Constraints used in the "constrained" optimization cases (MV=measured value).

	$L_{pff1,q}$	BNL	*asym*	*xmid*	*scale*	L	W	H	α	C_{rf}	CPT	A_p
Lower limit	48.0	MV*0.9	0.0	0.0	0.0	MV−1	MV−1	MV−1	MV−0.07	−5.0	1.0	0.1
Upper limit	70.0	MV*1.1	∞	∞	∞	MV+1	MV+1	MV+1	MV+0.07	5.0	6.0	1.0

tener is $L_s=L_{pff1,n}$. Each talker generates a pressure level $L_{rev,t}=L_{wt}+10\log(4/R)+C_{rf}$. This expression fundamentally assumes that the room sound field is diffuse; however, C_{rf} is added as a reverberant-field correction that corrects the reverberant level for nondiffuseness of the sound field. The total level L_D at D is the decibel sum of BNL and the total reverberant level from all talkers:

$$L_D = \mathrm{BNL} \oplus [L_{rev,t} + 10\log(n_t)], \tag{6}$$

in which $\oplus$ indicates decibel addition. Now each talker experiences a noise level L_n, which is the decibel sum of BNL and the total reverberant level of the other n_t-1 talkers:

$$L_n = \mathrm{BNL} \oplus [L_{rev,t} + 10\log(n_t - 1)]. \tag{7}$$

This is also the noise level L_n at each listener, so they experience total, A-weighted signal-to-noise level difference $\mathrm{SNA}=L_s-L_n$. The increased background-noise level experienced by the talkers causes them to raise their voices further, as described above, and so on.

An iterative algorithm based on this model was first programmed in a spreadsheet for testing. In general, it was found that, for typical EEs and numbers of customers, predicted levels converged to within 0.1 dB within about ten iterations. The algorithm was also programmed in R[18] in preparation for the optimization work.

VI. OPTIMIZATION

A. Objectives and procedures

The prediction model proposed above can be used to model speech and noise levels in an EE containing any number of customers/talkers, including the Lombard effect. However, it contains a number of parameters for which the values for EEs are not known. These include the parameters defining the Lombard effect—$L_{pff1,q}$, *asym*, *xmid*, and *scale* in Eq. (5), assumed the same in all EEs—and parameters defining the EE, some of which can be assumed to be approximately the same in all EEs (A_p, CPT) and some which would be expected to vary from EE to EE (BNL, L, W, H, α, C_{rf}). For each of the ten EEs described above, data derived from measurements in them (the average dining-period noise level $L_{eq,per}$ and the sound power per talker L_{wt}) were input as the target outputs of the prediction model. Optimization techniques were used to estimate the values of the 12 input parameters for each of the EEs. Note that values of several of the EE parameters—in particular, L, W, H, and α—are nominally known, since they were measured. Thus, the option exists to use the nominally known values in the optimization, or to allow them to vary and be found. In the latter case, the expectation is that the optimization procedure would find values close to the measured values.

There are four cases for which the optimal values of the 12 parameters of the prediction model were found: "fixed, constrained," "fixed, unconstrained," "constrained" and "unconstrained." The term fixed refers to the use of the nominally known values of the parameters L, W, H, and α; these parameters were held constant at the measured values in the fixed cases. The term constrained refers to cases when all parameters had their ranges restricted to values that were considered plausible. Table II lists the constraint ranges used.

The optimal values of the 12 parameters were found using two iterative optimization algorithms, implemented in R: the conjugate-gradient algorithm (CG) and the quasi-Newton algorithm (BFGS). Both of these algorithms work by improving on an initial guess (or starting value) using a linear or quadratic approximation of the function. Table III presents the starting values used for the ten EEs. Several different techniques were used to choose the starting values. L, W, H, and α were chosen to be the measured values. BNL was set equal to minimum level in the measured time histories. L_{pff1q}, C_{rf}, CPT, and A_p were chosen heuristically. *aysm* and *xmid* were estimated from preliminary work with the Sutherland, Lubman, and Pearson Lombard-effect model.[15] *scale* was chosen to take a value larger than it was reason-

TABLE III. Starting values used in the optimization procedure for the ten EEs.

Name	$L_{pff1,q}$	BNL	*asym*	*xmid*	*scale*	L	W	H	α	C_{rf}	CPT	A_p
C1	55.0	48.0	28.5	72.5	15.0	26.0	8.5	3.0	0.34	0.0	3.0	0.5
C2	55.0	60.0	28.5	72.5	15.0	21.0	7.0	3.0	0.19	0.0	3.0	0.5
B1	55.0	61.0	18.4	68.0	15.0	16.0	10.0	4.5	0.22	0.0	3.0	0.5
B2	55.0	63.0	18.5	68.0	15.0	15.0	8.5	4.0	0.22	0.0	3.0	0.5
B3	55.0	67.0	18.4	68.0	15.0	9.0	9.0	3.0	0.16	0.0	3.0	0.5
R1	55.0	58.0	28.5	72.5	15.0	9.0	5.0	4.0	0.18	0.0	3.0	0.5
R2	55.0	57.0	28.5	72.5	15.0	7.0	7.5	3.0	0.35	0.0	3.0	0.5
R3	55.0	57.0	28.5	72.5	15.0	30.0	6.0	4.0	0.24	0.0	3.0	0.5
S1	55.0	46.0	28.5	72.5	15.0	12.0	9.0	3.5	0.30	0.0	3.0	0.5
S2	55.0	49.0	28.5	72.5	15.0	17.5	10.5	4.0	0.29	0.0	3.0	0.5

TABLE IV. Individual values of the parameters for the ten EEs found by the "constrained" optimization procedure, as well as for $L_{p\mathrm{ff1,max}}$, *Lslope*, and *error* (see text).

EE	$L_{p\mathrm{ff1,q}}$	BNL	*asym*	*xmid*	*scale*	$L_{p\mathrm{ff1,max}}$	*Lslope*	*L*	*W*	*H*	α	C_{rf}	CPT	A_{p}	*error*
C1	59.9	52.8	17.5	73.6	3.2	77.4	1.37	26.1	9.2	3.9	0.32	−2.1	1.1	0.5	5.3
C2	55.4	63.6	28.2	73.7	8.4	83.6	0.84	21.4	7.7	3.8	0.20	−1.1	3.3	0.6	4.0
B1	55.7	60.0	18.9	67.2	5.6	74.6	0.84	15.5	9.3	3.6	0.15	−1.3	2.8	1.0	8.1
B2	57.5	60.9	19.2	64.6	5.3	76.7	0.91	14.8	8.1	3.5	0.15	−0.8	4.7	1.0	4.7
B3	54.9	69.8	19.4	67.6	12.1	74.3	0.40	9.2	9.2	3.6	0.23	−1.2	1.2	0.1	4.3
R1	53.1	52.3	29.1	70.3	8.4	82.2	0.87	10.0	5.0	3.7	0.15	−0.2	3.9	0.1	6.8
R2	55.3	58.3	27.8	71.4	9.6	83.1	0.72	7.6	8.1	3.5	0.41	−0.2	2.7	0.1	0.8
R3	55.1	56.7	28.3	70.9	7.9	83.4	0.90	30.2	6.3	4.3	0.17	−2.5	3.0	1.0	6.2
S1	57.2	49.9	28.2	74.2	2.7	85.4	2.61	12.8	9.9	4.5	0.23	0.0	2.4	0.1	2.0
S2	55.1	46.2	29.4	69.3	9.4	84.5	0.78	18.1	11.3	5.0	0.36	−3.5	2.7	0.3	8.6

able to assume that it could be found to be. This was because it was observed that the interactive algorithms had more difficulty increasing *scale* than decreasing it.

The objective (or score) function used in the optimization was the sum of the Pythagorean distances between the regression line through the measured $L_{\mathrm{eq,per}}$ values and the corresponding calculated L_{wt} values, and the $L_{\mathrm{eq,per}}$ vs. L_{wt} curve predicted by the model. Two additional special cases occurred during optimization. The first was when the current set of parameters for the prediction model produced nonreal values of $L_{\mathrm{eq,per}}$ and L_{wt}; the second was when the parameters exceeded their constraints. In the first case, the algorithm assigned the output value of the objective function (the error) a value of 150 000. This seemed sufficiently large that the algorithm would select almost any real point to proceed to and, indeed, using this optimization process, the algorithm never found a plateau of nonreal responses. In the second case, the score function was imprisoned in a parabaloid with its center at the starting values in Table III. The parabola's value was only added to the score function if one or more of the parameters were outside the constrained ranges. This technique not only encouraged the algorithm to keep the parameter values within the constraints, but caused the optimization algorithm to reenter the constrained ranges if it started outside of them.

B. Results

Table IV shows the individual values of the parameters for the ten EEs found by the constrained optimization procedure; Table V shows the mean values, averaged over the ten EEs, for the prediction parameters which are not clearly EE dependent, in each of the four optimization cases. The tables also shows the average error value (or score) associated with those values. Columns entitled $L_{p\mathrm{ff1,max}}$ and *Lslope* have been added for ease of interpretation of the results, and are explained below. Note that values that were implausible were not considered during the construction of Table V; implausible values occurred most often with unconstrained optimization. Note also that, on average, the error was lowest for the two fixed cases and was high in the unconstrained case.

In Table IV, BNL, *L*, *W*, *H*, and α varied from EE to EE, as expected. It is of considerable interest to consider how well the optimization methods identified the values of these parameters that were nominally known, since this evaluates the quality of the methods. BNL values were identified within several decibels, entirely plausible given the uncertainty with which BNL can be measured in an EE, and the fact that it likely did not, as was assumed, remain constant throughout the day. The EE dimensions were identified within 1 m, entirely reasonable since the dimensions are average values that effectively assume the EEs are parallelepipeds, which was not the case. Average absorption coefficients were identified within 0.07 (5% to 45%). That optimization identified these known values so well gives considerable confidence in the methods used. As for the unknown parameters, C_{rf} varied from 0 to −3.5 dB, which is an entirely credible range of values, since reverberant levels tend to decrease more rapidly with distance in nondiffuse fields.[19] CPT varied from 1.1 to 4.7 with an average of approximately 3, well within expectation, and explaining the value used in Sec. IV. The absorption per person, A_{p}, varied from 0.1 to 1.0 m^2 with an average of approximately 0.5 m^2, again consistent with previous results.[20,21] $L_{p\mathrm{ff1,q}}$ varied from 53.1 to 59.9 dB(A). *asym*, the decibel difference between the

TABLE V. Mean values, averaged over the ten EEs, for prediction parameters that are not clearly EE dependent, as well as for $L_{p\mathrm{ff1,max}}$, *Lslope*, and *error* (see text), in each of the four optimization case.

Case	$L_{p\mathrm{ff1,q}}$	*asym*	*xmid*	*scale*	$L_{p\mathrm{ffl,max}}$	*Lslope*	CPT	A_p	*error*
Constrained	55.9	24.6	70.3	7.3	80.5	0.84	2.8	0.5	5.1
Unconstrained	55.0	24.8	70.2	6.1	79.8	1.02	2.7	0.4	18.7
Fixed, constrained	55.7	27.1	70.5	7.4	82.8	0.92	2.9	0.5	4.7
Fixed, unconstrained	58. 2	28.4	70.9	8.5	83.4	0.84	3.5	0.6	4.0

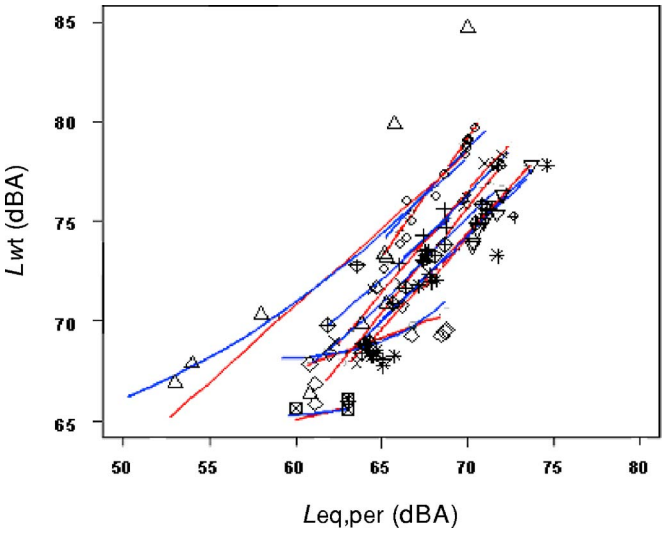


FIG. 10. Predicted variation of L_{wt} with $L_{eq,per}$ showing the individual data points and linear regression lines for each of the EEs, for the constrained case. Different symbols are used for the different EEs.

lowest $L_{pff1,q}$ and highest $L_{pff1,max}$ voice output levels, varied from 17.5 to 29.4 dB(A). Thus, $L_{pff1,max}$ varied from 74.3 to 85.4 dB(A). *xmid* varied from 64.6 to 74.2 dB(A). *scale* varied from 3.2 to 12.1; the corresponding values of *Lslope* $=asym/(4scale)$, the Lombard slope when the noise level equals *xmid*, varied from 0.4 to 2.61 dB/dB.

In Table V, $L_{pff1,q}$ varied from 55.0 to 58.2 dB(A), or from slightly above a casual voice level to slightly below a normal voice level.[6] *asym* varied from 24.6 to 28.4 dB(A). Thus, $L_{pff1,max}$ varied from 79.8 to 83.4 dB(A), or in the loud voice-level range, *xmid* varied from 70.2 to 70.9 dB(A). *scale* varied from 6.1 to 8.5; the corresponding values of *Lslope* varied from 0.84 to 1.02 dB/dB, within the range of previous determinations.[7–15]

The prediction model can also be evaluated by looking at the variation of L_{wt} with $L_{eq,per}$ that is predicted by the model using the optimal input-parameter values. Figure 10 shows the individual data points and linear-regression lines for each of the EEs, for the constrained case (which had a low error); results for the other cases were similar. As can be seen, the predicted values follow the measured data both in the extent/ranges and in the magnitudes of the values on both axes; moreover, eight of the ten EEs have curves of similar slopes. The differences in the vertical levels of the various curves can be explained by differences in the voice level in quiet, the absorption per person, and the customer-to-talker ratio, all of which might vary from one EE to another. The two EEs with somewhat different slopes were S1 and C1.

It is of interest to consider further the cases for which the optimal values were equal to the limits of the constrained ranges. A_p was the most affected by the constraints. This is likely largely due to scaling problems—the same step size in other parameters would be less likely to bring them to the limits of the constrained ranges. A similar effect of this scaling problem occurred with the α values in the constrained case. Two problems arose in the estimation of these parameters: the first was the scaling problem, the second was a lack a variance estimates or confidence intervals to go with the parameters. To solve the first problem, all of the parameters could have been adjusted such that the ranges between their upper and lower constraints were the same. Bootstrapping (equivalent to simulating data from an empirical estimate of the population distribution[22]) the EE samples and repeating the process could fix the second problem, and give reasonable estimates of the parameters.

VII. CONCLUSION

Measurements made in ten EEs found that levels to which diners and employees were exposed varied from 45 to 82 dB(A). From these levels, estimates of the number of customers, assumptions about the number of talkers per customer, and classical room-acoustical theory, talker voice output levels were found to vary from slightly above casual to loud. A new iterative model for predicting speech and noise levels in eating establishments, including the Lombard effect, was proposed. With measured noise levels as the targets for prediction, optimization techniques were used to find best estimates for the unknown prediction parameters, such as those defining the Lombard effect, the number of talkers per customer, and the average absorption per customer. Resulting values were highly credible. For example, the typical voice output level (free-field pressure level at 1 m) in quiet was about 56 dB(A), the maximum voice level was about 82 dB(A), the Lombard slope was 0.69 dB/dB, and the number of customers per talker was about 3. The prediction equations and optimal parameters constitute a novel model for predicting speech and noise levels—and thus speech intelligibility—in eating establishments, as a function of the number of customers, including a realistic model of the Lombard effect.

There are a number of limitations in the present work. These are mainly associated with the simplified assumptions made in the analysis, including the use of the minimum level in the occupied-EE time histories as the physical noise level and assuming that these levels do not vary, basing the prediction model on diffuse-field theory, and assuming that all of the nonphysical noise is speech. Future work should remove these limitations.

ACKNOWLEDGMENTS

The authors would like to thank the managements of the ten study eating establishments for their participation. Thanks also to all customers and employees who participated in the questionnaire surveys.

[1]R. Moulder, "Quiet Areas in Restaurants," report by Battelle to the U. S. Architectural and Transportation Barriers Compliance Board (1993).
[2]A. White, "The effect of the building environment on occupants: the acoustics of dining spaces," M.Phil. dissertation, University of Cambridge, (1999).
[3]A. Astolfi and M. Filippi, "Good acoustical quality in restaurants: a comparison between speech intelligibility and privacy," *Proc. EuroNoise 2003*, Naples, Italy, S102 (2003).
[4]J. Kang, "Numerical modelling of the speech intelligibility in dining spaces," Appl. Acoust. **63**(12), 1315–1333 (2002).
[5]L. H. Christie, "Psycho-to-building acoustics: are bars, cafes, and restaurants acceptable acoustic environments?" Research Report, Victoria University of Wellington (2004).
[6]ANSI S3.5-1997: *Methods for the Calculation of the Speech Intelligibility*

Index (Acoustical Society of America, New York, 1997).
[7]E. Lombard, "Le signe de l'élévation de la voix [The characteristics of the elevation of the voice]," Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx, **37**, 101–119 (1911).
[8]T. S. Korn, "Effect of psychological feedback on conversational noise reduction in rooms," J. Acoust. Soc. Am. **26**(5), 793–795 (1954).
[9]J. M. Pickett, "Limits of direct speech communication in noise," J. Acoust. Soc. Am. **30**(4), 278–281 (1958).
[10]J. C. Webster and R. G. Clumpp, "Effects of ambient noise and nearby talkers on a face-to-face communication task," J. Acoust. Soc. Am. **34**(7), 936–941 (1962).
[11]M. B. Gardner, "Factors affecting individual and group levels in verbal communication," J. Audio Eng. Soc. **19**(7), 560–569 (1971).
[12]S. K. Tang, D. W. T. Chan, and K. C. Chan, "Prediction of sound-pressure level in an occupied enclosure," J. Acoust. Soc. Am. **101**(5), 2990–2993 (1997).
[13]G. Dodd and J. Whitlock, "Auditory and behavioural mechanism influencing speech intelligibility in primary school children," Proc. 18th International Congress on Acoustics, Kyoto (2004), pp. 3581–3582.
[14]H. Sato and J. S. Bradley, "Evaluation of acoustical conditions for speech communication in active elementary school classrooms," Proc. 18th International Congress on Acoustics, Kyoto II. (2004), pp. 1187–1190.
[15]L. Sutherland, D. Lubman, and K. Pearsons, "Acoustic environment challenges for the unique communication conditions in group learning classes in elementary school classrooms," J. Acoust. Soc. Am. **117**(4, Pt. 2), 2366 (2005).
[16]Z. Razavi and M. Hodgson, "Evaluation and optimal design of acoustical environments in eating establishments," Proc. Inter-Noise 2005, Rio de Janeiro, Brazil (2005), p. 117.
[17]J. L. Flanagan, "Analog measurements of sound radiation from the mouth," J. Acoust. Soc. Am. **32**(12), 1613–1620 (1960).
[18]Created by the R Foundation, www.r-project.org, accessed 29 December 2006.
[19]M. R. Hodgson, "When is diffuse-field theory applicable," Appl. Acoust. **49**(3), 197–207 (1996).
[20]L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics*, Vol. 1 (Applied Science, New York, 1982), Sec. II.6.6.
[21]M. R. Hodgson, "Experimental investigation of the acoustical characteristics of university classrooms," J. Acoust. Soc. Am. **106**(4), 1810–1819 (1999).
[22]B. Efron and R. J. Tibshirani, "An Introduction to the Bootstrap," in *Monographs on Statistics and Applied Probability* (CRC, New York, 1998), Vol. 57.

Azimuthal sound localization using coincidence of timing across frequency on a robotic platform

Laurent Calmes[a)]
Knowledge-based Systems Group, Chair of Computer Science V and Institute for Biology II, RWTH-Aachen University, D-52056 Aachen, Germany

Gerhard Lakemeyer
Knowledge-based Systems Group, Chair of Computer Science V, RWTH-Aachen University, D-52056 Aachen, Germany

Hermann Wagner
Institute for Biology II, RWTH-Aachen University, D-52056 Aachen, Germany

(Received 24 November 2005; revised 20 January 2007; accepted 26 January 2007)

An algorithm for localizing a sound source with two microphones is introduced and used in real-time situations. This algorithm is inspired by biological computation of interaural time difference as occurring in the barn owl and is a modification of the algorithm proposed by Liu *et al.* [J. Acoust. Soc. Am. **110**, 3218–3231 (2001)] in that it creates a three-dimensional map of coincidence location. This eliminates localization artifacts found during tests with the original algorithm. The source direction is found by determining the azimuth at which the minimum of the response in an azimuth-frequency matrix occurs. The system was tested with a pan-tilt unit in real-time in an office environment with signal types ranging from broadband noise to pure tones. Both open loop (pan-tilt unit stationary) and closed loop experiments (pan-tilt unit moving) were conducted. In real world situations, the algorithm performed well for all signal types except pure tones. Subsequent room simulations showed that localization accuracy decreases with decreasing direct-to-reverberant ratio. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2709866]

PACS number(s): 43.60.Jn, 43.66.Pn, 43.66.Qp [AK] Pages: 2034–2048

I. INTRODUCTION

Sound localization is important in many behavioral situations. Examples are conversations among humans, orientation in space by animals and machines, avoidance of predators, and localization of prey. The barn owl has a localization precision of some 3° in both azimuth and elevation (Bala *et al.*, 2003; Knudsen *et al.*, 1979). Humans with their larger ear separation can localize sound sources with a precision of about 1° in azimuth [for a review see Blauert (1997)]. Artificial sound localization systems reach localization precisions in the range of 1°–10° (Birchfield and Gillmor, 2002; Huang *et al.*, 1999; Nakadai *et al.*, 2002; Ward and Williamson, 2002).

There are several ways of constructing artificial sound-localization systems. Engineering approaches mostly involve microphone arrays acting as beamformers [Ward and Williamson (2002); for a summary on beamforming arrays see van Veen and Buckley (1988)]. Such systems are usually computationally intensive in that they have to process a multitude of signals. Other approaches involve cross correlation between microphone pairs (Huang *et al.*, 1999; Nishiura *et al.*, 2002; Svaizer *et al.*, 1997).

Biologically inspired approaches restrict themselves to two inputs, equivalent to the two ears. One advantage of such systems is that computations may be done online with moderate computational costs. Additionally, for practical applications (especially on mobile robots), there is no need for special sound hardware providing more than two inputs.

In biological systems, binaural sound source localization relies on two major cues, interaural level differences (ILD) and interaural time differences (ITD). ITDs arise from the difference in conduction time a sound wave needs to reach the two ear drums. ILDs are caused by the acoustic shadow of the head, attenuating the sound arriving at the eardrum which is farthest from the source [for an overview on spatial hearing in humans, see Blauert (1997)]. Biologically inspired sound-localization systems have either implemented one of these cues [ITD: Albani *et al.*, 1994; Bodden, 1993; Braasch, 2002; Lindemann, 1986a, 1986b; Nix and Hohmann, 2001; Peissig, 1993; ILD: Spence and Pearson 1990], or both: Breebaart *et al.*, 2001; Gaik, 1993; Viste and Evangelista, 2004.

In searching for a simple, but effective algorithm operating online on a robotic platform, we followed Liu *et al.* (2000). These authors had taken a biological approach and had implemented a variant of the Jeffress model (Jeffress, 1948). The Jeffress model works in a frequency-specific manner and has two key elements: delay lines and coincidence detectors. The external ITDs are compensated in the brain by delaying the ipsi- and contralateral signals in delay lines formed by axons. The axon terminals synapse on coincidence-detector neurons, which are units that fire maximally if the inputs from the left and right ear arrive simultaneously. Strong neurological evidence for the realization of

[a)]Electronic mail: calmes@pool.informatik.rwth-aachen.de

 0001-4966/2007/121(4)/2034/15/$23.00

the Jeffress model in nature has been found in birds (Carr and Konishi, 1988, 1990; Parks and Rubel, 1975; Sullivan and Konishi, 1986). In these animals ipsi- and contralateral axons from the nucleus magnocellularis function as the delay lines, while laminaris neurons are the coincidence detectors. The ability to represent ITDs implies that the cells can measure relative time, which is achieved in the auditory system by locking of action potentials to stimulus phase (Sullivan and Konishi, 1984). Owls are specialists in this respect, as they can achieve phase locking at high frequencies (up to 9 kHz). This also implies that the neurons in nucleus laminaris, which are narrowly tuned to frequency, show a cyclic response to ITDs caused by phase ambiguities. These ambiguities are preserved in the auditory pathway up to the lateral shell of the inferior colliculus (ICc LS). It is only starting at the level of the external nucleus of the inferior colliculus (ICx) that neurons are broadly tuned to frequency and respond maximally to a specific ITD. This is achieved by integrating the responses of many narrowly frequency-tuned neurons with the same characteristic delay from ICc LS (Takahashi and Konishi, 1986). While it is clear that there are coincidence detectors in mammals, it is currently debated whether these animals have delay lines at all (McAlpine and Grothe, 2003).

Many successful models have dwelled on Jeffress' ideas [for reviews and discussion see Colburn and Durlach (1978), Joris *et al.* (1998), McAlpine and Grothe (2003), Stern and Trahiotis (1995)]. The most basic implementation of the Jeffress model consists of a correlation of the ear (or microphone) input signals. Such approaches have been used in robotics (Murray *et al.*, 2005). More biologically oriented models perform frequency separation, include inhibition, normalization or thresholding instead of multiplication, reintegrate across frequency, and detect either peaks or troughs [e.g., Albani *et al.* (1994), Cai *et al.* (1998), Colburn and Durlach (1978), Colburn *et al.* (1990), Lindemann (1986a, 1986b)]. Recent complex simulations include precise neuronal models and inhibition (Zhou *et al.*, 2005), to take into account the findings in the mammalian auditory system (McAlpine and Grothe, 2003).

The model by Liu *et al.* (2000) performs an operation similar to the correlation in the frequency domain by exploiting interaural phase differences (IPDs). This causes the same problems with phase ambiguities as in the barn owl. Thus frequency integration has to be performed over the whole frequency range in order to solve these ambiguities. We have modified the Liu *et al.* (2000) algorithm using the "direct" method of frequency integration and implemented it on a robotic platform. The main difference to the original method lies in taking into account the complete three-dimensional coincidence map for azimuth estimation. As with every model based solely on the evaluation of interaural time (or phase) differences, no statement on the elevation or the front/back position of a sound source can be made. Furthermore, it is assumed that the sound wave reaching the microphones is planar (far field assumption), meaning that the interaural time differences are independent of sound source distance.

In Sec. II the mathematical model is described. Section III describes the materials used as well as the experimental setup. In Sec. IV the results of testing the algorithm in ideal, real, and simulated environments are presented. The discussion and concluding remarks can be found in Sec. V. Part of this work has been published in abstract form (Calmes *et al.*, 2003).

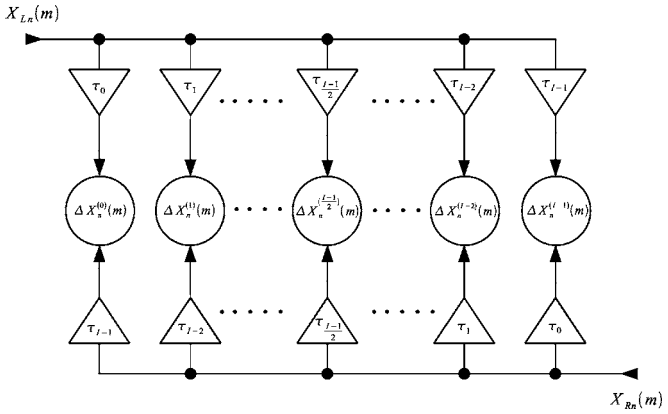


FIG. 1. Dual delay-line structure. $X_{Ln}(m)$ and $X_{Rn}(m)$ represent the spectral values of the mth frequency band of the left and right signals for time frame n after Fourier transformation. τ_i represent the axonal delay elements and $\Delta X_n^{(i)}(m)$ represent the coincidence detector neurons.

II. MATHEMATICAL MODEL

The method described is derived from the dual delay-line algorithm published by Liu *et al.* (2000). Figure 1 shows the dual delay-line structure. This is in essence an implementation of the Jeffress model (Jeffress, 1948). The axonal delays are represented by the triangular delay elements. The coincidence detector neurons are depicted by the circular elements. Note that Fig. 1 shows only one of many frequency bands.

The basic unit of computation of the model is a time frame from a digitized stereo audio signal encompassing N samples per channel. The first step is to transform the current time frame with index n (possibly zero-padded to the Fast Fourier Transform (FFT) size $M \geqslant N$) to the frequency domain using a short-time Fourier transform

$$x_{Ln}(k) \leftrightarrow X_{Ln}(m), \tag{1a}$$

$$x_{Rn}(k) \leftrightarrow X_{Rn}(m), \quad k = 0, \ldots, M-1, \tag{1b}$$

$$m = 0, \ldots, M/2 - 1.$$

Next, delaying in the frequency domain has to be performed. The complex Fourier points for each channel and frequency are delayed by

$$\tau_i = \frac{\mathrm{ITD}_{\max}}{2} \sin\left(\frac{i}{I-1}\pi - \frac{\pi}{2}\right), \quad i = 0, \ldots, I-1, \tag{2}$$

where $\mathrm{ITD}_{\max} = b/c$ is the highest possible interaural time difference given the microphone distance b (20.5 cm are used here) and the speed of sound c (340 m/s). This yields an $\mathrm{ITD}_{\max}$ of 602 μs. $\mathrm{ITD}_{\max}$ corresponds to 90° in azimuth. Thus, by Eq. (2) the azimuthal space is partitioned into I sectors of equal size. Azimuths ranging from $\alpha = -90°$ to $+90°$ are considered. Negative values indicate a

sound source positioned in the left hemisphere, while positive azimuths point to a sound source in the right hemisphere.

The actual delaying is performed by adding a phase shift corresponding to the delay τ_i to the original phase of the input signals in each frequency band

$$X_{Ln}^{(i)}(m) = X_{Ln}(m)e^{-j2\pi f_m \tau_i}, \tag{3a}$$

$$X_{Rn}^{(i)}(m) = X_{Rn}(m)e^{j2\pi f_m \tau_i}, \quad i = 0, \ldots, I-1, \tag{3b}$$

$$m = 0, \ldots, M/2-1,$$

where M is the FFT size, τ_i specifies the delay in s, $f_m = mf_s/M$ is the center frequency of the mth frequency band, and n is the number of the current time frame.

As this operation is performed in the frequency domain, subsample accuracy is achieved without any additional effort, because interpolation between samples is done implicitly. In the time domain, interpolation would have to be done explicitly in order to shift the signals by an amount smaller than one sample. Equation (2) may seem unintuitive as it allows for negative delays. However, this has no practical implications due to the periodic nature of the discrete Fourier transform and thus can be safely ignored, as long as the delayed signals are not meant to be transformed back into the time domain.

The delays are symmetric around the 0-valued delay $\tau_{(I-1)/2}$. For the left channel, the negative internal delays are situated to the left of the midline in the dual delay-line structure, while positive internal delay values are situated to the right. For the right channel, the reverse is true. τ_0 has the value $-\mathrm{ITD}_{\max}/2$, while τ_{I-1} has the value $\mathrm{ITD}_{\max}/2$. Thus, coincidence detection for external negative delays (sound sources positioned to the left) happens at the right side in the delay-line structure, while coincidence detection for positive external delays (sound sources positioned to the right) happens at the left side in the dual delay-line structure. As can be deduced from Fig. 1, the delay value for the right channel corresponding to the point i in the dual delay-line would be τ_{I-i-1}, whereas in Eq. (3b), $-\tau_i$ is used. It can easily be shown by substituting $I-i-1$ in Eq. (2), that $\tau_{I-i-1} = -\tau_i$.

The external time delay at the point i in the dual delay-line structure, which is compensated by the internal time delay, corresponds to

$$\mathrm{ITD}_i = -[\tau_i - (-\tau_i)] = -2\tau_i. \tag{4}$$

As the azimuth space has been partitioned into I sectors by Eq. (2), there exists a linear relationship between the azimuth α_i and the position i of a coincidence detector element

$$\alpha_i = 90 - \frac{i}{I-1}180. \tag{5}$$

In the ideal case, the time domain signals from both channels are identical except for a time difference. This results in identical amplitude spectra in the frequency domain, whereas the time difference leads to a difference in the phase spectra for both channels. Equations (3a) and (3b) induce a phase change in the left and right channels, respectively, for every point i in the dual delay line. At a given point (the coincidence location), the left and right phase spectra will be identical (in the ideal case) or at least minimally different (for signals recorded by the microphones). To detect that point, first a coincidence map is built with

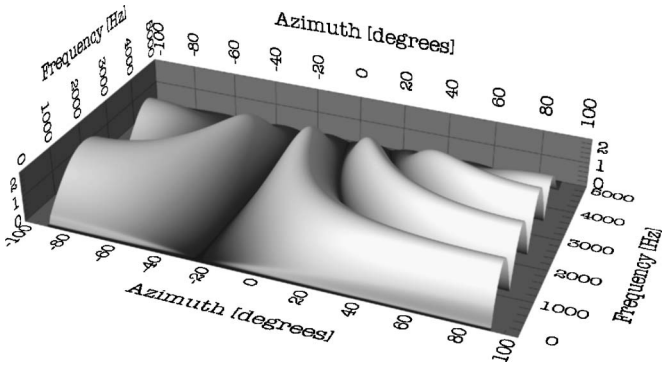


FIG. 2. Example of a three-dimensional coincidence map. It is computed by generating two unit samples (discrete-time Dirac delta function) in software and using them as input to the system. The left channel leads the right channel by an interchannel time difference of four samples corresponding to an azimuth of −24.5° (sampling frequency is set to 16 kHz, microphone distance is set to 20.5 cm). The algorithm returned a value of −24.5°, which corresponds to the frequency-independent minimum at −24.5° azimuth in the graph. *Z*-axis values denote dissimilarity between left and right signals. The lower the value, the higher the similarity at that azimuth. The map was computed using the data from the first time frame of the input signals.

$$\Delta X_n^{(i)}(m) = |X_{Ln}^{(i)}(m) - X_{Rn}^{(i)}(m)|, \quad i = 0, \ldots, I-1, \tag{6}$$

$$m = 0, \ldots, M/2-1.$$

This coincidence map shows the distribution of the response amplitude as a function of both the position in the dual delay-line structure (corresponding to azimuth) and frequency. Since we create a three-dimensional structure, we refer to the map as a three-dimensional coincidence map. As an example, Fig. 2 shows the ideal case derived from computer-generated inputs. As the azimuth α_i depends linearly on the index i in the delay-line structure, the i has been replaced by the corresponding azimuth in Fig. 2 for clarity. The map has a frequency independent minimum at or close to −24.5° of azimuth, which corresponds to the time shifts in the input signals. There are more response minima (caused by phase ambiguities), especially in the high-frequency region, but these minima change their location with frequency. Minima occurring in an azimuth-independent manner over the whole frequency range for a given i specify coincidence location. This is where the method deviates from that described in Liu *et al.* (2000), where the indices of the minima for each frequency band are computed from the map:

$$i_n(m) = \arg\min_i [\Delta X_n^{(i)}(m)]. \tag{7}$$

The coincidence map is integrated over time by performing a running average with time constant β on the coincidence maps computed for all time frames:

$$P_n(i,m) = \sum_{k=0}^{n} \beta^{n-k} \Delta X_k^{(i)}(m), \quad i = 0, \ldots, I-1, \tag{8}$$

$$m = 0, \ldots, M/2-1.$$

This again is in contrast to the algorithm used in Liu *et al.* (2000) where integration over time is done by accumulating the coincidence locations of the minima for each frequency band (here, δ refers to the Kronecker delta function):

$$P_n(i,m) = \sum_{k=0}^{n} \beta^{n-k} \delta(i - i_k(m)). \tag{9}$$

In our algorithm, integration over frequency is performed by summing up the coincidence map at the current time frame index n over all frequency bands

$$H_n(i) = \sum_{m=0}^{M/2-1} P_n(i,m), \quad i = 0, \ldots, I-1. \tag{10}$$

Liu *et al.* (2000) describe two methods for frequency integration. The first (called the "direct" method) is the same as Eq. (10). The second method (called the "stencil" filter) is more complex. While the "direct" method only sums up coincidence locations over frequency corresponding to a position i in the delay line, the stencil filter also takes into account coincidence locations corresponding to phase ambiguities for the index i. This is possible, because the pattern of high-frequency phase ambiguities is unique for each index i. To make this method computationally tractable, a broadband coincidence pattern has to be precomputed, providing the theoretical positions of coincidence locations. As the delay values τ_i vary in a nonuniform manner, the coincidence pattern varies with the index i, thus requiring storage space for I different patterns. To circumvent this disadvantage, Liu *et al.* (2000) chose to use uniform delays, thus requiring only one precomputed theoretical broadband coincidence pattern. A sliding window, centered at the position i in the dual delay line, provides the coincidence positions needed for frequency integration. The tradeoff of this method is, that with uniform delays, the angular resolution across azimuth positions is not constant. Positions close to the midline will have higher angular resolution than more lateral positions, thus requiring a higher number I of coincidence detectors to achieve a resolution equivalent to that obtainable by the direct method. We chose not to implement the stencil filter, because we wanted to keep a constant angular resolution and because the results obtained by using Eq. (10) were sufficient for our purposes.

The final localization function is obtained by normalizing the function $H_n(i)$ at the current time frame index n to the range 0–1 and by transforming the minima into maxima by subtracting from 1,

$$\mathrm{Loc}_n(i) = 1 - \frac{H_n(i) - \min(H_n(i))}{\max(H_n(i)) - \min(H_n(i))}, \tag{11}$$

$$i = 0, \ldots, I-1.$$

To determine the points of coincidence location, the indices i_n^{MAX} of the local maxima of $\mathrm{Loc}_n(i)$ have to be found satisfying the following properties:

$$\mathrm{Loc}_n(i_n^{\mathrm{MAX}}) \geqslant 0.5, \tag{12a}$$

$$\mathrm{Loc}_n(i_n^{\mathrm{MAX}}) > \mathrm{Loc}_n(i_n^{\mathrm{MAX}} - 1), \tag{12b}$$

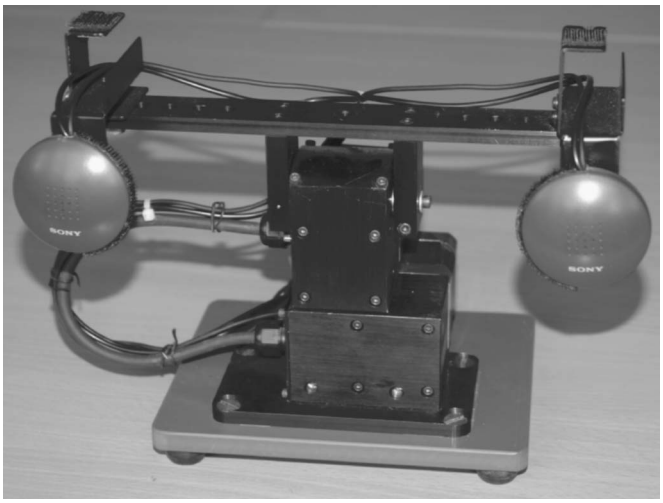


FIG. 3. Pan-tilt unit with microphones.

$$\mathrm{Loc}_n(i_n^{\mathrm{MAX}}) > \mathrm{Loc}_n(i_n^{\mathrm{MAX}} + 1). \tag{12c}$$

The threshold Eq. (12a) is necessary to suppress unwanted side peaks in the localization function which can result from high-frequency phase ambiguities. A value of 0.5 proved to be quite effective in suppressing side peaks not attributable to real sound sources. From i_n^{MAX}, the azimuth to the corresponding sound source can be computed with the help of Eq. (5).

The final output of the localizer is an array of pairs of azimuth with corresponding peak height

$$(\alpha_{i_n^{\max}}, \mathrm{height}_{i_n^{\mathrm{MAX}}}) = \left(90 - \frac{i_n^{\mathrm{MAX}}}{I-1} 180, \mathrm{Loc}_n(i_n^{\mathrm{MAX}})\right). \tag{13}$$

An implementation using Eqs. (7), (9), and (10) was initially tried, but this resulted in strong outliers at ±90° with noisy or nonbroadband signals. Finally the method of integrating the complete three-dimensional coincidence map (instead of coincidence locations) over frequency [Eqs. (6), (8), and (10)] was chosen and this solved the problem.

III. MATERIALS AND METHODS

A. Hardware setup

Throughout the experiments—with the exception of the pan-tilt unit (PTU)—standard, readily available, off-the-shelf components were used. The microphones were two Sony ECM-F8 omnidirectional electret condenser microphones (frequency range: ≈50 Hz–12 kHz), connected to two preamplifiers built around an LF351N op-amp (frequency range: ≈50 Hz–20 kHz; obtained as kit from an electronics supplier). The preamplifiers were connected to the line-in input of the on-board sound chip of a standard PC.

The microphones were mounted on a Directed Perception PTU-46 pan-tilt unit (Fig. 3), controlled by the same computer which was running the localization algorithm. The angular resolution of the PTU (0.0514°) is one order of magnitude higher than the angular resolution of the sound source localizer (0.5°), ensuring that the microphone assembly is able to pan toward the positions indicated by the algorithm.

All the experiments were conducted in a normal office environment, with background noise from computers and ventilation. The sound source (a Sony SRS-57 loudspeaker; frequency range: ≈100 Hz–20 kHz) was placed at a distance of approximately 1 m from the microphone assembly. The loudspeaker output volume was set in such a way that the signals recorded over the microphones (at the 0° azimuth setting) had a maximal amplitude of about −2 dB with respect to the maximal input amplitude of the analog-to-digital converters. This ensured a high signal to noise ratio while avoiding clipping in the input signals.

TABLE I. Signal types used in the experiments.

Signal type	Frequency range
Noise	Broadband
Click	Broadband
Bandpass noise	1–4 kHz
Bandpass noise	500 Hz–4 kHz
Bandpass noise	100 Hz–1 kHz
sine	1.5 kHz
sine	1 kHz
sine	500 Hz

B. Software configuration

The algorithm was implemented in C++ on a Linux OS. All signal processing was done in software. Whenever possible, the algorithm parameters mentioned in the Liu *et al.* (2000) article have been used. Due to hardware and real-time considerations, some parameters had to be changed. The sampling frequency was 16 kHz. Signals were quantized at 16 bits. The FFT size was 2048 points, yielding a frequency resolution of 7.8125 Hz per frequency band. The system was set up with 361 delay elements per delay line. With this configuration, a linear angular resolution of 0.5° is achieved. A value of 340 m/s was used for the speed of sound. The time-integration constant β from Eq. (8) was set to 0.8.

As early versions of the software processed a time frame in about 60 ms (on an AMD Athlon PC clocked at 1.3 GHz), the time frame size was set to 62 ms (992 samples at 16 kHz), with no overlap. In this way, real-time operation was achieved. Even though the latest, optimized version of the software completes the computation in less than 20 ms on a newer computer (AMD Athlon XP 1800+), the time frame size was not reduced, in order to keep the data from later experiments consistent with earlier measurements.

After A/D conversion, the time frames were filtered with a 12th-order Butterworth band-pass filter (passband approximately 100–4000 Hz) and weighted with a Hann window. The 992 samples were then zero-padded to the FFT size of 2048 points.

In the case of click signals, a simple signal detector was used to prevent the algorithm from producing azimuth estimations corresponding to background noise. Before the experiment, 2 s of background noise were recorded. As the click was very short, it could be that the variance of a whole time frame containing a click was still quite low. Therefore, for every time frame, the mean of the subframe (32 samples) variances was computed. The threshold was set to 1.7 (value determined empirically) times the mean of the individual time frame values. During the experiment, a time frame was accepted as containing a signal if the mean of the subframe variances was above the threshold computed from background noise. In that case, localization computations were performed, otherwise the time frame was dropped. The sole purpose of this signal detection system was to ignore time frames that did not contain samples belonging to the click stimulus. We did not intend to develop a signal detector suitable for practical applications.

Time was measured by reading out the processor cycle counter at appropriate locations in the program. By subtracting two readouts enclosing a part of the code which is to be timed and dividing by the clock frequency, accurate timing information was obtained (within the limits of a non-real-time multitasking operating system).

To obtain different sound source positions for the experiments, it was not the source loudspeaker that was moved with respect to the microphone assembly, but rather the microphone assembly was rotated with respect to the loudspeaker. This was done for practical reasons, as the spatial requirements for moving the loudspeaker in an arc from −70° to +70° around the microphones exceed the space available in most office environments. In the following, for simplicity, it will still be referred to a variation of the azimuth of the sound source, but it should be kept in mind that it was actually the microphones that were panned while the sound source position remained fixed.

IV. EXPERIMENTAL RESULTS

Four different types of tests of the algorithm were performed. Tests using the computer generated signals showed the general correctness of the implementation of the algorithm. Tests using signals transmitted via loudspeaker and recorded in open loop conditions demonstrated the robustness of the algorithm in real situations. Additionally, the performance of the algorithm was assessed in closed loop conditions on a robotic pan-tilt unit. Finally, to find an explanation for the poor performance with low-frequency, narrowband signals, room simulations were conducted.

To assess the influence of spectral content of a signal on the localization system, stimulus signals with increasing bandwidth were chosen, from pure tones to broadband noise and clicks (Table I). Because of the bandpass filter employed, broadband here actually refers to a frequency range of 100 Hz–4 kHz.

It should be noted that multiple azimuths can be returned by the system (especially for sine stimuli), although only one stimulus source is present. In this case, the position with maximal height was selected from the azimuths detected in Eq. (13). For sine stimuli, peaks not attributable to the real source position corresponded to phase ambiguities and could in some cases produce the maximal peak height. This resulted in the system localizing phase ambiguities instead of the real source position. For non-sine stimuli, spurious peaks above the threshold imposed by Eq. (12a) were sometimes detected. They were either caused by transient environmental

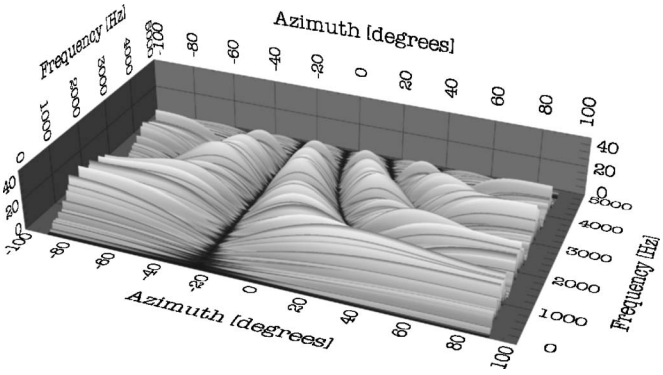


FIG. 4. Example of a three-dimensional coincidence map of the first time frame of a broadband noise signal with a time difference of 4 samples corresponding to an azimuth of −24.5°. The algorithm computes the correct azimuth of −24.5°. Sampling rate is 16 kHz, microphone distance 20.5 cm. *Z* axis represents dissimilarity as in Fig. 2.

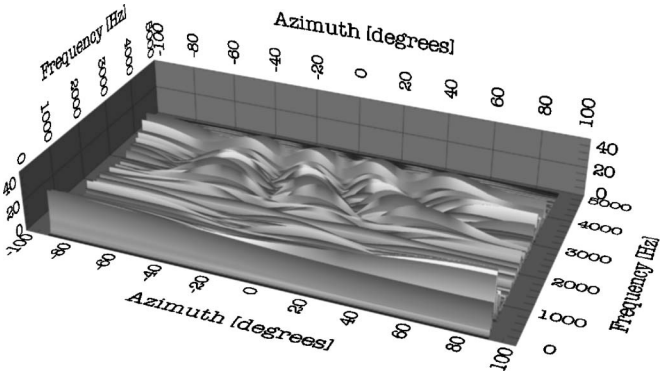


FIG. 5. Example of a three-dimensional coincidence map of a real signal (broadband noise, third time frame), recorded with microphones. Source position was −20°. The value returned by the algorithm is −20°. Sampling rate is 16 kHz, microphone distance 20.5 cm.

noise (e.g., door closing), or they corresponded to phantom sources caused by the room acoustics. In either case they never produced the azimuth with maximal peak height, so that the azimuth estimation from the sound source localizer corresponded to the real source.

A. Tests using computer-generated signals directly

The coincidence map in Fig. 2 was created with two time-shifted unit samples (the discrete-time version of the Dirac delta function) generated in software. The frequency-independent minimum at −24.5° represents the simulated position of the sound source. The output of the algorithm, indeed yielded a sound-source position of −24.5°. In Fig. 4 a similar coincidence map was created, but with broadband noise as stimulus. Again, one minimum, occurring at −24.5° was clearly frequency independent, while all other minima changed their position with frequency. Although the minimum was less well defined than in Fig. 2, the algorithm had no problem finding the position of the sound source.

Tests with many different ITDs and signal types were conducted with computer-generated signals. The algorithm always found the right peak corresponding to the original ITD with an error smaller than 1°. Moreover, the localization estimate remained stable during the whole experiment. Even for sinusoidal stimuli, the correct ITD could be extracted. However, as expected for this type of input, for frequencies above ≈830 Hz, virtual peaks corresponding to phase ambiguities were also detected.

B. Tests using microphone signals: Open loop experiments

In these tests, the computer generated sound was transmitted via a loudspeaker and recorded by a pair of microphones. Stimulus presentation was continuous (with the exception of the click stimulus). The sound-source azimuth remained fixed during a localization run. The output of the algorithm was not fed back to the PTU. Recorded data included source position, number of detected azimuths and the pairs of azimuth with corresponding peak heights from Eq. (13).

As an illustration for a coincidence map of real signals, Fig. 5 shows an example generated by the third time frame of a broadband noise stimulus recorded through the microphones. The source was positioned at an azimuth of −20° with respect to the microphone assembly. Whereas in simulations (cf. Fig. 4), there is a clear, frequency-independent minimum at the source azimuth, the frequency-independent minimum in Fig. 5 is much more diluted.

Figure 6 shows the azimuths as a function of time, for three different signal types and a source position of −60°. The algorithm was able to precisely localize the source at −60° for a broadband stimulus. When the bandwidth is limited to the 100 Hz–1 kHz range, a systematic localization error of some 20° occurred throughout the run. In a similar way, the algorithm gave a stable estimate when the stimulus was a 500 Hz sinusoid. However, it can be seen in Fig. 6, that at about +65°, the localization error is much larger. High-frequency phase ambiguities arise at wavelengths smaller than twice the microphone distance. With a microphone distance of 20.5 cm, this would be the case for frequencies above 830 Hz (speed of sound 340 m/s). Thus, the

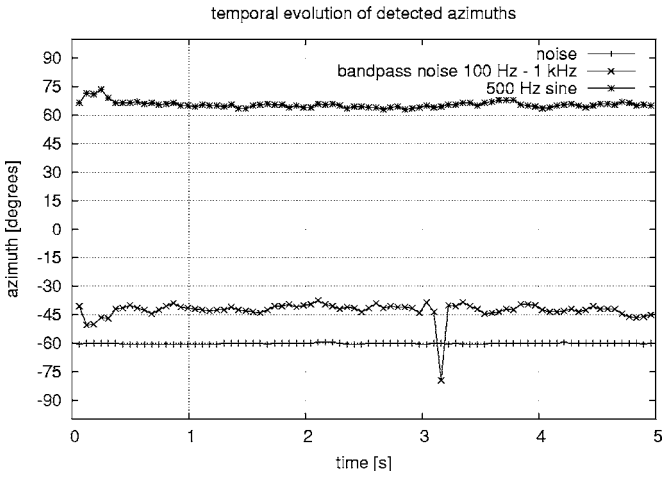


FIG. 6. Temporal sequence of estimated azimuths for three different signal types measured at a source position of −60°. One 80 time frame (≈5 s) run per signal.

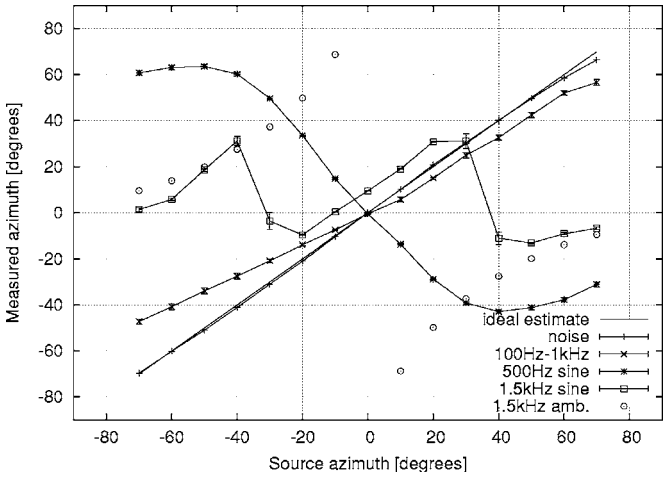


FIG. 7. Averages of measured azimuths for four different signal types (5 runs per signal, 80 time frames per run and azimuth). Error bars indicate 99% confidence interval for the given source position. Note the decrease in accuracy with decreasing bandwidth. For every new localization run, the source is positioned at a different azimuth (−70° to +70° in 10° steps). Open circles indicate the expected locations of phase ambiguities for the 1.5 kHz sine.

mislocalization of the 500 Hz sinusoid (wavelength 68 cm) cannot be caused by the system locking onto a high-frequency phase ambiguity.

Figure 7 shows the averages over five runs for random noise, 100 Hz–1 kHz bandpass noise, 1.5 kHz sine (along with the first phase ambiguities depicted as open circles) and 500 Hz sine stimulus types (80 time frames per run and azimuth for each signal type). Table II shows the minimum, maximum, and the mean of detected azimuths alongside the standard deviations for all the signal types used in the experiments at the −70°, 0°, and +70° source positions (5 runs, 80 time frames per run and azimuth for each signal type). In these representations, the impressions from the examples shown in Fig. 6 manifest themselves: The algorithm performed almost perfectly for broadband noise, and very well also for clicks, although with clicks some variation may be seen (Table II). High-frequency noise (1–4 kHz) could be localized as well as broadband noise, but problems arose with low-frequency, bandpassed noise as manifested by increased standard deviations.

The localization of sinusoids having a frequency of 1.5 kHz shows a periodic curve (Fig. 7). For azimuths close to zero the localization of sinusoids is quite acceptable, but for larger (smaller) stimulus positions a jump occurs. This is a consequence of the algorithm detecting the real peak for small azimuths and virtual peaks (offset by 1 or more periods) for larger azimuths. A similar observation was made with 1 kHz tones (Table II shows only data for ±70° azimuth). This explains the large errors seen in Fig. 7 and Table II for these frequencies and certain azimuths.

C. Tests using microphone signals: Closed loop experiments

During the closed loop experiments, the algorithm produced an estimate of the sound-source position [Eq. (13)] and transmitted this to the PTU that had to rotate toward that position within a time limit of about 5 s (in the following, "run" refers to this time period). As long as the PTU was moving, sound localization was suspended in order to avoid confusion from motor noise, but resumed after the panning movement ceased. This was done by ignoring time frames during PTU movement, so that the algorithm would only "see" time frames during which no movement took place. From the viewpoint of the sound localization system, a single closed loop experiment is actually a sequence of open loop experiments. Nevertheless, the whole system consisting of sensor (sound localizer) and actuator (PTU) can be considered as a closed loop controller due to the sensory feedback to the PTU, which is why we refer to these experiments as "closed loop" in the following.

Stimulus presentation for the nonclick signals was again continuous. Thus, the azimuth of the sound source changed during a run. In addition to the data described in the open loop experiments section, PTU positions with corresponding time stamps were recorded. Time zero was set to the moment the first pan command was issued to the unit. PTU positions were sampled from this moment on until the estimated source position was reached, by continuously requesting the current position from the PTU controller. As can be deduced from Figs. 8–10, the PTU could provide its current position approximately every 0.1 s. In order to let the motor noise reverberations die out, localization was only resumed approximately 0.8 s (value determined empirically) after movement stopped. Due to the time-measurement method employed (see Sec. III B), the software only checked how much

TABLE II. Open loop experiments results (5 runs and 80 time frames per run and azimuth).

		Measured Azimuths			
Signal type	Source	Min	Max	Mean	σ
Noise	−70	−71.50	−66.0	−69.61	0.57
	0	−0.50	0.00	−0.04	0.14
	70.0	65.50	68.00	66.42	0.37
Click	−70.00	−70.50	−69.50	−70.00	0.45
	0	−0.50	0.00	−0.40	0.20
	70.00	−5.00	68.00	53.20	29.10
Noise	−70.00	−70.50	51.50	−68.25	6.92
1–4 kHz	0	−30.50	0.50	−0.17	2.13
	70.00	66.00	69.00	67.19	0.43
Noise	−70.00	−71.50	33.50	−69.24	5.65
500 Hz–4 kHz	0	−0.50	0.50	−0.01	0.07
	70.00	64.50	69.50	66.51	0.49
Noise	−70.00	−90.00	0.00	−47.16	8.30
100 Hz–1 kHz	0	−4.50	3.00	−0.50	1.00
	70.00	0.00	89.00	56.55	10.34
Sinusoid 1500 Hz	−70.00	−90.00	4.50	1.52	5.12
	0	8.00	11.00	9.48	0.97
	70.00	−8.50	−4.00	−6.99	1.56
Sinusoid 1000 Hz	−70.00	−47.50	75.00	−30.34	38.16
	0	8.50	17.00	11.75	2.24
	70.00	−60.00	52.00	46.27	22.17
Sinusoid 500 Hz	−70.00	0.00	70.50	61.09	4.87
	0	−5.00	8.00	−0.60	1.73
	70.00	−87.00	0.00	−31.29	6.76

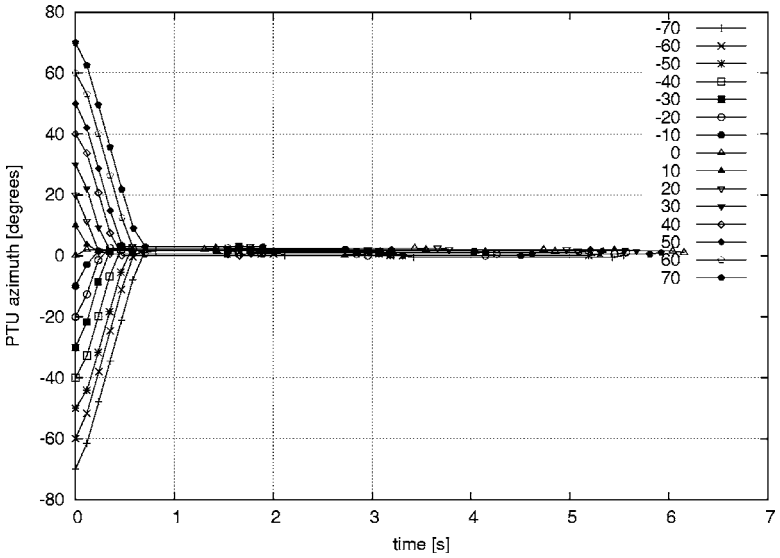


FIG. 8. PTU tracking noise (single run). Inset indicates initial position of sound source.

time had passed after the PTU stopped. This explains the intervals longer than 0.8 s between PTU movements and the run times longer than 5 s in the figures. In the case of clicks, the presentation of the stimulus happened some time after the experiment started. As the signal detector ignored every time frame before the click, the moment zero of the experiment could be well into the 5 s measurement interval, explaining the shorter run times. The run leads to a fixation, if the source moves toward zero azimuth and stays there. The localization precision was estimated by averaging the end positions of several runs. The PTU was able to move the sound-source toward zero azimuth independent of the starting position when the stimulus was broadband noise. This situation is shown in Fig. 8. The data points indicate PTU position and not the azimuth estimates from the localization algorithm. The standard deviation at the end of the run is only slightly larger than the spatial resolution of the algorithm. In general the localization of the click signals was excellent (Fig. 9). However, in some cases larger errors occurred and were not corrected throughout a run, leading to a wrong fixation. This becomes also manifest in the relatively large standard deviation for click signals as shown in Table III, which depicts the minimum, maximum, mean, and standard deviations of the end positions for the tested stimulus types over five runs. These outliers are caused by problems in the signal detector. Usually, only one click was presented for a given start position. However, if the signal detector decided to present a spurious transient (by, e.g., a door slamming shut) to the algorithm, a second click was presented (cf. starting position of 70° in Fig. 9). The panning movement starting at about 2.5 s was caused by the second click in an attempt to bring the PTU toward 0°. With low-frequency noise (100 Hz–1 kHz), an increased standard deviation is seen (Fig. 10 and Table III). Interestingly, signals with a starting position on the left were mislocalized to the right and vice versa.

D. Room simulation tests

To test the algorithm further in different sound field conditions and to learn more about the strong deviations for low-frequency bandpass stimuli (100 Hz–1 kHz, Fig. 7), room simulations were performed. The simulated, empty room consisted of six surfaces (floor, ceiling, walls) and had

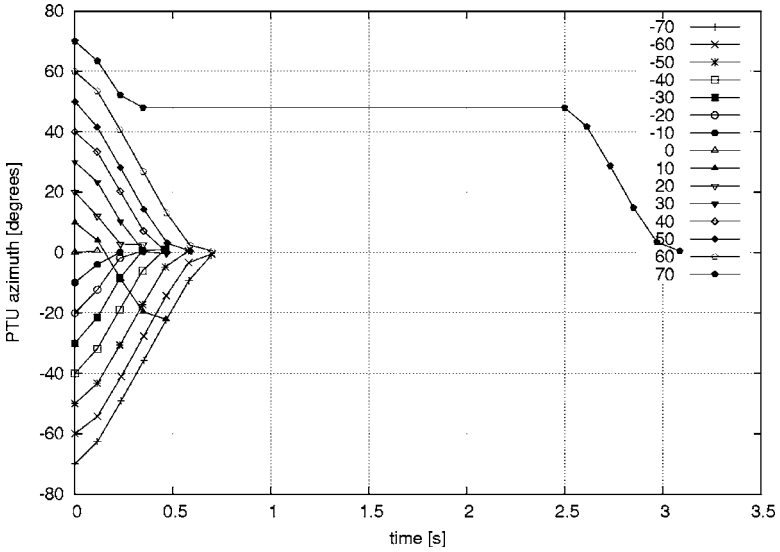


FIG. 9. PTU tracking a click (single run). Click duration is about 180 μs.

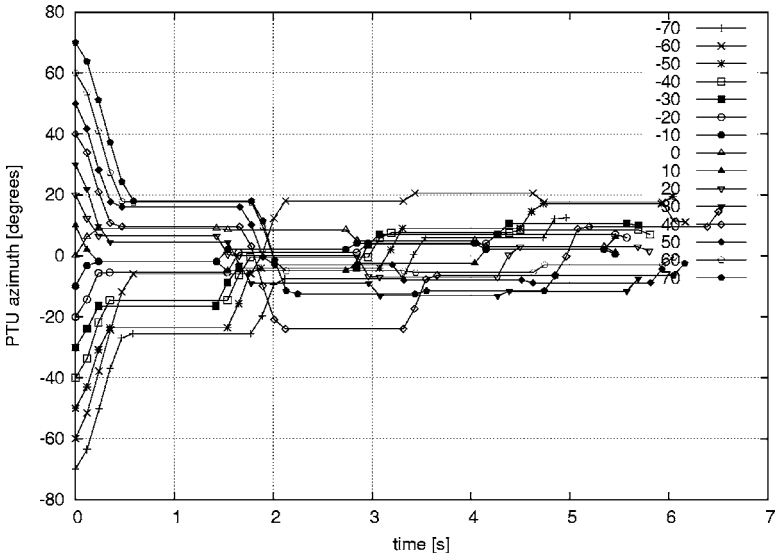


FIG. 10. PTU tracking bandpass noise (100 Hz–1 kHz; single run). Other conditions as in Fig. 8.

the same dimensions as the room in which the real experiments took place. Receiver position and configuration, as well as sound source position were also the same. In addition to the sound source distance of 1 m, a source distance of 3.5 m was simulated to assess the impact of direct-to-reverberant ratio on the localization estimates. As in the real room, the virtual microphone assembly was rotated whereas the source remained at the same position to generate the 15 different sound source positions ($-70° \cdots +70°$ in 10° steps). Three sets of absorption coefficients were used for all six surfaces of the room:

(1) anechoic (total absorption),
(2) 50% (50% absorption), and
(3) unpainted concrete (absorption coefficients corresponding to surfaces made of unpainted concrete).

With the help of the freely available MATLAB program ROOMSIM, 90 impulse responses (2 source distances, 3 sets of absorption coefficients, 15 source azimuths) were generated. These were convoluted with two audio files corresponding to the stimuli used (broadband random noise and 100 Hz–1 kHz bandpass noise), yielding 180 audio files which served as input to the algorithm. The actual parameter values used for generating the room impulse responses can be found in the Appendix.

To assess the impact of noise on localization precision, uncorrelated random noise was mixed into the left and right channels by additive superposition at 11 different signal to noise ratios (+30, +20, +10, +6, +3, 0, −3, −6, −10, −20, and −30 dB). Although this method of noisification does not represent an accurate simulation of noise in a room, it is useful for measuring the sensitivity of the algorithm to the quality of the input signals. Effectively, as the input signals due to the stimulus are gradually drowned in noise with decreasing signal to noise ratio, the correlation also will decrease. At a given signal to noise ratio, it will no longer be possible to produce a reliable localization estimate.

Except for the timing information, the same data as in Sec. IV B was collected. As an example, Fig. 11 shows the result of a simulation of 100 Hz–1 kHz bandpass noise in the room with the absorption coefficients set to "unpainted concrete" and a signal to noise ratio of +30 dB. Note the similarity with the results in Fig. 7 for the same type of stimulus in the real room.

Table IV shows the results for the simulations with random noise at the different signal to noise ratios and for source distance of 1 and 3.5 m. The values were obtained by first computing the difference of the simulated source posi-

TABLE III. Closed loop experiments results (5 runs).

	End positions			
Signal type	Min	Max	Mean	σ
Noise	−0.51	2.01	0.90	0.57
Click	−65.52	16.51	−0.77	8.22
Noise 1–4 kHz	−2.01	2.52	0.25	0.68
Noise 100 Hz–1 kHz	−93.09	26.02	2.19	14.01

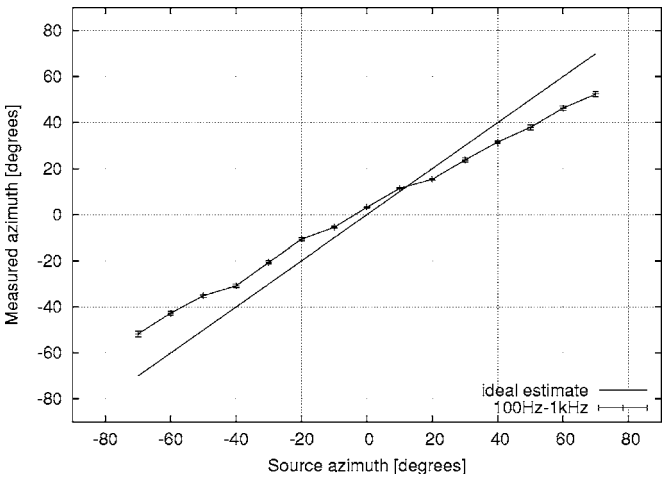


FIG. 11. Averages of measured azimuths for a bandpass noise (100 kHz–1 kHz) in the room simulation (81 time frames per azimuth). Source distance was 1 m. Absorption coefficients for all surfaces were set to unpainted concrete. Signal to noise ratio was +30 dB. Error bars indicate 99% confidence interval.

TABLE IV. Room simulation, broadband noise stimulus with varying signal to noise ratios. Values shown are localization errors obtained by computing the means over all source positions of the absolute values of the differences between "real" source position and the mean (over 81 time frames) of the localization estimates at that source positions.

	Source distance. 1 m			Source distance 3.5 m		
SNR (dB)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)
+30	2	2	2.1	2.3	2.3	6.5
+20	2	2	2.1	2.3	2.3	6.9
+10	2	2	2	2.3	2.3	10
+6	2	2	2.1	2.3	2.3	10
+3	2	2	2.1	2.3	2.3	10
0	2.1	2	2.1	2.3	2.2	15
−3	1.9	2.1	2	2.4	2.2	16
−6	2	2.1	2.8	2.2	2.7	26
−10	2.3	2.7	18	2.9	8.7	31
−20	33	38	36	32	38	42
−30	36	41	34	34	40	30

tion to the mean (over 81 time frames) of the localization estimates for that source position. The mean of the absolute values of the individual errors for each source position yielded the final error value shown in Table IV.

For the source distance of 1 m, the localization error starts to significantly increase at a signal to noise ratio of −20 dB with absorption coefficients set to "anechoic" and "50%." In the case of the unpainted concrete absorption coefficients, a major degradation in localization performance can be observed beginning at a signal to noise ratio of −10 dB.

For the simulations with a source distance of 3.5 m, performance in the anechoic case again worsens at a SNR of −20 dB, whereas a slight increase in localization error can already be observed at −10 dB for the 50% absorption coefficients setting. In contrast to the 1 m sound source distance, the localization error is already quite important at high SNR in the unpainted concrete case and degrades further beginning at −6 dB.

Table V shows the simulation results for the 100 Hz–1 kHz bandpass noise. The errors are generally higher when compared to the broadband noise stimulus shown in Table IV.

In the anechoic case, a major increase in error can already be observed at −10 dB for the source distance of 1 m and at −6 dB for the source distance of 3.5 m. In the 50% case, this already happens between −3 and −6 dB for both distances. The worst case is the one with the unpainted concrete absorption coefficients. Although a major increase in error happens at a lower SNR (at around −10 dB for both distances), this is due to the fact that the initial error at +30 dB is already about three times as high when compared to the anechoic and 50% cases (at both distances).

As a comparison, we computed error values for the data from Sec. IV B in the same way as in Tables IV and V. The error for the broadband noise stimulus was 0.79°. The result for the 100 Hz–1 kHz bandpass noise was 9.34°. Note the similarity of the error of the real-world 100 Hz–1 kHz bandpass noise measurements, carried out at high SNRs, to the

TABLE V. Room simulation, bandpass noise (100 Hz–1 Hz) stimulus with varying signal to noise ratios. Error values obtained the same way as in Table IV.

	Source distance 1 m			Source distance 3.5 m		
SNR (dB)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)
+30	2	2.7	10	2.3	6.7	26
+20	2.1	2.8	10	2.4	6.6	25
+10	1.9	2.7	10	3.1	6.7	25
+6	2.8	2.6	9.9	3.4	6.7	24
+3	2.6	2.5	9	3.5	6.8	25
0	4.8	3.7	9.5	4.7	6.7	28
−3	4.1	5.4	12	5.6	11	28
−6	7.5	11	16	12	16	26
−10	18	24	29	20	26	36
−20	29	35	36	39	40	38
−30	34	40	40	41	39	39

TABLE VI. Room simulation, Liu *et al.* (2000) algorithm (direct frequency integration method). Broadband noise stimulus with varying signal to noise ratios. Error values obtained the same way as in Table IV.

	Source distance 1 m			Source distance 3.5 m		
SNR (dB)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)
+30	2	2.3	10	2.4	8.6	31
+20	2	2.4	12	2.4	10	33
+10	2.1	2.4	16	2.5	11	34
+6	2.1	3.8	22	2.3	16	35
+3	3.1	8.2	28	3.9	23	36
0	8.7	16	29	10	27	37
−3	24	27	32	23	33	38
−6	29	35	36	32	36	37
−10	34	36	37	34	36	37
−20	37	37	38	37	37	37
−30	37	37	35	35	38	38

simulation values for high SNR and a sound source distance of 1 m in the unpainted concrete case (Table V).

Table VI and VII show the results of the room simulations using our implementation of the Liu *et al.* (2000) algorithm with the direct frequency integration method (see Sec. II). Although the issue of the outliers at ±90° mentioned in Sec. II could not be solved, a workaround was found by restricting the localization function $\mathrm{Loc}_n(i)$ [Eq. (11)] to the index range $i=1,\ldots,I-2$. This in effect reduces the available azimuths to the range from −89.5° to +89.5° (with $I=361$), but has the advantage of discarding the unwanted outliers.

Results for the broadband noise stimulus (Table VI) in the anechoic (1 and 3.5 m source distance) and 50% absorption cases (1 m source distance) are similar to those shown in Table IV, with the difference that major decreases in localization accuracy already appear at higher SNR. The 50% case for a source distance of 3.5 m as well as the unpainted concrete case (both source distances) exhibit much higher angular errors than those shown in Table IV.

The angular errors for the bandpass noise stimulus (100 Hz–1 kHz) shown in Table VII are significantly higher than those shown in Table V, except for a SNR of +30 dB in the case of the anechoic absorption coefficients (both source distances).

V. DISCUSSION

It was our goal to implement a robust, but computationally efficient sound-localization system on a robot for application in real-world situations. We did not want to simulate the various aspects of the biological system as was done in other models of binaural hearing (Breebaart *et al.*, 2001; Chung *et al.*, 2000; Jin *et al.*, 2000; Nix and Hohmann, 2001; Zhou *et al.*, 2005). The focus of the present work was on practical applicability (i.e., real-time performance). The mathematical model presented by Liu *et al.* (2000) with the direct method of frequency integration fulfilled our basic requirements. In the following we discuss first our method and the changes to the original algorithm. Next we compare the

TABLE VII. Room simulation using Liu *et al.* (2000) algorithm (direct frequency integration method). Bandpass noise (100 Hz–1 kHz) stimulus with varying signal to noise ratios. Error values obtained the same way as in Table IV.

	Source distance 1 m			Source distance 3.5 m		
SNR (dB)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)	Anechoic (deg)	50% (deg)	Unpainted concrete (deg)
+30	2.4	17	30	2.8	26	35
+20	7.3	22	29	8.2	27	35
+10	22	26	32	22	30	35
+6	26	28	33	29	30	36
+3	29	31	34	30	33	36
0	32	32	34	34	31	39
−3	33	34	35	34	36	38
−6	36	35	35	36	37	37
−10	35	35	37	38	36	37
−20	38	38	37	38	37	40
−30	38	38	39	37	37	38

results of our tests with the localization performance of other systems. Finally we present an outlook for further improvements of our system.

A. Method and control tests

While Liu *et al.* (2000) was a good starting point, we noted that this algorithm, by taking into account only the minima of the coincidence map, does not use all of the information available. We minimized information loss by performing the frequency integration over the whole three-dimensional coincidence map. The modified algorithm produced excellent results without any indications of failures with computer-generated signals. The ambiguities observed for pure tones with a frequency higher than about 830 Hz were expected, given the structure of the algorithm. We chose not to implement the stencil filter method of frequency integration, because then we would have lost the constant angular resolution over the whole azimuth range. Furthermore, we did not want to incur the additional computational overhead associated with the method.

It is difficult to compare the performance of our algorithm with the performance of the original method, because Liu *et al.* (2000) restricted their experiments to simulations and anechoic chamber tests, and mainly conducted multi-source measurements. However, the one-speaker tests conducted in an anechoic chamber by Liu *et al.* (2000) seem to have produced a similar localization accuracy as our open-loop tests in a laboratory environment. Our own tests with the Liu *et al.* (2000) algorithm initially produced outliers at ±90° (using the direct method), which overshadowed the correct source azimuth (if it was present at all) in all cases except high SNR broadband stimuli. By restricting the range of estimated azimuths from −89.5° to +89.5° (thus ignoring the outliers), a workaround was found which could produce usable data. With high signal-to-noise and direct-to-reverberant ratios, the precision is quite good and seems to reflect the data from the original publication. Nevertheless, the Liu *et al.* (2000) algorithm with the direct method of frequency integration showed higher sensitivity to SNR and reverberation. We suppose that this is related to the minimum operation performed on the coincidence values prior to frequency integration [Eq. (7)]. For every frequency band, one minimum is returned, indicating the location of coincidence. This assigns equal weights to all frequency bands. This is not a problem with high SNR broadband signals. But at low signal-to-noise ratios or with narrowband stimuli, giving equal weight to frequency bands containing little or no energy pertaining to the signal seems to seriously corrupt the localization estimate.

One advantage of this type of algorithm is that they can achieve subsample accuracy for interaural delays without requiring explicit interpolation between samples. This is a consequence of carrying out all computations in the frequency domain. Moreover, a high number of frequency bins may be implemented without an increase in the data size. Algorithms working in the time domain are using filter banks for frequency separation [e.g., Roman and Wang (2003)]. These generate a high number of additional signals for the left and right channels, which is computationally intensive. The algorithm presented here allows for efficient frequency filtering and, thus, restricting the computation of the coincidence map to frequency ranges relevant to the intended practical application.

B. Open-loop and closed-loop tests

In the open-loop tests in real-world conditions, performance was excellent for broadband signals, but decreased with narrowing bandwidth of the stimuli. Specifically, problems in the low-frequency range were observed. As initial simulations (cf. Sec. IV A) showed that the software was able to accurately determine the correct azimuth for all signal types, these high localization errors are not due to the algorithm but to reverberation as subsequent room simulations showed (cf. Sec. IV D). Adding an echo-avoidance system (Huang *et al.*, 1999) might improve the situation. These authors used three omni-directional microphones on a mobile robotic platform. The localization was restricted to a single frequency band centered at 1 kHz and with a bandwidth of 600 Hz in order to avoid phase ambiguities. ITDs were computed by the zero crossings of the wave forms from microphone pairs and from these, the direction to the sound source could be computed. The localization accuracy was tested with a 1 kHz sinusoid and a hand-clapping noise. The error for the sinusoid stimulus was within ±1° whereas for the hand-clapping noise, the accuracy was within ±7°. Although this system is able to perform sound localization in three dimensions as well as resolving front-back confusions, this is only possible through the use of three microphones. The restriction to a single frequency band in order to avoid phase ambiguities in the ITD computation seems to be too much of a constraint for practical applications. Additionally, extracting ITDs from several frequency bands with this method would entail a considerable additional computational overhead. In this respect, the algorithm described here is much more robust and suitable for future extensions.

In Nakadai *et al.* (2000, 2002) a frequency-domain algorithm for the sound localization subsystem of the humanoid torso SIG was used. The method performs ITD extraction by directly computing the phase difference between the left and right channels from FFT frequency peaks. Additionally, interaural level differences were included in the azimuth estimation. The error of the sound localization system was within ±5° from 0° to 30° and deteriorated for more lateral positions (Nakadai *et al.*, 2002). Compared to this system, the method proposed here performs better for broadband noise.

The closed-loop experiments were performed to test the algorithm in an environment closer to its later application on a mobile robotic platform. The results confirmed those obtained during the open loop tests. An excellent localization was achieved with broad-band signals. High-frequency signals were localized better than low-frequency signals. This demonstrated that the algorithm may be applicable to dynamic, real-world situations.

Although comparisons are difficult, we have the impres-

sion that our system, despite its simplicity, does not perform much worse in azimuth estimation than microphone arrays with more than two microphones.

Omologo and Svaizer (1994) used 4 equispaced microphones with a separation of 15 cm. Three different localization algorithms were tested, with the so-called crosspower-spectrum phase algorithm providing the best results. For the experiments, 97 stimuli were used with frequency content ranging from narrowband to wideband at various azimuths and distances ranging from 1 to 3.6 m. Half of the stimuli had a noise component with an average SNR of 15 dB. Localization accuracy was 66% with a tolerance $<2°$, 88% with a tolerance $<5°$ and 96% with a tolerance $<10°$.

Brandstein and Silverman (1997) used a bilinear array of 10 microphones with an intermicrophone separation of 25 cm. Their system used a frequency-domain time-difference of arrival estimator designed for speech signals combined with a speech source detector. For experiments with single, nonmoving sources, 18 different source positions were tested. Speech stimuli were used. The angular error was approximately 2.5° over a range of 3 m.

Valin *et al.* (2003) used 8 microphones arranged on the summits of a rectangular prism of dimensions 50 cm $\times 40$ cm $\times 36$ cm. The acoustic environment was noisy with moderate reverberation. The localization system used the crosspower-spectrum phase algorithm enhanced with a spectral weighting scheme. The angular error was approximately 3° over a range of 3 m. The stimuli used for the experiments consisted of snapping fingers, tapping foot, and speaking.

C. Room simulations

The simple "shoebox" room model helped with understanding the acoustic environment in which the real experiments took place. Three conclusions can be drawn from these simulations. First, the algorithm is relatively robust against noise, as important changes in localization error can only be observed beginning at signal to noise ratios between −3 dB in the worst case and −20 dB in the best case. Second, the most important parameter degrading localization performance is direct-to-reverberant ratio. This becomes particularly apparent with the sound source at a distance of 3.5 m from the microphones and highly reflective surfaces (unpainted concrete), where even the azimuth estimation of a broad-band stimulus produces rather large errors. Third, the room simulations suggested that the systematic deviations observed with the low-frequency narrowband noise stimulus were due to room reverberations. This is in accordance with findings that binaural cues vary depending on the acoustic environment and noise conditions (Nix and Hohmann, 2006).

D. Conclusions and outlook

The relatively simple algorithm we used here with online capability performed surprisingly well in real-world situations. It was less sensitive to adverse acoustical conditions than the algorithm by Liu *et al.* (2000) using the direct method of frequency integration, while not being computationally more complex. A further decrease in computation time and memory requirements without a dramatic loss in localization accuracy can be reached by reducing the critical parameters of FFT size and number of delays per delay line. Thus the algorithm is easily adaptable to environments with reduced computational resources such as mobile robots. The current implementation is a good starting point for future extensions. One aspect that will have to be considered is interaural level differences caused by the mismatch of the left and right microphone/preamplifier combinations. Although they are of an identical make, some asymmetries are introduced by manufacturing and preamplifier adjustment tolerances. By Eq. (6) we assume that the only differences between the left and right signals will be phase differences, an assumption that is clearly violated in real environments. Even if in our experiments, localization accuracy was quite high, this mismatch could lead to problems in acoustically more challenging environments. Specifically, source discrimination in the presence of multiple sources could be affected, requiring a mismatch compensation (Liu *et al.*, 2001).

We are currently working on a statistical source tracking module using a Bayes filter, which is expected to increase the robustness against motor noise from the PTU or the robot, among other things. This would make it possible to continue sound-source localization through motor activity of the microphone platform. In the implementation presented here the localizer had to be interrupted during movement. Another extension would be a speech detector and a speech recognizer. The localizer would be working continuously, but its output would be ignored as long as no speech was detected. As soon as there is speech, computation of the coincidence map could be reduced to the relevant frequency components. This may be achieved, because all computations are done in

TABLE VIII. ROOMSIM parameters. Only those parameters differing from the program defaults are shown. Source coordinates are specified relative to receiver coordinates.

Sampling frequency	16 kHz
Room depth (Lx)	4.95 m
Room width (Ly)	3.48 m
Room height (Lz)	4 m
Receiver x position	0.8 m
Receiver y position	1.5 m
Receiver z position	1.0 m
Receiver type	Two sensors
Receiver sensor separation	0.205 m
Receiver sensor directivity	Omnidirectional
Receiver azimuth offset	from −70° to +70° (10° steps)
Source radial distance	1 or 3.5 m
Source azimuth	0°
Source elevation	0°

TABLE IX. Surface absorption coefficients used in the room simulations. Values shown are those provided by the ROOMSIM package.

	Standard measurement frequencies					
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
Anechoic	1	1	1	1	1	1
50%	0.75	0.75	0.75	0.75	0.75	0.75
Unpainted concrete	0.4	0.4	0.3	0.3	0.4	0.3

TABLE X. Estimated reverberation times (s) (RT_{60}) for the different configurations.

	Standard measurement frequencies					
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz
Anechoic	0	0	0	0	0	0
50%	0.08	0.08	0.08	0.08	0.08	0.08
Unpainted concrete	0.216	0.216	0.309	0.309	0.216	0.309

the frequency domain. In this way, localization accuracy of the source could be improved. The directional information could be used to steer a robot closer to the source and/or to perform directional filtering in order to increase signal-to-noise ratio for the speech recognizer.

ACKNOWLEDGMENTS

We thank Albert Feng, Chen Liu, and their co-workers for discussions. This work was supported by the German Science Foundation (LA747/11).

APPENDIX: ROOMSIM SETUP

Tables VIII and IX show the parameter values used in the ROOMSIM program during the room simulation experiments. Table X shows reverberation times for the different simulation setups, estimated using the Eyring formula (Eyring, 1933).

Albani, S., Peissig, J., and Kollmeier, B. (**1994**). "Echtzeitimplementation und Test eines binauralen Lokalisationsmodells ("Realtime implementation and test of a binaural localization model")," in *Fortschritte der Akustik-DAGA 1994* (DPG Kongreß-GmbH, Bad Honnef, Germany), pp. 1393–1396.

Bala, A., Spitzer, M., and Takahashi, T. (**2003**). "Prediction of auditory-spatial acuity from neural images on the owl's auditory space map," Nature (London) **424**, 771–774.

Birchfield, S. T., and Gillmor, D. K. (**2002**). "Fast Bayesian acoustic localization," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bodden, M. (**1993**). "Modeling human sound source localization and the cocktail-party-effect," Acta Acust. **1**, 43–55.

Braasch, J. (**2002**). "Localization in the presence of a distracter and reverberation in the frontal horizontal plane. II. Model algorithms," Acust. Acta Acust. 88, 956–969.

Brandstein, M. S., and Silverman, H. F. (**1997**). "A practical methodology for speech source localization with microphone arrays," Speech Commun. **11**, 91–126.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Cai, H., Carney, L. H., and Colburn, H. S. (**1998**). "A model for binaural response properties of inferior colliculus neurons. I. A model with interaural time difference sensitive excitatory and inhibitory inputs," J. Acoust. Soc. Am. **103**, 475–493.

Calmes, L., Lakemeyer, G., and Wagner, H. (**2003**). "A sound-localization algorithm for a mobile robot," in *Abstractband der 96. Jahresversammlung der Deutschen Zoologischen Gesellschaft* (Humboldt-Universität zu Berlin, Berlin).

Carr, C. E., and Konishi, M. (**1988**). "Axonal delay lines for time measurement in the owls brain stem," Proc. Natl. Acad. Sci. U.S.A. **85**, 8311–8315.

Carr, C. E., and Konishi, M. (**1990**). "A circuit for detection of interaural time differences in the brainstem of the barn owl," J. Neurosci. **10**, 3227–3246.

Chung, W., Carlile, S., and Leong, P. (**2000**). "A performance adequate computational model for auditory localization," J. Acoust. Soc. Am. **107**, 432–445.

Colburn, H. S., and Durlach, N. I. (**1978**). "Models of Binaural Interaction," in *Handbook of Perception*, edited by Carterette, E. C. and Friedman, M. P. (Academic, New York), Vol. **4**, Chap. 11.

Colburn, H. S., Han, Y. A., and Culotta, C. P. (**1990**). "Coincidence model of MSO responses," Hear. Res. **49**, 335–346.

Eyring, C. F. (**1933**). "Methods of calculating the average coefficient of sound absorption," J. Acoust. Soc. Am. **4**, 178–192.

Gaik, W. (**1993**). "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," J. Acoust. Soc. Am. **94**, 98–110

Huang, J., Supaongprapa, T., Terakura, I., Wang, F., Ohnishi, N., and Sugie, N. (**1999**). "A model-based sound localization system and its application to robot navigation," Robotics and Autonomous Systems **27**, 199–209.

Jeffress, L. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.

Jin, C., Schenkel, M., and Carlile, S. (**2000**). "Neural system identification model of human sound localization," J. Acoust. Soc. Am. **108**, 1215–1235.

Joris, P. X., Smith, P. H., and Yin, T. C. T. (**1998**). "Coincidence detection in the auditory system: 50 years after Jeffress," Neuron **21**, 1235–1238.

Knudsen, E. I., Blasdel, G. G., and Konishi, M. (**1979**). "Sound localization by the barn owl (tyto alba) measured with the search coil technique," J. Comp. Physiol. **133**, 1–11.

Lindemann, W. (**1986a**). "Extension of a binaural cross-correlation model by means of contralateral inhibition. I. Simulation of lateralization of stationary signals," J. Acoust. Soc. Am. **80**, 1608–1622.

Lindemann, W. (**1986b**). "Extension of a binaural cross-correlation model by means of contralateral inhibition. II. The law of the first wave front," J. Acoust. Soc. Am. **80**, 1623–1630.

Liu, C., Wheeler, B. C., O'Brien, W. D. Jr., Bilger, R. C., Lansing, C. R., and Feng, A. S. (**2000**). "Localization of multiple sound sources with two microphones," J. Acoust. Soc. Am. **108**, 1888–1905.

Liu, C., Wheeler, B. C., O'Brien, W. D. Jr., Lansing, C. R., Bilger, R. C., Jones, D. L., and Feng, A. S. (**2001**). "A two-microphone dual delay-line approach for extraction of a speech sound in the presence of multiple interferers," J. Acoust. Soc. Am. **110**, 3218–3231.

McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines—Do mammals fit the model?," Trends Neurosci. **26**, 347–350.

Murray, J. C., Wermter, S., and Erwin, H. (**2005**). "Auditory robotic tracking of sound sources using hybrid cross-correlation and recurrent networks," Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE, Piscataway, NJ).

Nakadai, K., Lourens, T., Okuno, H. G., and Kitano, H. (**2000**). "Active audition for humanoid," in Proceedings of the of 17th National Conference on Artificial Intelligence (AAAI-2000), pp. 832–839.

Nakadai, K., Okuno, H. G., and Kitano, H. (**2002**). "Real-time sound source localization and separation for robot audition," in Proceedings of the Seventh International Conference on Spoken Language Processing (ICSLP-2002), Denver, Co, pp. 193–196.

Nishiura, T., Nakamura, M., Lee, A., Saruwatari, H., and Shikano, K. (**2002**). "Talker tracking display on autonomous mobile robot with a moving microphone array," in Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan.

Nix, J., and Hohmann, V. (**2001**). "Enhancing sound sources by use of binaural spatial cues," in Proceedings of the Eurospeech 2001 Workshop on Consistent and Reliable Acoustical Cues (CRAC), Aalborg, Denmark.

Nix, J., and Hohmann, V. (**2006**). "Sound source localization in real sound fields based on empirical statistics of interaural parameters," J. Acoust. Soc. Am. **119**, 463–479.

Omologo, M., and Svaizer, P. (**1994**). "Acoustic event localization using a crosspower-spectrum phase based technique," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Adelaide, Australia.

Parks, T. N., and Rubel, E. W. (**1975**). "Organization of projections from n. magnocellularis to n. laminaris," J. Comp. Neurol. **164**, 435–448.

Peissig, J. (**1993**). *Binaurale Hörgerätestrategien in Komplexen Störschallsituationen (Binaural hearing aid strategies in complex noise environments)*, Fortschr.-Ber. VDI (VDI, Düsseldorf).

Roman, N., and Wang, D. (**2003**). "Binaural tracking of multiple moving

sources," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, Vol. **5**, pp. 149–152.

Spence, C., and Pearson, J. C. (**1990**). "The computation of sound source evaluation in the barn owl," in *Advances in Neural Information Processing Systems 2*, NIPS Conference, Denver, Co, 27–30 November 1989, edited by D. S. Touretzky (Morgan Kaufmann, San Francisco, CA), pp. 10–17.

Stern, R. M., and Trahiotis, C. (**1995**). "Models of binaural interaction," in *Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, New York), Vol. **6**, pp. 347–386.

Sullivan, W. E., and Konishi, M. (**1984**). "Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl," J. Neurosci. **4**, 1787–1799.

Sullivan, W. E., and Konishi, M. (**1986**). "Neural map of interaural phase difference in the owl's brain stem," Proc. Natl. Acad. Sci. U.S.A. **83**, 8400–8404.

Svaizer, P., Omologo, M., and Matassoni, M. (**1997**). "Acoustic source location in a three-dimensional space using crosspower spectrum phase," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Munich, Germany.

Takahashi, T., and Konishi, M. (**1986**). "Selectivity for interaural time difference in the owl's midbrain," J. Neurosci. **6**, 3413–3422.

Valin, J.-M., Michaud, F., Rouat, J., and Létourneau, D. (**2003**). "Robust sound source localization using a microphone array on a mobile robot," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV.

van Veen, B. D., and Buckley, K. M. (**1988**). "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**, 4–24.

Viste, H., and Evangelista, G. (**2004**). "Binaural source localization," in Proceedings of the Seventh International Conference on Digital Audio Effects (DAFx'04), Naples, Italy, pp. 145–150.

Ward, D. B., and Williamson, R. C. (**2002**). "Particle filter beamforming for acoustic source localization in a reverberant environment," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, Vol. **II**, pp. 1777–1780.

Zhou, Y., Carney, L. H., and Colburn, H. S. (**2005**). "A model for interaural time difference sensitivity in the medial superior olive: Interaction of excitatory and inhibitory synaptic inputs, channel dynamics, and cellular morphology," J. Neurosci. **25**, 3046–3058.

Manatee position estimation by passive acoustic localization

Paulin Buaka Muanke and Christopher Niezrecki[a)]
Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, Massachusetts 01854

(Received 10 August 2006; revised 19 December 2006; accepted 5 January 2007)

Passive sound source localization with sensor arrays is based on the estimation of the time difference of arrival (TDOA), and precise TDOA is required to achieve accurate position estimation. For a majority of practical localization systems (based on TDOA estimation with four sensors in two dimensions), only three time delays are computed to determine the location of interest. This paper presents an approach to determine the position of a manatee by using four hydrophones and all the combinations of the TDOAs available. With four hydrophones, six TDOAs are computed and then combined three by three to get 20 possible points for each position to estimate. Experimental results using the Hilbert envelope peak technique to estimate the TDOAs and the least square method to estimate the position are presented. For the tests conducted it is shown that for a manatee call having a high signal-to-noise ratio, the individual position estimated for each of the 20 combinations of TDOAs lies on a straight line, providing a good estimation of the direction of arrival approximately 85% of the time. However, a good estimation of the position is obtained for a manatee near the hydrophone array approximately 55% of the time. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2532210]

PACS number(s): 43.60.Jn, 43.66.Qp [AIT] Pages: 2049–2059

I. INTRODUCTION

A significant amount of research has been conducted on the localization of a sound source during the past several decades. The objective is to locate a sound source with a set of sensors combined as arrays. The majority of practical sound source localization systems are based on the estimation of the time delay or the time difference of arrival (TDOA) of the propagating signal, between sensors (Harris and Ledwidge, 1974). Generally the source localization technique is done in two steps. First, time delay estimates for different groups of sensors are determined. Then these time delays are combined to determine the source location.

Although the theory for the source localization technique using time delays is well documented, its application in the field must be tuned to the particular characteristics of the local environment. In the TDOA strategy, accurate computation of time delay is the basis for accurate source location. Accurate time delay estimates are strongly dependant on the signal-to-noise ratio (SNR) at each sensor of the localization array and good time delay estimates are usually obtained for a high SNR. The TDOA strategy has been successfully applied to radar, sonar systems (Carter, 1981), whale detection and localization (Jarvis and Moretti, 2002) or gunshot localization (Torney and Nemzek, 2005; Deligeorges *et al.*, 2006) where the signal band is narrow and SNR is high. For example, some whale vocalization measurements had a source level of approximately 180 dB (Širovic, 2006) providing a relatively high SNR. For manatees the source levels have been shown to be much lower, approximately 112 dB, thus providing a lower SNR for most measurements (Phillips and Niezrecki, 2004). However, for the manatee localization problem, detection must occur in an environment where the background is noisy and the SNR is poor. The reflection of the sound can also be more complex in a shallow water environment compared to an open sky or a deep ocean.

To find the time delay for each pair of sensors, the data from those sensors are transformed into a function that exhibits a peak in the location corresponding to the source. Many techniques are used to find the time delays. The most commonly used method for time delay estimation is the generalized cross-correlation function (GCC) (Knapp and Carter, 1976; Chan and Ho, 1994; Agius and Saunders, 2000; Moore *et al.*, 2002). The time delay estimate is obtained as the time lag that maximizes the cross-correlation function between filtered versions of the received signal. This method performs well with signals that have a high SNR and has usually the property of breaking down for low SNR (Zeira and Schultheiss, 1991). In order to improve the estimation performance of the correlator, many methods propose to prefilter the received signals prior to performing the cross correlation. With the addition of the pre-filters, the GCC is achieved. Examples of well-known pre-filters include the Wiener processor (Hem and Sehwm, 1985), the Roth processor (Roth, 1971), the smoothed coherent transform (Lapp and Carter, 1976), and the phase transform (Caner, 1972). In general, it is the aim of these pre-filters to attenuate the frequency ranges where the noise is dominant and enhance the frequency ranges where the signal is dominant.

Another method for the estimation of time delays is the adaptive eigenvalue decomposition (Huang *et al.*, 1999). The approach exploits the cross correlation between two microphone outputs to construct an error signal. Then the mean value of the error signal power is minimized by an adaptive

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

filter designed in the frequency domain to search for the channel impulse response. The model assumes that the system is linear time invariant and the acoustic channel is characterized by a finite impulse response filter.

A method based on using the Hilbert transform in the correlation between two signals for the time delay estimation is used in (Thrane, 1984; Thrane *et al.*; Grennberg and Sandell, 1994; Kurz, 2004). In comparison with the generalized cross-correlation method, it was shown that the Hilbert transform method is fast, simple and has the same accuracy as conventional methods for signals that have a high SNR, while it performs better for signals with a low SNR.

Once the time delays between each pair of sensors are computed, one of the methods used to calculate the position of the source is a hyperbolic location method. This approach uses the estimated TDOA from each pair of microphones to create hyperbolic curves. The point where the curves intersect represents the position of the sound source. The exact localization necessitates solving a set of nonlinear hyperbolic equations, which can be computationally demanding. Different strategies can be used to find the intersecting points and some of the principal strategies are described below.

Friedlander's method utilizes a least squares (LS) and a weighted LS error criterion to solve the set of nonlinear hyperbolic equations (Friedlander, 1987). It has been shown (Stark and Woods, 1994) that the LS solution provides the maximum likelihood estimate, if the range difference errors are uncorrelated and Gaussian distributed, with zero mean and equal variances. The method assumes R_1 (the distance between the first sensor and the sound source) is independent of x and y coordinates of the sound source and does reduce the computational complexity as compared to other solutions, but it is suboptimal as compared to some of them, because it assumes that R_1 is constant.

Another method used to obtain an estimate of the sound source position is the Taylor-series method (Foy, 1976; Torrieri, 1984). The Taylor-series method linearizes the set of nonlinear hyperbolic equations by Taylor-series expansion, and then uses an iterative method to solve the system of linear equations. The iterative method begins with an initial guess and improves the estimate at each iteration by determining the local linear least square solution. The Taylor series can provide accurate results and is robust. However, linearization can introduce significant errors when a relatively small ranging error can result in a large position location error. It has been shown by Bancroft (Bancroft, 1985) that eliminating the second order terms in the Taylor-series expansion can lead to significant errors in this situation. The effect of linearization of hyperbolic equations on the position location solution has been studied by Nicholson (Nicholson, 1976).

For arbitrarily placed hydrophones and a system of equations in which the number of equations equals the number of unknown source coordinates, Fang's method (Fang, 1990) provides an exact solution to the set of nonlinear hyperbolic equations. However, Fang's solution does not use the redundant measurements made at additional receivers to improve position location accuracy. In comparison with the Taylor-series method, this method provides a closed form and exact solution and it is also computationally less intensive.

Chan and Ho (1994) proposed a noniterative solution to the hyperbolic position estimation problem that is capable of achieving optimum performance for arbitrarily placed sensors. This method provides an explicit solution that is not available in the Taylor-series method. It is also better than Fang's method as it can take advantage of redundant measurements like the Taylor-series method.

An overview of several other algorithms that have different complexity and accuracy, proposed to solve the set of nonlinear hyperbolic equations for sensors placed in a straight line, can be found in Foy, 1976, and Torrieri, 1984.

In this paper, an acoustic method combining the time delay estimations by using the Hilbert envelope peak approach and the hyperbolic location by a least square approach is used to locate manatees. The acoustic approach discussed uses four hydrophones and all the combinations of the time delays available. When four hydrophones are used, six time delays are computed and then combined three by three to get 20 possible points for each position to be estimated. In a noisy environment, there is uncertainty in the time delay estimation. So, some combinations of time delays generate position locations that are too far away to be realistic, and these points are considered outliers. These outliers are rejected by a statistical outlier test. This paper represents the first effort to locate the position of vocalizing manatee by using acoustic methods. Additionally, the use of numerous time delay combinations to compute multiple possible source locations is seldom performed.

With a growing number of collisions between manatees and boats, the need to develop a system to warn boaters of the presence of manatees, that can signal to boaters that manatees are present in the immediate vicinity, could potentially reduce these boat collisions. So, a manatee localization system can be used in the development of a manatee avoidance system. Likewise, a localization system can be used to study the communication patterns of the animals and to identify the manatees' movement in turbid waters where they are otherwise undetectable.

The paper is organized as follows. Section II presents the sound source position calculation by using the time delay estimation method. The hyperbolic equations and the statistic extreme value test to reject the outliers are presented. Section III presents the experimental setup. In Sec. IV, the experimental results obtained and a discussion of the influence of noise in the position estimation is addressed.

II. POSITION CALCULATION WITH TIME DELAYS OF ARRIVAL

A. Time delay estimation methods

For any two sound sensors, the delay between the arrival times at the sensors determines a hyperbolic curve on which the sound source must lie. With three or more sensors, the intersecting hyperbolas determine the sound source position although sometimes there can be ambiguity. To eliminate the ambiguity, at least four sensors are needed to find the source

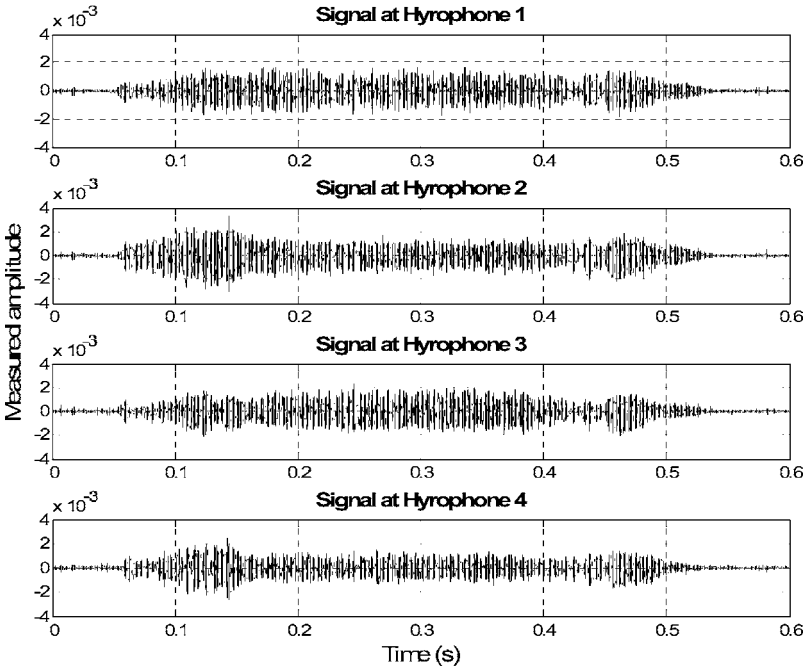


FIG. 1. Sample manatee call measured using four hydrophones.

position in two dimensions (Spiesberger, 2001). The location estimation using TDOA is achieved in two stages (Rappaport *et al.*, 1996). First, the time differences of arrival of sound between hydrohones are estimated and in the second stage, estimated TDOAs are transformed into a set of nonlinear hyperbolic equations which, upon solving, will provide the estimated position of the sound source. Since the equations derived are nonlinear, solution of those hyperbolic equations require use of efficient algorithms. In this section, the techniques used to perform the TDOA estimation and solving hyperbolic nonlinear equations are presented. First, the Hilbert envelope technique to estimate the time delays is introduced and then the least square method for solving the hyperbolic equations to find the position is presented.

The concept of an envelope function is most useful in analyzing rapidly oscillating signals, such as correlation functions resulting from a narrowband signal. The signals should be well behaved in the sense that instantaneous changes in the amplitude of the input signal should be avoided. Cross correlations from narrowband sources, or sources with tonal components, may contain rapidly oscillating components. The slow time variation of the amplitude of these oscillations is referred to as the "envelope" of the function. Using the Hilbert transform, it is possible to remove rapid oscillations and identify the envelope of the function. So, a relatively simple form of signal conditioning (Hilbert transform) is used in this paper to calculate the signal's envelope. The Hilbert envelope of the cross-correlation function is defined as the magnitude of the so-called "analytic signal" of the cross-correlation function. The analytic signal of the cross-correlation function is the complex function that has the cross-correlation coefficient as the real part and the Hilbert transform of the cross-correlation coefficient as the imaginary part. For the time dependent signal $x(t)$ with cross-correlation function $s(t)$, $S(t)$ is the analytic signal of $s(t)$ given by Eq. (1)

$$S(t) = s(t) + j\check{s}(t), \tag{1}$$

where $\check{s}(t)$ is the Hilbert transform of $s(t)$ and is defined by Eq. (2) (Bracewell, 1986)

$$\check{s}(t) = H\{s(t)\} = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{s(\rho)}{\rho - t} d\rho = h(t) * s(t), \tag{2}$$

where the integral is a Cauchy principal value, ρ is the integration variable, "*" denotes convolution, and the Hilbert kernel is denoted by $h(t) = -1/\pi t$.

The Hilbert transform is represented by a convolution integral, i.e., the Hilbert transform is a causal transfer function that behaves like a filter. Thus, the envelope time function $E(t)$ of the signal $s(t)$ is calculated as the magnitude of the analytic function $S(t)$ (Buttkus, 1991)

$$E(t) = |S(t)| = \sqrt{(s(t))^2 + (\check{s}(t))^2}. \tag{3}$$

Squaring and norming of the envelope of the signal leads to a suppression of noise of lower amplitude and to an increase of the signal content of higher amplitude. The high frequency noise is suppressed and the signal is accented. According to Thrane (1984), the correct time delay can be found from the peak of the envelope of the Hilbert cross-correlation function, whether or not the peak of that envelope corresponds to the peak of the cross-correlation function. The time at which this value occurs is then taken as an estimate of the time delay of arrival of the sound at the sensor pair considered. For a manatee call signal measured using four hydrophones (shown in Fig. 1), the Hilbert envelope function to use for estimation of the time delay is shown in Fig. 2. In this case, the envelope function has removed the rapidly varying amplitude fluctuations observed on the cross-correlation function. The peak of the cross-correlation coefficient function and the peak of the Hilbert envelope lie at slightly different points in time, as indicated by the circled

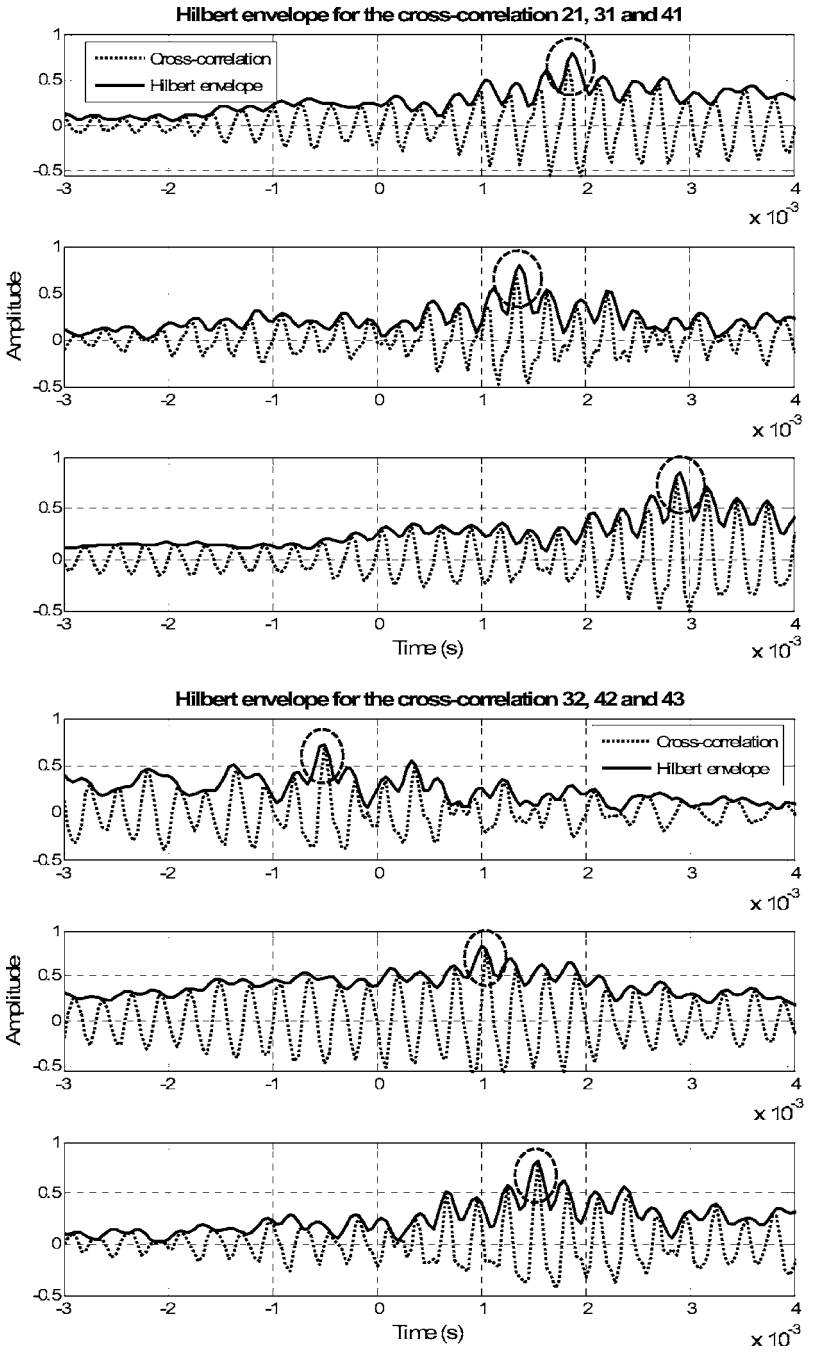


FIG. 2. Cross-correlation function and the corresponding Hilbert envelope for the hydrophone signals.

data. A small deviation in the time delay difference will significantly affect the computation of the sound source position. See Moore *et al.*, 2002 for the influence of error estimation of the time delays in the accuracy of source position estimation.

B. Hyperbolic equations and localization

After the TDOA estimates are obtained, a set of nonlinear hyperbolic equations is defined by transforming TODAs into range difference measurements. Since these equations are nonlinear, solving them is not a trivial procedure (Chan and Ho, 1994). For the two-dimensional (2D) position location of a sound source using four hydrophones, let (X_i, Y_i) with $i=1,2,3,4$, be the known location of the ith hydrophone and (x,y) be the sound source location to be computed. Hydrophone 1 is considered the reference sensor. The range difference between the ith and the jth hydrophone is given as

$$R_{ij} = \sqrt{(X_i - x)^2 + (Y_i - y)^2} - \sqrt{(X_i - x)^2 + (Y_j - y)^2}. \quad (4)$$

The range difference between the ith hydrophone with respect to the reference hydrophone 1 is

$$R_{i1} = cTD_{i1} = R_i - R_1 = \sqrt{(X_i - x)^2 + (Y_i - y)^2} - \sqrt{(X_1 - x)^2 + (Y_1 - y)^2}, \quad (5)$$

where c is the speed of sound in the water (approximately 1500 m/s), TD_{i1} is the TDOA estimate between the hydrophone i and the hydrophone 1, R_1 is the distance between the reference hydrophone 1 and the sound source, and R_{i1} is the range differences between the reference hydrophone

1 and the ith hydrophone. These equations together define the set of nonlinear hyperbolic equations, and solving these equations will yield the estimated position of the sound source in terms of x and y coordinates. A problem to be addressed here is how to find the minimum distance between the different hyperbolas when the intersection of the hyperbolas does not coincide.

There are many algorithms to find the minimum distance between the hyperbolas, for example, the minimum least square error algorithm. The method used the Levenberg-Marquardt nonlinear least squares optimization procedure (Mellinger, 2002a; Mellinger, 2002b) to solve the hyperbolic equations for searching the sound source location. The Levenberg-Marquardt algorithm is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of nonlinear real-valued functions. It has become a standard technique for nonlinear least squares problems (Levenberg, 1944; Marquardt, 1963; Madsen *et al.* 2004), and has been widely adopted in a broad spectrum of disciplines. The Levenberg-Marquardt algorithm can be thought of as a combination of steepest descent and the Gauss-Newton method. When the current solution is far from the correct one, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. When the current solution is close to the correct solution, it becomes a Gauss-Newton method.

The Levenberg-Marquardt algorithm considers an assumed functional relation which maps a parameter vector to an estimated measurement. Here, the range difference between the ith hydrophone (for $i=2,3,4$) with respect to the reference hydrophone 1 are the components of the parameter vector $R=[R_{12}R_{13}R_{14}]$. The measurement vector is computed with the time delay estimates and the speed of sound; $\hat{R}=[c\mathrm{TD}_{21}c\mathrm{TD}_{31}c\mathrm{TD}_{41}]$. So, an arbitrary initial position is chosen, a measured position is determined, and it is desired to find the vector that tries to minimize the square distance $\varepsilon T\varepsilon$ where $\varepsilon=R-\hat{R}$. The goal is to find a location point that minimizes the sum of the square of difference between the actual TDOA and measured TDOA. Figure 3 shows a sample of the three hyperbolic curves with an intersection that gives the sound source position corresponding to the manatee call shown in Fig. 1.

Within this paper four hydrophone receivers are used. As a result, for each individual vocalization six hyperbolas can be intersected to compute a position because there are six time delays between the four receivers. However, only three time delays are sufficient to compute a source location without ambiguity in two dimensions (Spiesberger, 2001). Increasing the number of time delays in a group will reduce the number of combinations that can be used to determine a position estimate. For example, by using a group of six time delays instead of three, six hyperbolas are used to compute a single position. Regarding the number of time delays used to compute a position, it is not clear which approach is more accurate. However, by using more combinations (i.e., 20), a better estimate of the direction of arrival is obtained.

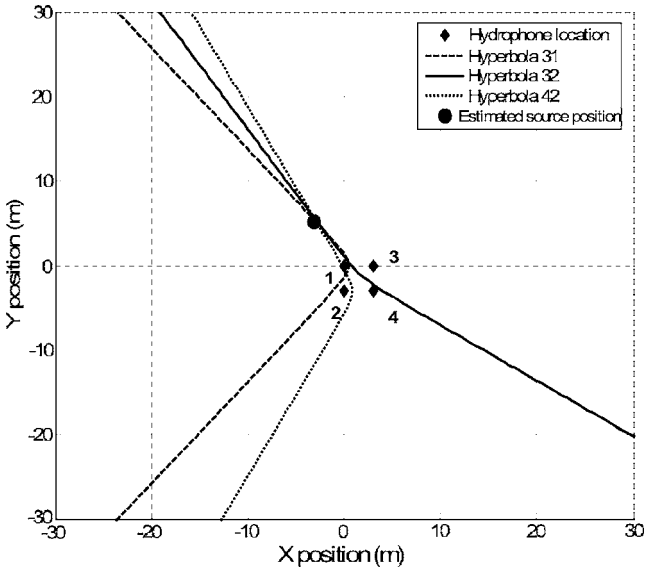


FIG. 3. Illustration of three hyperbolas intersecting to obtain a sound source location. Hyperbola 31 is created between hydrophones 3 and 1, hyperbola 32 is created between hydrophones 3 and 2, and hyperbola 42 is created between hydrophones 4 and 2.

C. Extreme value test

Due to the presence of noise in the environment, the time delays are estimated with some uncertainty. Some time delay combinations provide position estimation far from the remaining group and can be considered outliers to be eliminated in order to get a good estimation of a manatee position. The rejections can be done using the extreme value test presented in this section.

The extreme value test, also known as "*Dixon's test*" provides an appropriate method for examining whether an observation is an outlier (Gibbons, 1994). If more than one outlier is considered to be possible (from visual inspection of the data using frequency and/or normal probability plots) then the presumed outliers must be tested in sequence from the most severe error to the least error. This will avoid the possibility of one data value masking another. Before the test can proceed, a test of normality on the data not suspected of being outliers must be performed. Normality of the main data is required for the application of the method. Dixon's test is generally used for detecting a small number of outliers (Gibbons, 1994). The data are ranked in ascending order, and then based on the sample size; the τ statistic (as defined in Gibbons, 1994) for the highest value or lowest value is computed. The computed τ statistic is then compared to a critical value at a chosen level of confidence. If the τ statistic is less than the critical value, the current point is not rejected, and the conclusion is that no outliers are present. If the τ statistic is greater than the critical value, the current point is rejected, and the conclusion is that the most extreme value is an outlier. To check for other outliers, the Dixon test can be repeated; however, the power of this test decreases as the number of repetitions increases.

For the 20 different combinations of localizations (using four sensors), on occasion, several of the calculations (described in Sec. IV) generate results that are not physically

possible. These errors result from noise, computational error, reflections, time delay estimation error (choosing the wrong cross-correlation peak), etc. These positions are located on land and are not possibly correct. The statistical extreme value test used helps to eliminate these spurious results. Theoretically all 20 calculated locations should be identical; however they are not.

D. Position location algorithm steps

In order to find the position location, the implementation of the algorithm that combines the time delay estimations and solving the hyperbolic equations by using all the time delays to be computed has the following stages:

1. Compute the six time delays TD_{21}, TD_{31}, TD_{41}, TD_{32}, TD_{42} and TD_{43}.
2. Use each combination of time delays and the Levenberg-Marquardt algorithm in order to solve a set of nonlinear hyperbolic equations to find the position. The 20 combinations to be used are $\{TD_{21}, TD_{31}, TD_{41}\}$; $\{TD_{21}, TD_{31}, TD_{32}\}$;…$\{TD_{32}, TD_{42}, TD_{43}\}$.
3. With the background noise that leads to some uncertainty in the time delay estimations, some combinations of time delays may generate an outlier. The extreme value statistic test is applied to reject outliers and the average of the remaining points provides an estimation of a manatee's position.

III. EXPERIMENTAL SETUP

The algorithm developed was tested with manatee vocalizations recorded in Homosassa Springs Wildlife State Park, Florida, on January 18th and March 28th, 2006 and in Crystal River, Florida, on January 19th, 2006. For each test, four hydrophones were placed in the water. For the tests conducted on March 28th and January 19th, a square hydrophone array configuration was used. For the test on January 18th the array was placed in an approximate diamond configuration. In the square array configuration, the hydrophones were placed 10 ft (3.05 m) apart while in the diamond configuration, a point at the edge of the spring was used as a reference point and the distances to each hydrophone were recorded. The distance between the hydrophones was also recorded, so that the relative positions could then be calculated. Manatee vocalizations were recorded using a Teac RD-135T DAT recorder with a sample rate of 48 kHz. At the same time, a video camera was used to visualize the manatees when it was possible to see them. The video camera time was synchronized with the Teac recorder time so the position computed by the algorithm could be compared with that seen in the video. The video tape is made from a platform located in the springs for the Homosassa Springs tests, while all the data acquisition equipment and the video camera are mounted on a boat for the Crystal River test.

For the data processing, the recordings were replayed in the laboratory and the time windows with the audible manatee vocalizations were identified and then processed offline to estimate the manatee position.

IV. EXPERIMENTAL RESULTS

The manatee vocalizations are triangulated using the method previously described, and the estimated positions obtained are compared with the position from ISHMAEL software (Mellinger, 2002a). The original data recordings are first filtered by an eighth order Butterworth band pass filter having a pass band from 0.6 to 11.4 kHz. The SNR computed for all the measurements collected shows that the maximum SNR obtained does not exceed 25 dB.

It should be noted that the Ishmael estimated position is not the correct position of the manatees but is used as a benchmark for comparison. The true manatee positions are determined through visual observation. Although the author's algorithm and the algorithm used by ISHMAEL both use a least square method to solve the set of nonlinear hyperbolic equations, the primary differences between the two algorithms include: (1) The Hilbert envelope peak method is used by the author to compute the time delays, while ISHMAEL uses the peak cross-correlation method. (2) All 20 combinations of the time delays available are used in the author's algorithm. This algorithm provides both an estimated manatee position and the direction of arrival of the manatee call, while ISHMAEL only provides an estimated position by using one time delay combination. (3) The author's algorithm utilizes an outlier statistical test to reject extraneous computational results. This provides a measure of the quality of the estimated position. The output generated by ISHMAEL produces only a single data point in which the quality of the result cannot be interpreted.

A. Crystal River—January 19th, 2006

Figure 4(a) shows the result of a manatee located approximately 18.0 m away from the array while Fig. 4(b) corresponds to a case near the array during the Crystal River test. In these two cases, five of the 20 points corresponding to 20 combinations of time delays are rejected as outliers after the statistic test. The 15 remaining points lie on the straight line that gives a good estimated value of the direction of arrival while the average of the 15 remaining points gives manatee location which differs with the position from ISHMAEL to less than 5.0 m. Figure 5 illustrates a position deviation (distance between the individual estimated position and the position obtained from ISHMAEL software) for each time delay combination in comparison with the ISHMAEL software results for a set of ten manatee calls. An individual position estimate refers to a position obtained for a specific time delay combination. For example, an individual position for one combination is the position estimated by using the three hyperbolas from three time delays, such as: TD_{21}, TD_{31}, and TD_{41}. One can observe that most of the outliers occur at the time delay combinations (21, 41, 42) and (31, 41, 43), which corresponds to the cases where only three instead of four hydrophones are used in the position estimation. When the outliers are rejected, for the remaining points, the position deviation is less than 3.0 m as shown in Table I. The table also shows that better signal-to-noise ratio generally reduces the average deviation. The SNR is computed by taking the root mean square (rms) value of the time domain

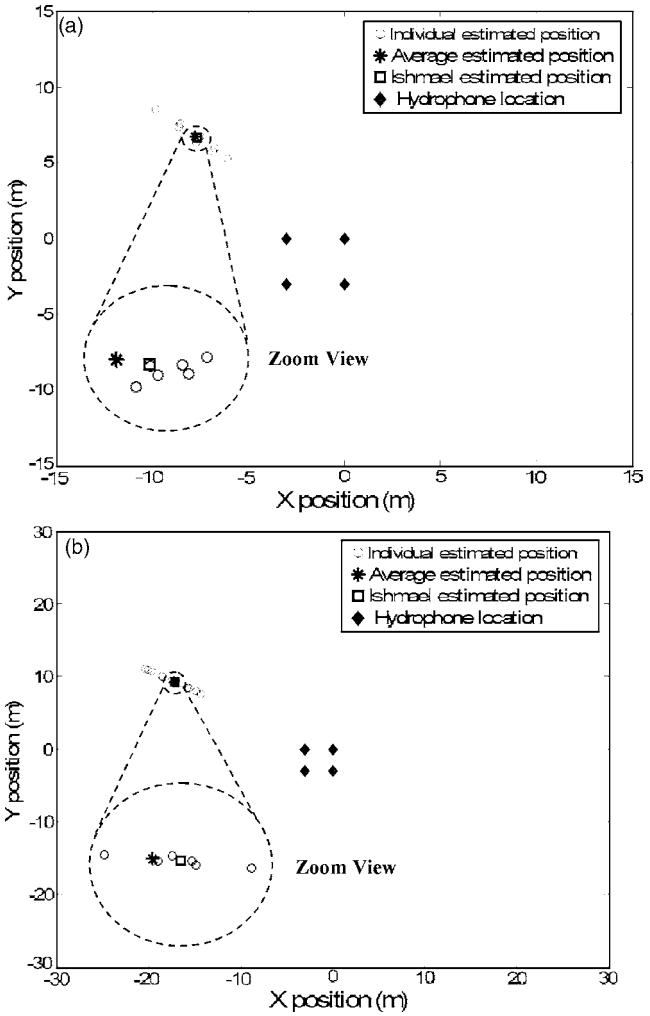


FIG. 4. Manatee's call located near the array (a) and at approximately 18.0 m away from the hydrophone array (b) during the Crystal River test.

signal in the region where the pure manatee call is present and dividing that value by the rms value over the same time interval just prior to the call where only the background noise is present. It is assumed that the background noise levels do not vary significantly over the duration of a manatee call (Yan *et al.* 2006).

TABLE I. Average deviation from ISHMAEL for various signal-to-noise ratios for the Crystal River test.

Call No.	Average deviation (m)	Hydrophone SNR (dB)			
		1	2	3	4
1	1.87	6.14	6.98	3.20	4.85
2	2.18	9.31	7.90	5.64	6.83
3	2.02	9.46	8.87	6.89	6.93
4	1.72	9.30	8.98	9.40	8.20
5	1.70	10.21	10.32	11.85	9.70
6	1.21	15.55	12.10	12.34	7.66
7	1.38	17.32	15.78	14.80	13.93
8	1.28	19.50	15.40	10.92	12.73
9	1.60	17.32	15.78	14.80	13.93
10	1.14	21.32	21.61	19.43	16.32

For positions far away from the hydrophone array, the individual position estimated from each combination of the time delays shows a good estimation of the direction of arrival (DOA) of the manatee's call, when the manatee is over 10.0 m from the hydrophone array. It is evident that these points align in a straight line, indicating the direction of arrival (see Fig. 4). For the positions nearby the hydrophone array, the individual estimated positions for each combination of time delays also remain on a straight line. During the Crystal River test, only one camera has been used to videotape the manatees, so it is difficult to match the position estimated with the video image. Because the area at Crystal River is not confined and has numerous manatees, it was not possible to identify visually the location of all the manatees in the immediate vicinity.

B. Homosassa Springs—January 18th, 2006

Figures 6(a) and 6(b) show the result of a sample manatee location call for two different vocalizations at Homosassa

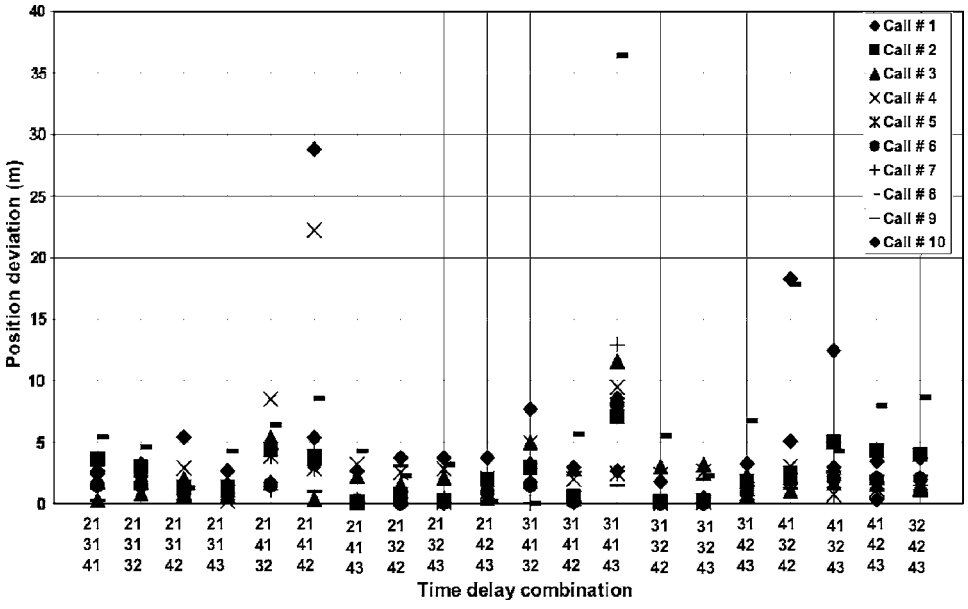


FIG. 5. Deviation between the positions estimated by the authors' software and from the ISHMAEL software for a set of ten manatee calls during the Crystal River test.

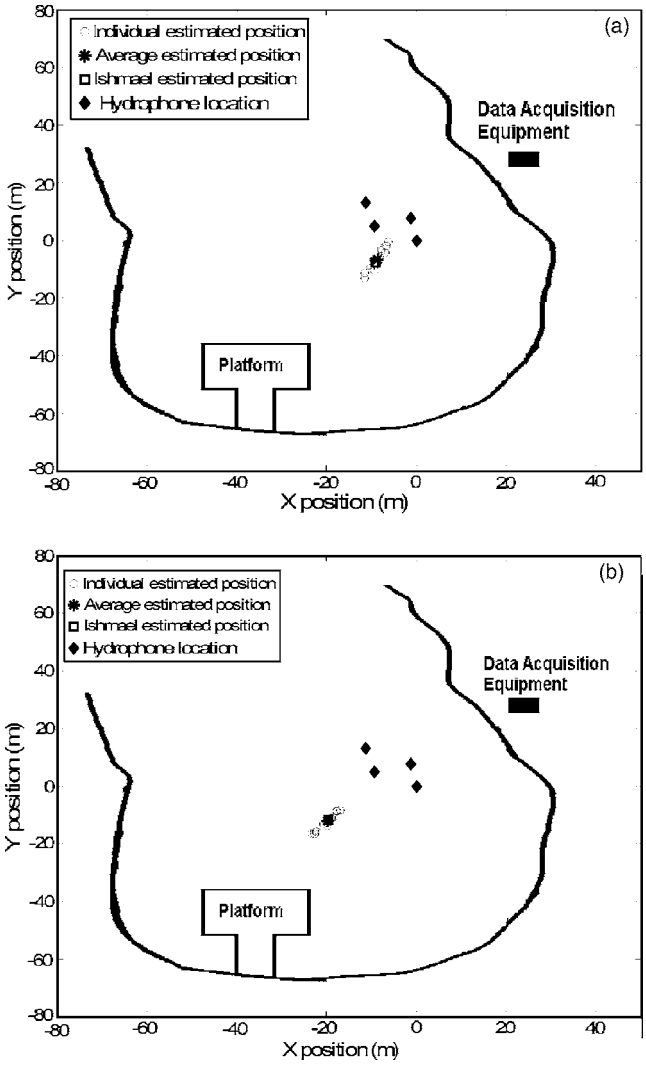


FIG. 6. Results of a sample manatee call location for two different vocalizations at Homosassa Springs January 18th, 2006 test.

Springs—January 18th, 2006 test. During this test, only one manatee was present in the spring run near the array during the recording.

The average deviation between the position obtained from the ISHMAEL software and each of the individual position estimates is shown in Table II with the corresponding signal-to-noise ratio for each hydrophone. The empirical observation shows that the diamond configuration array used during this test is less sensitive to the signal-to-noise ratio.

TABLE II. Average deviation from ISHMAEL for various signal-to-noise ratios for the Homosassa Springs January 18th, 2006 test.

Call No.	Average deviation (m)	Hydrophone SNR (dB) 1	2	3	4
1	4.08	7.15	3.10	6.37	8.04
2	4.00	5.73	3.50	4.80	7.15
3	1.56	8.40	7.08	13.83	9.93
4	1.30	4.86	3.85	3.52	6.70
5	4.40	4.57	3.48	3.92	1.76
6	3.90	7.35	0.96	8.25	6.25
7	2.02	4.96	3.08	3.58	3.03
8	1.20	4.63	3.13	3.60	9.58
9	3.25	4.17	3.97	3.60	10.25

The plot of the deviation between each individual position estimated for each combination of time delays is shown in Fig. 7. For most of the combinations the deviation remains under 5.0 m. Once again, a larger deviation is observed for the combinations (21, 41, 42), (31, 41, 43) and (32, 42, 43) that corresponds to the cases where only three instead of four hydrophones are used in the position estimation. However, due to the low SNR, there are some combinations in which the deviation is more than 5.0 m, even when all four hydrophones are used.

C. Homosassa Springs—March 28th, 2006

For the test at Homosassa Springs on March 28th, 2006, approximately 2 h of manatee vocalizations were recorded. The recordings started at 10:46 a.m. and continued until 12:44 p.m. There were four manatees in the immediate vicinity. To distinguish them, the manatees are labeled manatee A, B, C, and D. Two video cameras were used to track the manatees. The manatee vocalizations heard on the recording were located primarily at the beginning of the recording when a diver was present in the spring and at the time when a manatee show took place, in which the manatees were receiving carrots. Outside of these time windows, most of the time the manatees remained stationary and appeared to sleep. A total of 30 manatee calls were recorded during this period.

A sample of the localization results is shown in Fig. 8(a) with the call coming from manatee B. In Fig. 8(b), three manatees (B, C, D) are grouped together and the call is coming from manatee A, which is moving to join the others manatees.

The manatee vocalizations are triangulated using the method previously described, and the estimated positions from the methods are compared with the position obtained with the ISHMAEL software. Six of the 30 calls have a SNR less than 3 dB and the positions obtained for these six cases are located outside of the waterway. This error is attributed to the noise that affects the time delays estimates. For the 15 of the 24 remaining calls, the position estimated by the method developed match the position seen in the video. The best results are obtained when the manatee is less than 15.0 m away from the hydrophone array and the manatee calls have approximately a SNR at least 8 dB for each hydrophone channel. The difference observed with the position estimated with ISHMAEL software is less than 3.0 m as shown in Table III.

For the locations near the array (distance $<$ 15.0 m) estimated position and direction of arrival obtained are accurate. For distance larger than 15.0 m, the experimental results also show a good estimation of the DOA. It is possible to achieve improved performance for distances far from the hydrophones if two hydrophone arrays are used for which the intersection of the DOAs would give a good estimate of the position.

Table III presents the results for the comparison with the ISHMAEL software for 14 of the 30 calls that had good posi-

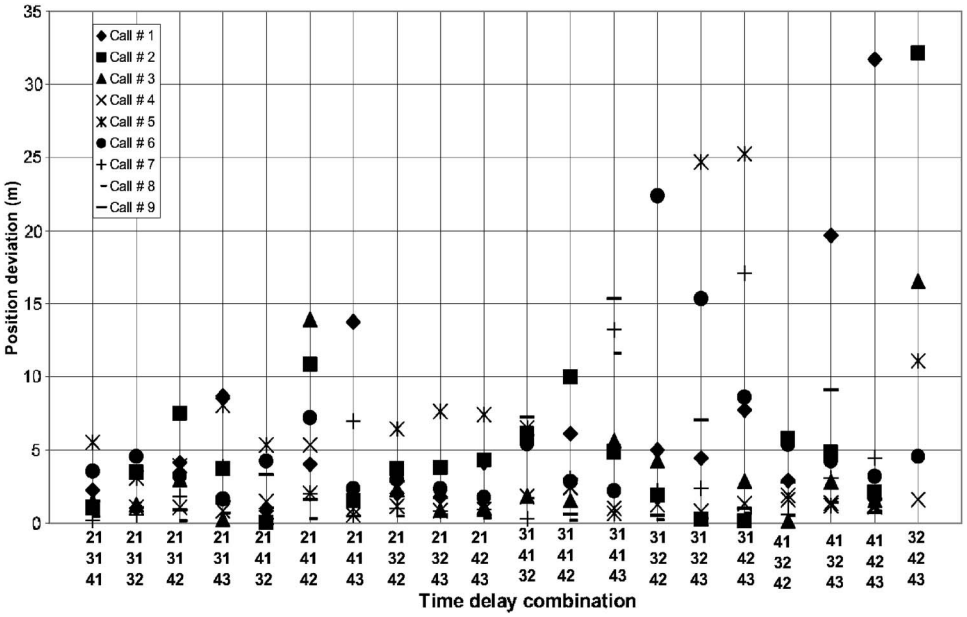


FIG. 7. Deviation between the positions estimated by the code developed and from the ISHMAEL software for a set of nine calls during Homosassa Springs January 18th, 2006 test.

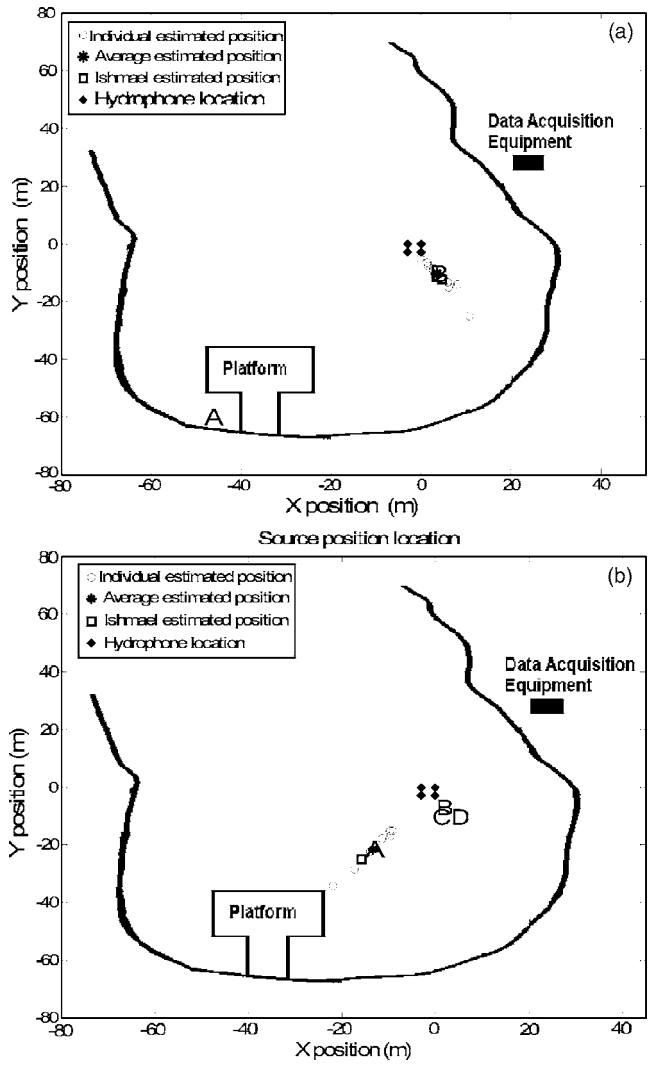


FIG. 8. Results of a sample manatee call location for two different vocalizations at Homosassa Springs March 28th, 2006 test.

tions and DOA estimations. The remainder had a poor SNR (six calls) or only a good DOA estimation (ten calls). The deviation presented in Table III corresponds to the Euclidian distance between the positions estimated with the approach developed in this paper and that obtained from the ISHMAEL software. It can be observed from Table III that for some calls the ISHMAEL software diverges while the author's approach does not. Likewise, unlike ISHMAEL an accurate DOA is still generated by the method presented for most of the cases.

V. DISCUSSION

The SNR is one of the most important sources of TDOA estimation error in the convolution method and the time difference error decays exponentially as the SNR of the two signals increases. When examining the impact of noise on the experimental accuracy, the three intersection points close to the manatee position do not coincide exactly in the presence of noise. In general, the results from Homosassa springs and those from Crystal River have shown good position and/or direction of arrival estimation of the manatee's call when the data have a high signal-to-noise ratio (more than 8 dB on each hydrophone). The location estimation for a call far away from the hydrophone array is more sensitive to the presence of noise. The other effect of noise can also be seen in the number of the outliers in the individual estimated positions for each time delay combination. A low signal-to-noise ratio generates many outliers for the estimation of the position. When all the measurement data contained high levels of noise, the localization estimates were not consistent and a localization estimate was not obtainable.

It should be noted that additional calculations were performed by the authors on the measurement signals that generated a good position estimation. The analysis was done on the second and third highest peak in the cross-correlation envelope using all 20 combinations for a single vocalization signal. The results indicated that choosing the second or third highest peak for the time of flight difference between sensors

TABLE III. Manatee calls position for the Homosassa Springs March 28th, 2006 test.

	Position location (m)					Hydrophone SNR (dB)			
	Authors Software		Ishmael		Deviation				
Call No.	X	Y	X	Y	(m)	1	2	3	4
				Distance from the array <15.00 m					
1	3.30	−9.73	4.70	−12.30	2.95	8.30	14.23	4.37	5.91
2	1.55	−5.69	1.56	−5.50	0.20	5.31	4.33	6.51	3.73
3	1.73	−7.50	2.40	−8.50	1.20	9.27	5.91	4.05	1.23
4	−0.57	6.05	0.30	6.84	1.17	2.01	1.18	1.78	1.33
5	0.83	8.17	0.45	7.05	1.18	5.82	1.27	3.79	0.41
6	1.80	3.10	2.54	2.17	1.20	7.71	2.45	6.04	1.37
7	0.78	−10.40	1.00	−11.60	1.50	23.07	24.00	22.70	17.56
				Distance from the array >15.00 m					
8	−40.04	−22.37		*Diverge*	⋯	10.16	10.30	12.20	13.56
9	−36.60	−40.84	−40.60	−45.85	6.41	11.17	10.57	10.28	12.87
10	−35.76	−35.82		*Diverge*	⋯	8.27	4.55	5.73	9.42
11	−21.92	−30.26	−22.30	−30.42	0.55	14.27	15.66	15.67	14.83
12	−41.10	−22.71		*Diverge*	⋯	9.15	15.81	8.40	10.90
13	−45.48	−58.60	−45.71	−58.81	0.31	12.40	15.00	10.27	14.10
14	−17.72	−28.35	−15.84	−25.00	3.84	14.75	15.50	13.76	14.78

did not significantly change the location estimation because these peaks are so closely spaced in time. This implies that the propagation of sound between a manatee and each hydrophone in the environment where measurements have been conducted is dominated by the straight path and not by echoes due to reflection.

VI. CONCLUSIONS

This paper focuses on identifying the location of manatees by using acoustic methods. Tests are conducted with a hydrophone array placed in two different configurations at Crystal River and at Homosassa Springs, Florida. Manatee vocalizations were recorded simultaneously as video recorded the manatee positions. The video is used to correlate the manatee vocalizations with individual manatee positions. The measured signals were processed offline with the code developed that uses the Hilbert envelope peak method to estimate the time delays and the least square method to estimate the manatee position by solving the set of nonlinear hyperbolic equations. Then the position obtained is compared with the results predicted from the ISHMAEL software. For positions not near the array, the individual position estimated from each combination of the time delays shows a good estimation of the DOA of the manatee's call. This can be observed when the call is more than 15.0 m from the hydrophone array. It is evident that these points fall in a straight line, indicating the direction of arrival. For positions near the array (distance <15.0 m from the array), when the signals have the SNR higher than 8 dB, a good estimation of the position is determined.

ACKNOWLEDGMENTS

The authors would like to express their sincere appreciation to the Florida Fish and Wildlife Conservation Commission, Florida Sea Grant, and the University of Florida College of Veterinary Medicine, Marine Mammal Program, in supporting this research. The authors would also like to thank Socrates Deligeorges, at Boston University, for providing insight into the time delay of arrival estimation.

Agius, A. A., and Saunders, S. R. (2000). "Design and development of a methodology for efficiently tracing the source of intermittent EMC disturbances to radio reception, final report," Radio Communications Agency project reference AY 3639, Centre for Communications Systems Research, University of Surrey, United Kingdom.

Bancroft, S. (**1985**). "An algebric solution of the GPS equations," IEEE Trans. Aerosp. Electron. Syst. **AES-21**, 56–59.

Bracewell, R. N. (**1986**). *The Fourier Transform and its Applications*, 2nd ed. (McGraw–Hill, New York).

Buttkus, B. (**1991**). *Spektralanalyse und Filtertheorie in der angewandten Geophysik* (Spectrographic Analysis and Filter Theory in Applied Geophysics), (Springer, Verlag, Berlin).

Caner, G. C. (1972). "The smoothed coherent transform SCOT," Technical Report No. TM TC 159-72, Naval Undersea Warfare Center, Newport, RI.

Carter, G. C. (**1981**). "Time delay estimation for passive sonar signal processing," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-29**, 463–470.

Chan, Y. T., and Ho, K. C. (**1994**). "A simple and efficient estimator for hyperbolic location," IEEE Trans. Signal Process. **42**, 1905–1915.

Deligeorges, S., Zosuls, A., Mountain, D., and Hubbard, A. (**2006**). "A biomimetic robotic system for localizing gunfire," J. Acoust. Soc. Am. **119**, 3271.

Fang, B. T. (**1990**). "Simple solutions for hyperbolic and related fixes," IEEE Trans. Aerosp. Electron. Syst. **26**, 748–753.

Friedlander, B. (**1987**). "A passive localization algorithm and its accuracy analysis," IEEE J. Ocean. Eng. **OE-12**, 234–244.

Foy, W. H. (**1976**). "Position-location solutions by Taylor-series estimation," IEEE Trans. Aerosp. Electron. Syst. **AES-12**, 187–194.

Gibbons, R. D. (**1994**). *Statistical Methods for Groundwater Monitoring* (Wiley, New York).

Grennberg, A., and Sandell, M. (**1994**). "Estimation of subsample time delay differences in narrowband ultrasonic echoes using the Hilbert transform correlation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 588–595.

Harris, R. W., and Ledwidge, T. J. (**1974**). *Introduction to Noise Analysis* (Pion Limited, London).

Hem, A., and Sehwm, S. (**1985**). "A new generalized cross correlator," IEEE Trans. Acoust., Speech, Signal Process. **33**, 38–54.

Huang, Y., Benesty, J., and Elko, G. W. (**1999**). "Adaptive eigenvalue de-

composition algorithm for real time acoustic source localization system," *Acoustics, Speech, and Signal Processing, ICASSP'99. Proceedings*, Phoenix, AZ, p. 38-54.

Jarvis, S., and Moretti, D. (**2002**). "Passive detection and localization of transient signals from marine mammals using widely spaced bottom mounted hydrophones in open ocean environments," *Conference Proceedings, International Workshop on the Applications of Passive Acoustics in Fisheries*, Cambridge, MA.

Knapp, C. H., and Carter, G. C. (**1976**). "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**, 320–327.

Kurz, J. H. (**2004**). "Signal conditioning of acoustic emissions and ultrasound signals mind the traps," Otto-Graf J. **15**, 59–76.

Lapp, C. H., and Carter, G. C. (**1976**). "The generalized correlation method for estimation of time delays," IEEE Trans. Acoust., Speech, Signal Process. **24**, 320–327.

Levenberg, K. (**1944**). "A method for the solution of certain nonlinear problems in least squares," Q. Appl. Math. **2**, 164–168.

Madsen, K., Nielsen, H. B., and Tingleff, O. (2004). "Methods for nonlinear least squares problems," Technical University of Denmark. Lecture notes, available at http://www.imm.dtu.dk/courses/02611/nllsq.pdf.

Marquardt, D. W. (**1963**). "An algorithm for the least-squares estimation of nonlinear parameters," SIAM J. Appl. Math. **11**, 431–441.

Mellinger, D. (**2002a**). "ISHMAEL 1.0 user's guide. ISHMAEL: Integrated system for holistic multi channel acoustic exploration and localization," Tech. Report OAR PMEL-120, NOAA Technical Memorandum.

Mellinger, D. (**2002b**). http://www.pmel.noaa.gov./pubs/PDF/mell2434/mell2434.pdf Accessed on 3/14/07.

Moore, P. J., Glover, I. A., and Peck, C. H. (**2002**). "An impulsive noise source position locator, Final report," Radio communications Agency project reference AY3925, Department of Electronic and Electrical Engineering, University of Bath.

Nicholson, D. L. (**1976**). "Multipath and ducting tolerant location techniques for automatic vehicle location systems," in *IEEE Vehicular Technology Conference*, Washington, D.C., March 24–26, pp. 151–154.

Phillips, R., and Niezrecki, C. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428.

Roth, P. R. (**1971**). "Effective measurements using digital signal analysis," IEEE Spectrum **8** 62–70.

Rappaport, T. S., Reed, J. H., and Woerner, B. D. (**1996**). "Position location using wireless communications on highways of the future," IEEE Commun. Mag. **34**, 33–41.

Širovic, A. (2006). "Blue and fin whale acoustics and ecology off Antarctic Peninsula," Ph.D. thesis, University of California, San Diego.

Stark, H., and Woods, J. W. (**1984**). *Probability, Random Processes and Estimation Theory for Engineers*, (Prentice-Hall, Inc., 2nd edition).

Spiesberger, J. L. (**2001**). "Hyperbolic location errors due to insufficient numbers of receivers," J. Acoust. Soc. Am. **109**, 3076.

Thrane, N. (1984). "Hilbert Transform," Bruel & Kjaer Technical Review, **3**, pp. 3–15.

Thrane, N., Wismer, J., Konstantin-Hansen, H., and Gade, S., "Practical use of the Hilbert transform," Bruel & Kjaer Application Note, BO 0437-11.

Torney, D. C. and J. Nemzek, R. J. (**2005**). "Least-error localization of discrete acoustic sources," Appl. Acoust., **66**, 1262–1277.

Torrieri, D. J. (**1984**). "Statistical Theory of Passive Location Systems," IEEE Trans. Aerosp. Electron. Syst., **AES-20**, 183–198.

Yan, z., Niezrecki, C., Cattafesta, L., and Beusse, D. O. (2006). "Background Noise Cancellation of Manatee Vocalizations Using an Adaptive Line Enhancer," J. Acoust. Soc. Am., in press 2006.

Zeira, A. and Schultheiss, P. M. (**1991**). "Thresholds and related problems in time delay estimation," In Proceedings of international conference on acoustics, speech and signal processing Toronto, Canada, 1261–1264.

Multiple angle acoustic classification of zooplankton

Paul L. D. Roberts[a] and Jules S. Jaffe
Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238

(Received 8 August 2006; revised 23 January 2007; accepted 24 January 2007)

The use of multiple angle acoustic scatter to discriminate between two taxa of fluid-like zooplankton, copepods and euphausiids, is explored. Using computer modeling, feature extraction, and subsequent classification, the accuracy in discriminating between the two taxa is characterized via computer simulations. The model applies the distorted wave Born approximation together with a simple system geometry, a linear array, to predict a set of noisy training and test data. Three feature spaces are designed, exploiting the relationship between the shape of the scatterer and angularly varying scattering amplitude, to extract discriminant features from these data. Under the assumption of uniform random length and uniform three-dimensional orientation distributions for each class of scatterers, the performance of several classification algorithms is evaluated. Simulations reveal that the incorporation of multiple angle data leads to a marked improvement in classification performance over single angle methods. The improvement is more substantial using broadband scatter. The simulations indicate that under the stated assumptions, a low classification error can be obtained. The use of multiple angle scatter therefore holds promise to substantially improve the *in situ* acoustic classification of fluid-like zooplankton using simple observation geometries. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2697471]

PACS number(s): 43.60.Np, 43.30.Sf, 43.60.Fg [EJS] Pages: 2060–2070

I. INTRODUCTION

Zooplankton play a major role in the global ecosystem and the employment of remote sensing techniques for measuring abundance and behavior continues to be a venerable goal. Compared with optically based methods, acoustic ones have an inherent advantage in that sound is attenuated less than light. This leads to both larger detection distances as well as sampling volumes. Unfortunately, problems associated with a lack of specificity have hindered the use of acoustic techniques on a routine basis. Work by McNaught[1] and the subsequent development of multiple frequency methods by Holliday and colleagues have revealed both the great advantages and also challenges that exist when using this technique.[2–6] Although additional work with broadband sound to discriminate between three groups of zooplankton[7,8] indicated that it was possible to correctly classify specific examples from each group with reasonable success (80% overall average correct classification), the goal of robustly mapping acoustic volume scattering to biophysical parameters of zooplankton under various oceanographic conditions has remained elusive. One problem has been the confounding influence of both orientation and material properties on backscatter magnitude. This often prevents investigators from making the necessary link between animal size and backscatter magnitude. Additional complications have been that there is substantial scatter from nonbiological sources such as suspended sand, bubbles, and perhaps even microstructure.[9] If a way could be found to discriminate among various taxa acoustically, in spite of these problems, it would be of great value.

In this article the potential increase in classification accuracy that results from observation of reflected sound from multiple angles is considered. It has recently been proposed that sound scattered at multiple angles can be used in order to both size and measure the orientation of fish bladders.[10] The underlying concept is that the spatial structure of the sound field from a single, strong scattering target, has a characteristic pattern related to its size. The success of the method was illustrated with a well-known data set[11] and various sampling theorems were proved to obtain unaliased sampling of the scattered sound field. Here, the use of sound scattered at multiple angles in order to discriminate among two zooplankton taxa is explored via forward modeling and subsequent classification.

Many have considered the formulation of acoustic models to predict backscatter as an important component of a program to characterize animals *in situ*. A family of scattering models can be successfully used to predict the acoustic reflectivity of several different types of zooplankton.[12–15] The situation with respect to crustacean zooplankton is especially good as use of the distorted wave Born approximation (DWBA) has been validated.[16,17] A website maintained by Benfield[18] provides public access to several zooplankton models and their morphologies.

Outside of the realm of ocean ecology, recent work in the acoustic classification of stationary targets from multiple views has demonstrated that the multiple views can significantly improve target classification when combined with suitable feature extraction and classification algorithms.[19–22] Applications such as underwater mine detection,[20,22] airborne

[a]Also at: Electrical and Computer Engineering Department, U.C.S.D., La Jolla, California 92093; electronic mail: paulr@mpl.ucsd.edu

 0001-4966/2007/121(4)/2060/11/$23.00

target identification,[23,24] and unexploded ordinance detection[25] have been considered. Most algorithms apply a hidden Markov model (HMM) to account for either the unknown sensor-target aspect,[20,21,26] or the unknown target type.[22] One approach decomposes the target reflections into a set of discrete angular regions yielding a set of possible states in the HMM.[26] Alternatively, a nonlinear backpropagated neural network has been used to fuse the classification results for multiple views in a wavelet packet based feature space.[19] This formulation demonstrated very good performance in discriminating between mine and nonmine like targets from multiple aspect scattering measurements.

Adaptation of multiple angle scatter techniques to zooplankton classification has promise to confer benefits when used in conjunction with the more traditional backscatter techniques. However, animals are dynamic and therefore require an observation system in which multiple views are obtained almost simultaneously. One solution is to use simultaneous multiple angle scatter measurements. A second issue is related to the feature space used to represent the data. Previous work in target classification considered rigid objects and therefore applied wavelet packets[19,22,27] or matching pursuit with an elastic scattering based dictionary.[20,26] However, the resulting feature spaces are not appropriate for the fluid-like weak scatterers considered here. A more appropriate idea for this problem is to exploit the relationship between the shape of the scatterer and the angularly varying scatter amplitude.

This paper explores, through simulation, the use of a one-dimensional array to collect multiple angle scatter and subsequently use these data to discriminate among zooplankton taxa. The case treated is that of differentiating between two taxa of crustacean zooplankton: copepods and euphausiids. The motivation for treating these animals stems from their significance in zooplankton populations of the California Current. As shown here, the large morphological difference between the two groups[28] will allow this discrimination.

Section II summarizes the theoretical basis[14,17] for forward model computations which are used to generate the synthetic data. Section III defines the feature spaces that are applied to reduce the raw data to a small set of discriminant parameters. Section IV describes the nearest neighbor and multilayer perceptron classification algorithms that have been used for classification. Section V discusses the performance of the classifiers as a function of the various parameters that are available. Section VI summarizes the results and identifies future research areas.

II. FORWARD MODELING: THEORY AND NUMERICAL IMPLEMENTATION

In this section, an acoustic forward model for generating synthetic data is proposed using linear system theory and the DWBA. The forward model permits the prediction of the received signal for a known transmit signal using the impulse response of the scatterer. This depends on both the physical properties of the scatterer such as size, shape, and material and also the orientation of the scatterer and the geometry of transmitters and receivers. Under the assumption of linearity, and neglecting effects of spreading and medium attenuation, the received signal $p(t)$ is given by the convolution of the transmitted signal $s_0(t)$ with the impulse response of the scatter $s(t,\mathbf{k}_i,\mathbf{k}_s,\theta,\phi,\mathbf{\Gamma})$,

$$p(t) = \int_{-\infty}^{\infty} s_0(\tau) * s(t-\tau,\mathbf{k}_i,\mathbf{k}_s,\theta,\phi,\mathbf{\Gamma})d\tau, \tag{1}$$

where $\mathbf{k}_i$ and $\mathbf{k}_s$ are the incident and scattered wave vectors, θ and ϕ define the orientation of the scatterer, and $\mathbf{\Gamma}$ is a parameter matrix describing the size, shape, and material properties. Assuming values for these parameters permits the prediction of the impulse response of the scatterer using the DWBA. This model does not include propagation effects, however the effect of scatterer position in the beam is included. A description of the multiple angle DWBA is given in Sec. II A. Section II B defines the size and orientation distributions that are used to generate synthetic data. Section II C describes the procedure for generating synthetic data.

A. Multiple angle DWBA scattering model

The scattering model used to obtain the impulse response of the scatterer is the DWBA.[14,17] This model relates the size, shape, and material properties of the scatterer to the complex scattering amplitude at a particular frequency. The impulse response of the scatterer can be obtained from the complex scattering amplitude by an inverse Fourier transform. The expression for the complex scattering amplitude $S(\mathbf{k})$ is

$$S(\mathbf{k}) = \frac{k_1^2}{4\pi}\int_R\int\int (\gamma_\kappa(\mathbf{r}_0) - \gamma_\rho(\mathbf{r}_0)\cos\alpha)e^{i\mathbf{k}\cdot\mathbf{r}_0}dv_0, \tag{2}$$

where

$$\mathbf{k} = k_2(\mathbf{e}_s - \mathbf{e}_i), \tag{3}$$

and

$$\cos\alpha = \mathbf{e}_s \cdot (-\mathbf{e}_i). \tag{4}$$

The scalars $k_1 = 2\pi f/c_1$ and $k_2 = 2\pi f/c_2$ are the wave numbers in the medium and body of the scatterer, respectively, $\mathbf{e}_i$ and $\mathbf{e}_s$ are unit vectors in the direction of the incident and scattered sound waves, and α is the angle between the negative incident wave vector and the scattered wave vector. The term $\gamma_\kappa(\mathbf{r}_0) - \gamma_\rho(\mathbf{r}_0)\cos\alpha$ is the gamma contrast.[17,29]

The gamma contrast inside of the volume integral is a function of the density and sound speed of the surrounding medium and the body of the scatterer where (omitting the explicit dependence on position in the body) $\gamma_\kappa = (1 - gh^2)/gh^2$ and $\gamma_\rho = (g-1)/g$. The term $g = \rho_2/\rho_1$ is the ratio of the density of the scatterer to the density of the surrounding medium and $h = c_2/c_1$ is the ratio of sound speed in the scatterer to the sound speed in the surrounding medium. Equation (2) provides the basis for the forward model used in the numerical experiments presented in this paper.

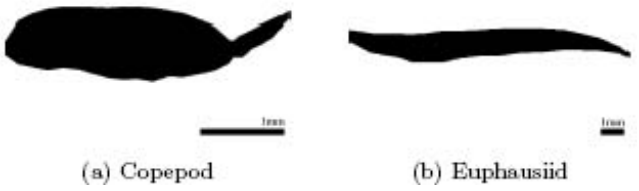


FIG. 1. Cross-sectional view of the scatterer shapes used to define each class. The copepod shape (a) and euphausiid shape (b) are displayed at different scales. The scale is defined in the lower right corner.

Using Eq. (3), the resultant wave vector $\mathbf{k}$ can be written as

$$\mathbf{k} = k_2 \frac{\sin(\pi - \alpha)}{\sin(\alpha/2)} \mathbf{d}, \tag{5}$$

where

$$\mathbf{d} = \frac{\mathbf{e}_s - \mathbf{e}_i}{\|\mathbf{e}_s - \mathbf{e}_i\|_2}, \tag{6}$$

is the unit vector that points in the direction of the difference between scattered and incident wave vectors. It is apparent that the multiple angle DWBA is closely related to the DWBA for backscatter[17] only now with a scaled and rotated wave vector. This important relationship allows the DWBA for multiple angle scatter to be computed using existing numerical methods for backscatter with only minimal modification.

B. Scatterer size and orientation distributions

An important aspect of the simulations is the choice of size and orientation distributions for the ensemble of scatterers. These data have been generated using a single shape for each class, scaled in volume and rotated in three dimensions. The shapes used for the copepod and euphausiid classes were taken from an online database of zooplankton scattering models.[18] Pictures of the cross section of the base shape used for each class are shown in Fig. 1.

The volume scaling is parametrized by a length parameter L, the length of the scatterer from head to tail. In order to simplify the treatment both length classes were drawn from uniform distributions according to $\mathcal{U}[2\text{ mm}, 4\text{ mm}]$ for the copepods and $\mathcal{U}[4\text{ mm}, 15\text{ mm}]$ for the euphausiids. Note that the length distributions overlap slightly and the distribution for the euphausiids is centered around medium length juveniles rather than the larger adults.

Similarly, for the orientation distributions, a simple approach was taken. Representing the orientation of the scatterer by a θ and ϕ angle where θ is the angle relative to the z axis, and ϕ the angle between the x and y axis (Fig. 2) the orientations were sampled uniformly in three dimensions according to

$$\theta \sim \arcsin(\mathcal{U}[-1, 1]), \tag{7}$$

$$\phi \sim \mathcal{U}[-\pi, \pi]. \tag{8}$$

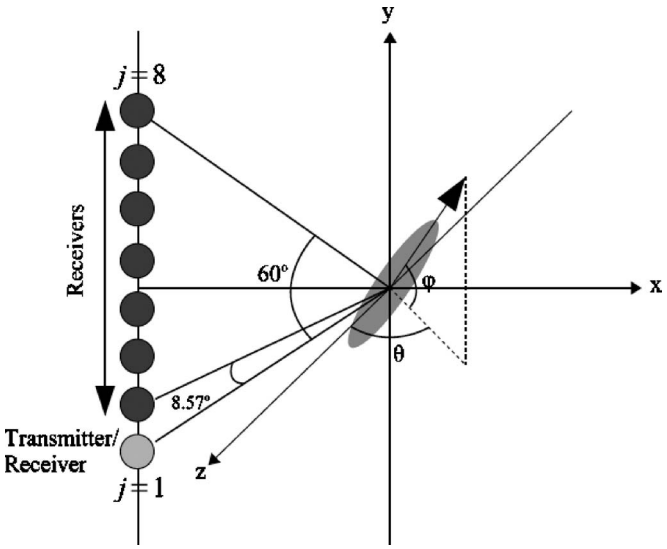


FIG. 2. View of the array configuration used to generate data. The scatterer is represented by the gray ellipse where the nose of the scatterer is directed along the unit vector defined by the angle ϕ in the x-y plane and θ from the z axis. The angular span of the array is 60° with an angular spacing of 8.57° between elements. The $j=1$ element functions as both transmitter and receiver, while the $j=2,\ldots,8$ elements receive only.

C. Creation of model realizations

The simulation of a single realization of received scatter on the array is described in this section. The configuration of the array is shown in Fig. 2. There are eight total elements, $M=8$. The $j=1$ element acts as both transmitter and receiver while the $j=2,\ldots,M$ elements act only as receivers. The total angular span of the array is 60° with an angular sampling frequency of one sample per 8.57°. The orientation of the scatterer relative to the array is shown in Fig. 2 and is defined by the angles ϕ and θ as mentioned previously.

The synthetic data are generated by predicting the received pressure signal on each of the eight array elements for a given scatterer orientation. The data generation process is represented graphically in Fig. 3. The first step is the selection of a model: copepod or euphausiid. Each three-dimensional scatterer shape is represented as a series of cylindrical segments of thickness 0.016 mm and location x, y, and z corresponding to the center. The segments have radius a, and relative density and sound speed g and h. The ith segment can be represented as the vector $\boldsymbol{\gamma}_i = (x_i, y_i, z_i, a_i, g_i, h_i)^T$ and the entire model of the body by a matrix $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_S)$ with S being the total number of segments. Ambient sound speed c is fixed at 1500 m/s. For all simulations performed in this work, the values of g_i and h_i are held constant throughout the body of the scatterer such that $g_i = 1.035\,\forall\, i$ and $h_i = 1.027\,\forall\, i$. More information about these parameters, the scattering models, and the algorithm used to compute the DWBA is available from the Acoustic Scattering Models of Zooplankton website. [18]

For each realization, a random sample from the distributions for ϕ, θ, L is selected. These parameters are combined with the sound speed c, the incident and scattered wave vectors for each array element: $\mathbf{k}_i^j$ (incident) and $\mathbf{k}_s^j$ (scattered) for $j=1,\ldots,M$, and the model for the scatterer $\boldsymbol{\Gamma}$. The

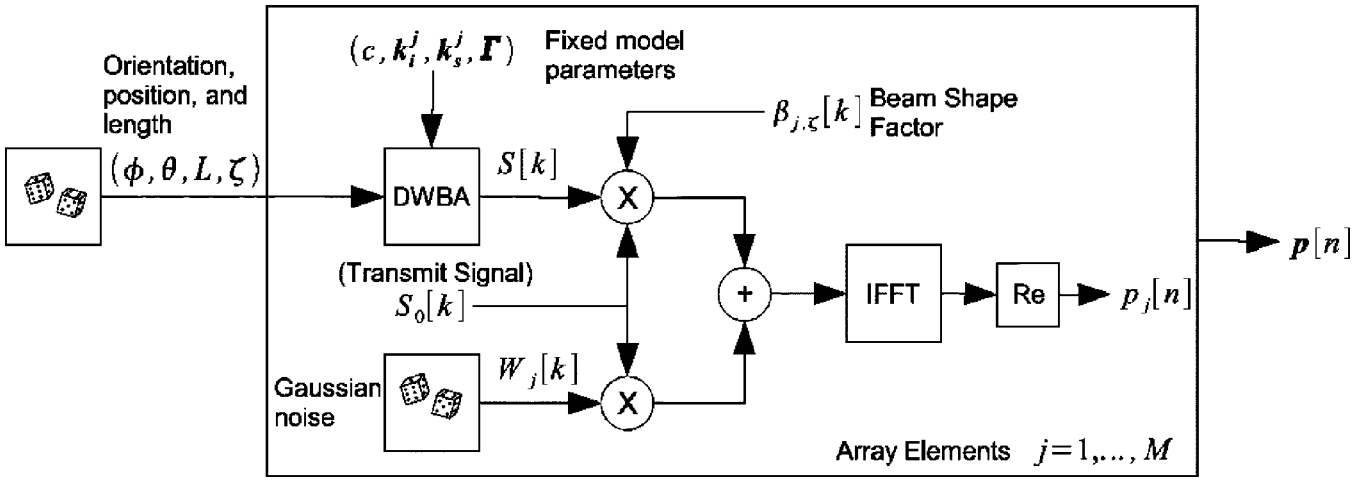


FIG. 3. Block diagram of the creation of a single realization of scattering on the array. Randomness is included in the model through the parameters $\phi, \theta, L, \boldsymbol{\zeta}$, and the Gaussian noise $W[k]$. The model parameters which are constant for all realizations are the sound speed c, the incident and scattered wave vectors for each array element $\mathbf{k}_i^j$ and $\mathbf{k}_s^j$ for $j=1,\ldots,M$ and the scatterer model $\mathbf{\Gamma}$. The beam shape factor $\beta_{j,\boldsymbol{\zeta}}[k]$ is computed for each array element and accounts for the position of the scatter in the transmit and receive beam pair. The Gaussian noise $W[k]$ is added to the product of the scattering amplitude $S[k]$ and the beam shape factor $\beta_{j,\boldsymbol{\zeta}}[k]$. The result is then multiplied by the FFT transmit signal $S_0[k]$. The real part of the inverse FFT of the product yields the pressure on the jth array element. The pressure is computed for each of the M array elements. This process is repeated 1000 times for each scatterer class to generate a set of training and test data.

DWBA is evaluated for the given model parameters at each frequency yielding the complex scattering amplitude

$$S_j[k] = \mathrm{DWBA}(\phi, \theta, L, c, \mathbf{k}_i^j, \mathbf{k}_s^j, \mathbf{\Gamma}), \tag{9}$$

where k represents the index of a particular wave number bin.

To incorporate the effect of a range-dependent sample volume in the simulation, for each realization, the scatterer is assigned a uniformly random three-dimensional position relative to the array. The position is defined by the parameter $\boldsymbol{\zeta}$.

The range-dependent sample volume is studied by calculating the position of the scatterer in the transmit and receive beam. As the beam shape changes with frequency, the incident sound intensity, and received sound intensity will vary in a predictable way. For the simulations considered here, the transducers are assumed to be disk shaped in which case the product of the transmit and receive beam shapes is given by

$$\beta_{j,\boldsymbol{\zeta}}[k] = \left| \frac{2J_1(kr\sin(\eta))}{kr\sin(\eta)} \frac{2J_1(kr\sin(\mu))}{kr\sin(\mu)} \right|, \tag{10}$$

where r is the radius of the transducer, $J_1(x)$ is the Bessel function of the first kind of order 1, and η and μ are the angles between the vector from the transducer to the scatterer position, and the incident and scattered wave vectors, respectively, for a particular transmitter-receiver pair. For a wave vector $\mathbf{k}$ and scatterer position $\boldsymbol{\zeta}$, the angles are given by

$$\eta = \arccos\left(\frac{\mathbf{k}_i^T \mathbf{k}_i - \mathbf{k}_i^T \boldsymbol{\zeta}}{\|\mathbf{k}_i\|_2 \|\mathbf{k}_i - \boldsymbol{\zeta}\|_2} \right), \tag{11}$$

and

$$\mu = \arccos\left(\frac{\mathbf{k}_s^T \mathbf{k}_s - \mathbf{k}_s^T \boldsymbol{\zeta}}{\|\mathbf{k}_s\|_2 \|\mathbf{k}_s - \boldsymbol{\zeta}\|_2} \right). \tag{12}$$

The geometry for the above-presented calculations is shown in Fig. 4. For all of the simulations, the parameter r is set to 12 mm, and the components of $\boldsymbol{\zeta}$ selected according to $\mathcal{U}[-5\ \mathrm{mm}, 5\ \mathrm{mm}]$. The horizontal distance from $\boldsymbol{\zeta}=\mathbf{0}$ to the array is defined to be 3 m, and thus the majority of scatterer positions are within the −6 dB beam width of the array elements at the highest frequency. To compare the effect of the sample volume on the classification performance, the simulations are performed with and without including the beam shape factor. The case without the beam shape factor is equivalent to setting $\beta_{j,\boldsymbol{\zeta}}[k]=1$ for all realizations.

The received echoes in any practical system will be corrupted by noise due to reverberation, electronics, and other sound sources. Noise due to reverberation will be in the same frequency band as the received echo whereas noise from electronics and other sound sources will have energy in other frequency bands as well as the band of the received echo. Since out of band noise can be reduced by filtering, rever-

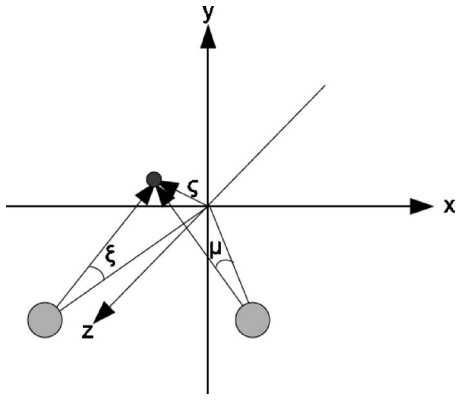


FIG. 4. View of the method for calculating beam shape for a scatterer positioned at $\boldsymbol{\zeta}$ and a transmitter and receiver pair. The angles between the scatterer position vector, and the vector to each element are defined by η and μ. The angles are used to compute the change in sound intensity as a function of frequency.

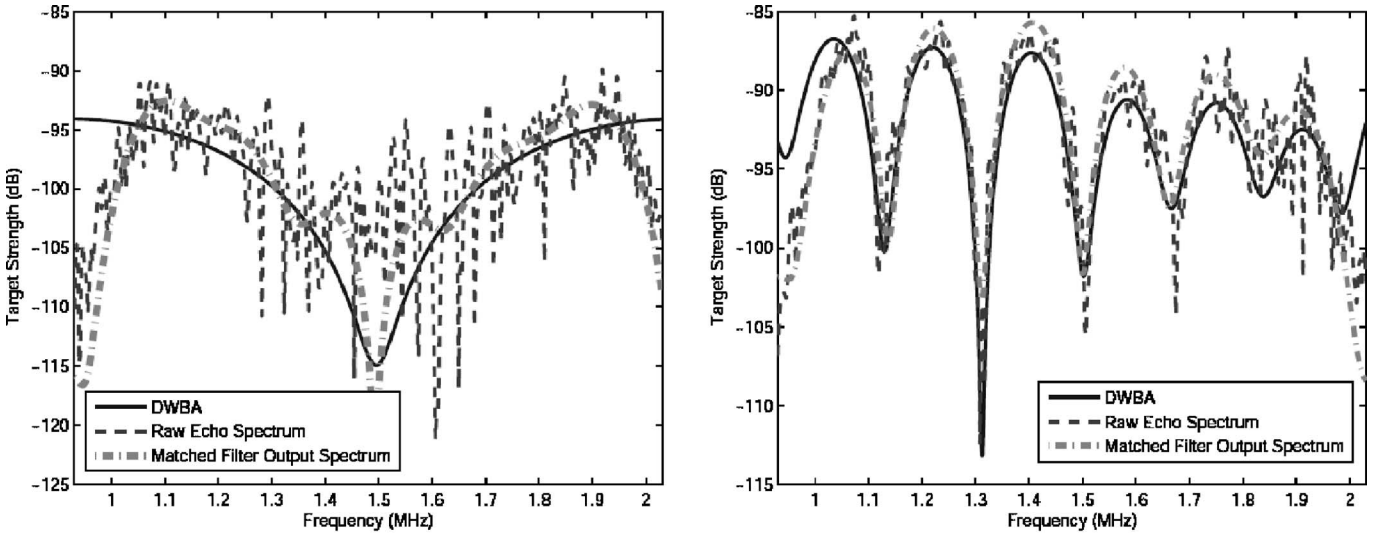


FIG. 5. Typical examples of simulated data for both the copepod and euphausiid class. The estimated target strength using the raw echo, and the matched filter output are plotted on top of the target strength as modeled by the DWBA. The high level of noise is clearly visible, as is the improvement obtained from the matched filter.

beration noise in the same frequency band as the transmit signal is added to the scattering amplitude. The noise is generated by taking the product of the fast Fourier transform (FFT) of the transmit signal $S_0[k]$ with the FFT of a realization of white Gaussian noise $W_j[k]$.

A constant reverberation level is used and thus the signal to noise ratio (SNR) varies as a function of length and orientation of the scatterer. The SNR for the copepod data ranges from −11 to 24 dB whereas the SNR for the euphausiid data ranges from −20 to 35 dB. These ranges for SNR were selected such that the average SNR was close to 7 dB for the copepods and 15 dB for the euphausiids. These values were deemed to be comparable to what is achieved in practical systems [30,31]. The noise level can also be defined in terms of an equivalent target strength of −110 dB.

In generating the data, the noise term $W_j[k]$ is first added to the product of the scattering amplitude and the beam shape factor at each wave number bin to yield the noisy scattering amplitude

$$\hat{S}_j[k] = S_j[k]\beta_{j,\zeta}[k] + W_j[k]. \tag{13}$$

The convolution defined in Eq. (1) is accomplished in the frequency domain by computing the product of the FFT of the transmit signal $S_0[k]$ with the noisy scattering amplitude $\hat{S}_j[k]$. Using Eq. (13), the pressure signal on the jth element is then obtained by an inverse FFT (IFFT),

$$p_j[n] = \text{Real}\left[\frac{1}{N}\sum_{k=0}^{N-1} S_0[k]\hat{S}_j[k]e^{2\pi ink/N}\right]. \tag{14}$$

This process is repeated for each of the M array elements building up the vector

$$\mathbf{p}[n] = (p_1[n], \ldots, p_M[n])^T. \tag{15}$$

In order to explore the classification success as a function of carrier frequency and bandwidth several different types of signals were used. For the narrowband signals (10% bandwidth) frequencies of 1 and 2 MHz were selected. For the broadband signal, a linear frequency modulated (LFM) chirp was used with a starting frequency of 1 MHz and ending frequency of 2 MHz. The signal duration and energy was kept constant for all signals. The range of frequencies was selected based on past experience with measuring scatter from animals of the size considered here.

To improve the SNR of the data input to the feature extraction algorithms, the raw echo data resulting from the simulation is passed through a matched filter.[32] For the transmit signal $s_0[n]$, and a received echo on array element j defined by $p_j[n]$, the output of the matched filter is

$$\mathcal{M}_j[n] = \sum_{p=-\infty}^{\infty} s_0[p]p_j[p-n]. \tag{16}$$

The matched filter output is windowed around the peak in the output with a window size of $W=50$, corresponding to a time of 5 μs or a distance of 7.5 mm. For the broadband signal the time-bandwidth product is 120, yielding a processing gain of roughly 20 dB. For the narrowband signals, the time-bandwidth product is much lower, and thus the processing gain is low as well, around 10 dB.

To visualize the type of noise, and its effect on the estimation of the scattered signal, the model target strength is displayed along with the estimated target strength for the broadband signal type in Fig. 5.

III. FEATURE EXTRACTION

In order to facilitate the classification procedure the data are mapped to a feature space which dramatically reduces the dimensionality of the data while simultaneously highlighting interclass differences. The three feature spaces used in this work are described in the following.

A. Single frequency based feature space

To explore the result of using only a single frequency rather than a broad spectrum of frequencies, a single frequency based feature space is defined in which the FFT of the matched filter output on each array element is computed and its magnitude squared is integrated over a small bandwidth. Only the narrowband data are used in this feature space. The resulting sum squared magnitudes on each array element are then combined to form a single feature vector. Specifically, assuming that the jth array element collects N samples, the power in the narrow bandwidth of the signal ($k_{\min}$ to $k_{\max}$) is

$$P_j = \sum_{k=k_{min}}^{k_{max}} \left| \sum_{n=0}^{N-1} \mathcal{M}_j[n] e^{-2\pi i n k/N} \right|^2. \tag{17}$$

For the simulations presented here, $N=1200$. The single frequency feature vector is then defined as

$$\mathbf{y} = (P_1, P_2, \ldots, P_M)^T. \tag{18}$$

The bandwith ($k_{\min}$ to $k_{\max}$) is set equal to the transmit signal bandwidth of 10% of the center frequency. The center frequencies used are 1 and 2 MHz.

B. Discrete cosine transform based feature space

The discrete cosine transform (DCT) has numerous qualities that make it attractive as a feature mapping. For one, the coefficients of the DCT are uncorrelated. It can also be shown that the DCT can embed most of the energy in the data into a small number of coefficients. While there is no guarantee that such an embedding will yield a discriminant feature space, this is often the case in practice. The DCT based feature space uses the power spectrum of the matched filter output for the broadband 1–2 MHz data. The power spectrum is computed as

$$P_j[k] = \left| \sum_{n=0}^{N-1} \mathcal{M}_j[n] e^{-2\pi i n k/N} \right|^2. \tag{19}$$

For the results presented here $N=1200$. Having computed $P_j[k]$ for each array element, the DCT of the power spectrum is computed as

$$E_j^{\mathrm{DCT}}[l] = \sqrt{\frac{2}{N}} \beta[l] \sum_{k=0}^{N-1} P_j[k] \cos\left(\frac{\pi l (2k+1)}{2N} \right), \tag{20}$$

where

$$\beta[l] = \begin{cases} \dfrac{1}{\sqrt{2}}, & l = 0 \\ 1, & l = 1, \ldots, N-1 \end{cases}. \tag{21}$$

The values of the K largest (ordered by magnitude) coefficients in the transform are retained in the feature vector for the jth array element

$$\mathbf{y}_j = (E_j^{\mathrm{DCT}}[l^1], \ldots, E_j^{\mathrm{DCT}}[l^K])^T, \tag{22}$$

where the features are arranged such that $l^1 \leqslant l^2 \leqslant \cdots \leqslant l^K$. This procedure can be interpreted as an adaptive threshold of the DCT of the power spectrum. Finally, the feature vectors at each array element are combined into a single feature vector,

$$\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_M^T)^T. \tag{23}$$

The feature vector, $\mathbf{y}$ captures the K most energetic wave number bins in the power spectrum of the received signal at each element of the array. A range of values for K were analyzed. It was found that the values $K=1$, $K=2$, and $K=4$ yield the best performance. As K increases, the feature vector is able to capture more of the variability in the frequency response of the scatterer at the cost of a larger feature space dimension. A wave number bin width of $\Delta l = 35$ rad/m was used throughout the simulation.

C. Frequency correlation based feature space

One of the major drawbacks of the previous two feature spaces is that they do not naturally combine the multiple angle data. Instead, the features from each angle of the multiple angle data are lumped together as one big feature vector. This can cause problems in the case of the DCT feature space as the dimensionality of the feature space grows as K times the number of angles M. In this section, a feature mapping which combines the multiple angle data systematically while extracting the features is defined. The features are the eigenvalues of the frequency correlation matrix.

The frequency correlation matrix is obtained by computing the correlation between all pairs of received wave forms in the frequency domain. Specifically, the correlation matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \mathbf{F}^H \mathbf{F}, \tag{24}$$

where

$$\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_M), \tag{25}$$

and

$$f_j[k] = \sum_{n=0}^{N-1} \mathcal{M}_j[n] e^{-2\pi i k n/N} \tag{26}$$

is the FFT of the matched filter output for the received data on element j. Thus, the individual elements of the frequency correlation matrix are the cross correlations between the Fourier transforms of the data on the array elements

$$C_{ij} = \mathbf{f}_i^H \mathbf{f}_j. \tag{27}$$

The features are extracted from the frequency correlation matrix using an eigenvalue decomposition[33]

$$\mathbf{\Lambda} = \mathbf{Q}^H \mathbf{C} \mathbf{Q}, \tag{28}$$

where $\mathbf{\Lambda} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_M)$ is the diagonal matrix of eigenvalues. The feature vector is then formed by taking the diagonal elements of $\mathbf{\Lambda}$,

$$\mathbf{y} = (\lambda_1, \ldots, \lambda_M)^T. \tag{29}$$

The eigenvalues (features) are not simply related to the data at each angle as in the case of the previous two feature spaces. Each eigenvalue is derived from data at all angles. This is the key benefit over the other two feature spaces. The number of nonzero eigenvalues is upper bounded by the number of angles, yet may be lower depending on the degree to which the echoes received at each angle are correlated with one another in the frequency domain. For example, if

$$\mathbf{f}_i^H \mathbf{f}_j \approx \begin{Bmatrix} 0 & \text{for } i \neq j \\ \kappa & \text{for } i = j \end{Bmatrix}, \tag{30}$$

the frequency correlation matrix $\mathbf{C} \approx \kappa \mathbf{I}$ and the eigenvalue spread is nearly flat. In contrast, if

$$\mathbf{f}_i^H \mathbf{f}_j \approx \kappa \ \ \forall\, i,j, \tag{31}$$

the frequency correlation matrix is approximately rank one, and the eigenvalue value spectrum will be highly peaked at the first eigenvalue. The first example can be thought of as representing a complex shape, where the spectrum of the received signal varies substantially as a function of angle. This second example corresponds to scattering from an angularly symmetric shape. Therefore, in the presence of noise, the variability at each array element is due only to noise.

IV. CLASSIFICATION OF FEATURES

Given a set of features that have been extracted from these data, the next task is to develop a method for assigning a class label to each feature so as to minimize a particular loss function. As is commonly done in pattern classification, the "0-1" loss function is applied which assigns equal penalties to classification errors made for either class.[34] In the zooplankton classification problem considered here, this is a reasonable loss function due to the fact that each class has, in effect, equal significance. It can be shown[34] that the classification rule which minimizes the "0-1" loss function is the Bayes decision rule (BDR)

$$i^* = \underset{i}{\operatorname{argmax}}\, P_{C|\mathbf{Y}}(i|\mathbf{y}), \tag{32}$$

where the class i^*, having the maximum *a posteriori* probability given the feature vector $\mathbf{y}$, is chosen. The BDR can be written in terms of the class conditional density (CCD) $p_{\mathbf{Y}|C}(\mathbf{y}|i)$ using Bayes rule, and assuming a prior class distribution $P_C(i)$, as

$$i^* = \underset{i}{\operatorname{argmax}}\, p_{\mathbf{Y}|C}(\mathbf{y}|i) P_C(i). \tag{33}$$

In practice, the prior probability may or may not be known. For the procedure considered here, it is assumed that the priors for each class are equal. As a result, that term drops from the maximization. The remaining task is that of maximizing the CCD which is equivalent to computing the maximum likelihood estimate of the class label.

Unfortunately, the CCD is almost always unknown. In the best case, only the form of the density is known, but not the parameters that define the actual shape. This is one of the fundamental difficulties encountered in pattern classification and is the point at which *a priori* knowledge or training data must be used to learn about the structure of $p_{\mathbf{Y}|C}(\mathbf{y}|i)$.

Here, two popular classifiers are considered: the nearest neighbor (NN) classifier, and the multilayer perceptron (MLP) classifier. The properties of each of these classifiers are briefly reviewed as the implementations used here are standard.

The NN classifier assigns a class label to a new pattern based on the label of the training pattern which is "nearest" to the new pattern according to a particular distance metric. For a given training set $\mathcal{D} = \{(\mathbf{y}_1, i_1), \ldots, (\mathbf{y}_N, i_N)\}$ where $\mathbf{y}_n$ is a feature extracted from the data according to the methods defined in Sec. III and i_n is the associated class label, the NN classifier under the 2-norm assigns the label i_k where k is the index of the nearest neighbor,

$$k = \underset{i}{\operatorname{argmin}} (\mathbf{y} - \mathbf{y}_i)^T (\mathbf{y} - \mathbf{y}_i). \tag{34}$$

In contrast to the NN method, the MLP tries to learn the mapping from feature space to class label space using multiple levels of weighted combinations of the components of the features rather than using the training data explicitly to represent the underlying CCDs. In essence, the MLP learns to approximate $P_{C|\mathbf{Y}}(i|\mathbf{y})$ via experience gained from analyzing numerous examples. It has been shown that this type of classifier can yield very good results in underwater target classification[19] as it is one of the best methods for approximating a high dimensional function.

Given the two class problem, the MLP has two output nodes. The number of input nodes is the same as the number components of the feature vector $\mathbf{y}$. A single hidden layer is used with the number of nodes selected to be twice the number of input nodes. The network is compactly expressed[35] as

$$c_k = U\left(\sum_{j=0}^{2M} \tilde{w}_{kj} V\left(\sum_{i=0}^{M} w_{ji} y_i \right) \right), \tag{35}$$

where U and V are nonlinear mapping functions, $\tilde{w}_{kj}$ and w_{ji} are network weights, y_i is the ith component of the feature vector, and c_k is the kth component of the classification vector. The weight matrices of the network are initialized randomly at the start of training, and updated at each iteration so as to minimize the error on the training set. Both mapping functions are selected to be the softmax function,[34] and the network is trained using the scaled conjugate gradient method. Prior to training and testing, all inputs to the network are z-scaled in the log domain. The training is implemented in MATLAB (The Mathworks; Natick, MA) using the NETLAB toolbox.[36]

V. RESULTS AND DISCUSSION

Classifier performance. The classifiers defined in Sec. IV are now evaluated quantitatively on a set of test data

mapped into each of the feature spaces defined in Sec. III. The results are displayed as the absolute probability of error as a function of the number of angles (or array elements) that are combined in the classifier. Specifically, the number of angles is equal to the number of array elements included in order starting from element 1. So, for example, three angles would correspond to using array elements 1, 2, and 3, and four angles would correspond to using elements 1, 2, 3, and 4. The probability of error is computed according to

$$p(\text{error}) = \frac{p(c|e) + p(e|c)}{2}, \tag{36}$$

where $p(c|e)$ is the probability of classifying an euphausiid as a copepod, and $p(e|c)$ is the probability of classifying a copepod as an euphausiid. Here, the fact that each class is equally likely in this simulation has been used. A consequence of the equal representation for each class is that a system which randomly guesses the class would have a probability of error of 50%. Therefore, 50% probability of error can be achieved with no effort, and any classification strategy should have an error below 50%. The classification experiment is performed both for the case where beam shape is neglected from the simulation, and where the beam shape is included. The results of the classification for both cases of data are shown in Fig. 6.

The 1 and 2 MHz curves result from using the single frequency feature space and the respective narrowband data. The DCT and CM curves result from using the DCT and frequency correlation feature spaces with the broadband 1–2 MHz LFM chirp data. The DCT(1), DCT(2), and DCT(4) curves apply the DCT method outlined in Sec. III with $K=1$, $K=2$, or $K=4$, respectively. Figure 6 illustrates that there is a general trend of decreasing probability of error as more angles are used in the classification. The amplitude and frequency response of acoustic scatter from crustacean zooplankton is directly related to the scatterer shape, size, and orientation. The addition of more angles in the classifier can be interpreted physically as observing the scatterer from multiple views. The probability of error is reduced as more angles are used in essence because there is less uncertainty about the shape of the scatterer.

When only one angle is used, observing scatter over a broad range of frequencies (1–2 MHz) reduces the probability of error substantially over the single frequency case. This is a consequence of the fact that single frequency scatter is much more sensitive to scatterer orientation and size then broadband scatter due to the effect of coherent interference at a given frequency. As broadband data excites many frequencies, it is far more robust to changes in scatterer shape and orientation.

The effect of including beam shape on the probability of error can be clearly seen by comparing the top and bottom rows of Fig. 6. The effect is essentially to shift the curves toward higher probability of error. This is a consequence of the fact that the random position of the scatterer in the beam adds another kind of noise to the data. However, it is possible this that kind of noise can be corrected by exploiting the angular diversity of the array to locate the scatterer in the beam, and correct for the beam shape. While not considered in this paper, this process will be investigated in future work.

The frequency correlation feature space clearly outperforms all of the other feature spaces for both classifiers. This is to be expected since it is the only feature space which naturally combines the echoes at each angle when extracting features. It may be inferred from the poor performance when only a single angle is used that the variation in the echo as a function of angle is most discriminant between the two classes. The frequency correlation feature space efficiently extracts this information from these data. The performance (in terms of correct and misclassification) of each classifier using all eight angles with, and without including beam shape is displayed in Tables I and II. Interestingly, there is a wide variation in the correct and misclassification results for the different feature spaces. The improvement in classification performance as additional angles, and as a result increased array aperture, are used is a direct consequence of having additional independent views of the scatterer. The additional views reduce the uncertainty in the shape of the scatter by way of the intimate link between scatterer shape, and angularly varying scattering amplitude as defined in Eq (2). Since the two classes of scatterers have distinct shapes, the reduced uncertainty in shape leads to improved classification performance. In general, all of the feature spaces, and classifiers have slightly higher accuracy for the copepod class rather than the euphausiid class except for the frequency correlation feature space which is inconsistent between the two classification algorithms. This is likely a consequence of the large number of orientations for which scattering from the euphausiid is very weak due to the elongated body. There is also a systematic increase in probability of error for the 2 MHz data over the 1 MHz data for both classifiers. This is likely caused by a greater similarity in scattering amplitude at 2 than 1 MHz between the two classes. This could be caused by the fact that at 2 MHz, the scattering is further into the geometric regime, and thus the scattering amplitude is less sensitive to the scatter size. Finally, the best results, are misclassifications of 9.0% and 6.2% for the copepod and euphausiid, respectively. As a result, the total absolute probability of error is 7.6% in the best case. This gives an improvement over random guessing of 84.8%.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, the use of multiple angle acoustic scatter to discriminate between two classes of ecologically important zooplankton has been explored using simulations. The research is motivated by the current need for more descriptive acoustic sensors for studying zooplankton *in situ*. Past work in this area has been limited by the inherent ambiguity in discrimination ability due to the sensitivity of acoustic scattering to material properties and scatterer orientation. These difficulties have been confirmed here where it has also been shown that it is possible to use scatter measured over a multiplicity of angles to achieve a higher rate of correct classification. Using synthetic data, generated via the use of the distorted wave Born approximation, two ecologically impor-

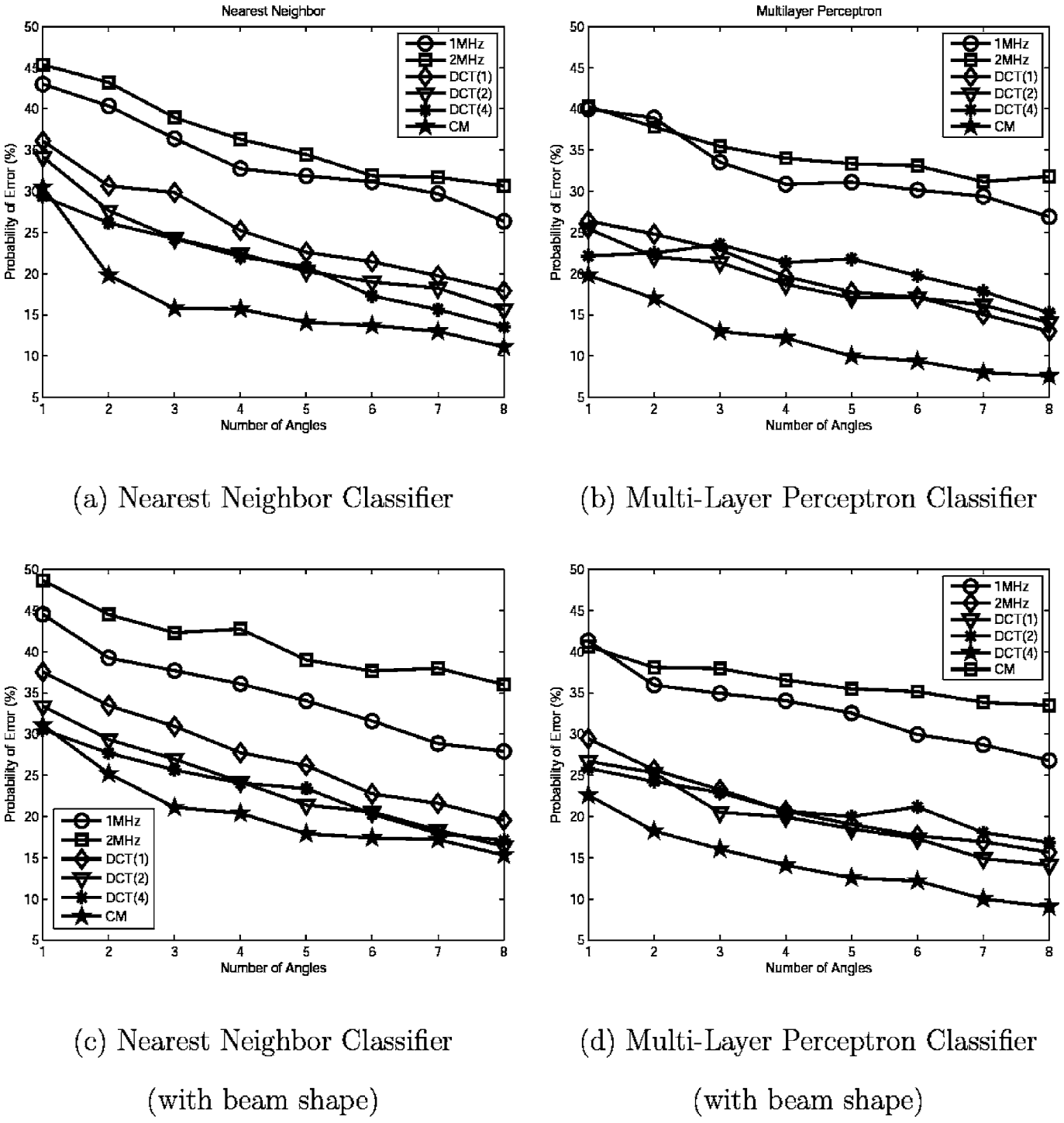


FIG. 6. Comparison between the NN classifier and the MLP classifiers for the case of no beam shape (a), (b) and beam shape (c), (d), in six different feature spaces. The 1 and 2 MHz curves correspond to the single frequency based feature space. The DCT(1), DCT(2), and DCT(4) feature spaces use the DCT method outlined in Sec. III with $K=1$, $K=2$, or $K=4$, bin indices included at each angle. The CM curves result from the frequency correlation feature space outlined in Sec. III.

tant classes of zooplankton—copepods and euphausiids—were classified. The classification performance, measured in terms of probability of error, is dramatically improved over single angle observation methods via the use of additional angles. This improvement is even more substantial when broadband scatter is used.

The simulations performed here were geared toward a practical system which could be deployed in the field. Therefore, constraints were placed on the bandwidth of the transmit signal, and the angular distribution of the receivers in this context. The length distributions for both classes were chosen to be typical of those encountered in the Southern California region.[37,38] In order to understand the ramifications of the proposed method in the presence of noise, a constant level of noise that resulted in an average SNR of 7 dB for the copepod and 15 dB for the euphausiid, or an equivalent target strength of −110 dB, was used. This noise level is consistent with practical systems that have been used

TABLE I. The correct and misclassification probabilities for each classifier when beam shape is excluded from the simulation.

	Copepod		Euphausiid		
Classifier	Correct (%)	Mis (%)	Correct (%)	Mis (%)	Total p (error) (%)
NN/1 MHz	76.1	23.9	71.2	28.8	26.3
NN/2 MHz	74.1	25.9	64.6	35.4	30.6
NN/DCT (4)	87.5	12.5	85.4	14.6	13.5
NN/CM	90.8	9.2	87.0	13.0	11.1
MLP/1 MHz	76.3	23.7	70.4	29.6	26.6
MLP/2 MHz	70.3	29.7	66.1	33.9	31.8
MLP/DCT (4)	85.0	15.0	84.5	15.5	15.2
MLP/CM	91.0	9.0	93.8	6.2	7.6

in the field.[30,31] In addition, although a strong effort was expended in order to make the work realistic, the performance of a field system may be limited by issues that have not been considered in this work. Specifically, the models used here, while accurate for weak sound scattering, do not include variability due to individual shape, or body pose. Furthermore, uniform orientation distributions were used here, where as in the field, the distributions may be different. Given the promising results observed here, these additional degrees of freedom certainly warrant further investigation through more complex simulations, as well as observation of live animals.

A curious, but potentially very helpful aspect of our result is that a one-dimensional array is capable of capturing enough information from a random three-dimensional orientation to yield good classification performance. It may therefore be that simple array geometries, which can dramatically reduce the development and deployment cost associated with such systems, constitute a pragmatic solution to the *in situ* classification of zooplankton after all.

ACKNOWLEDGMENTS

The authors would like to thank D. E. McGehee, M. Benfield, D. V. Holliday, and G. Greenlaw for development and maintenance of the Advanced Multifrequency Inversion Methods for Classifying Acoustic Scatters website, two anonymous reviews for helpful comments on the manuscript, and California Sea Grant for funding this research.

TABLE II. The correct and misclassification probabilities for each classifier when beam shape is included in the simulation.

	Copepod		Euphausiid		
Classifier	Correct (%)	Mis (%)	Correct (%)	Mis (%)	Total p (error) (%)
NN/1 MHz	74.1	25.9	70.1	29.9	27.9
NN/2 MHz	67.8	32.2	60.2	39.8	36.0
NN/DCT (4)	83.5	16.5	82.3	17.7	17.1
NN/CM	85.8	14.2	81.5	18.5	16.3
MLP/1 MHz	75.8	24.5	74.2	25.8	25.0
MLP/2 MHz	69.9	30.1	63.1	36.9	33.5
MLP/DCT (4)	85.3	14.7	81.0	19.0	16.8
MLP/CM	91.2	8.8	90.7	9.3	9.0

[1]D. McNaught, "Acoustical determination of zooplankton distributions," in The 11th Annual Conference on Great Lakes Research, 1968, pp. 76–84.
[2]D. V. Holliday (1977), "Extracting biophysical information from the acoustic signatures of marine organisms," in *Oceanic Sound Scattering Prediction*, edited by N. R. Anderson and B. J. Zahuraned (Plenum, New York), pp. 162–211.
[3]C. Greenlaw, "Acoustic estimation of zooplankton populations," Limnol. Oceanogr. **24**, 226–242 (1979).
[4]D. Holliday, R. Pieper, and G. Kleppel, "Determination of zooplankton size and distribution with multi-frequency acoustic technology," J. Cons., Cons. Int. Explor. Mer **41**, 226–238 (1989).
[5]D. McGehee, D. Demer, and J. Warren, "Zooplankton in the Ligurian Sea. I. Characterization of their dispersion, relative abundance and environment during summer 1999," J. Plankton Res. **26**, 1409–1418 (2004).
[6]M. McManus, O. Cheriton, P. Drake, D. Holliday, C. Storlazzi, P. L. Donaghay, and C. Greenlaw, "Effects of physical processes on structure and transport of thin zooplankton layers in the coastal ocean," Mar. Ecol.: Prog. Ser. **301**, 199–215 (2005).
[7]L. Martin, T. Stanton, P. Wiebe, and J. F.Lynch, "Acoustic classification of zooplankton," ICES J. Mar. Sci. **53**, 217–224 (1996).
[8]L. Traykovski, T. Stanton, P. Wiebe, and J. Lynch, "Model-based covariance mean variance classification techniques: Algorithm development and application to the acoustic classification of zooplankton," IEEE J. Ocean. Eng. **23**, 344–364 (1998).
[9]L. Goodman, "Acoustic scattering from ocean microstructure," J. Geophys. Res., [Atmos.] **95**, 11557–11573 (1990).
[10]J. S. Jaffe, "Using multiple-angle scattered sound to size fish swim bladders," ICES J. Mar. Sci. **63**, 1397–1404 (2006).
[11]K. G. Foote, "Rather high frequency sound scattering by swimbladdered fish," J. Acoust. Soc. Am. **78**, 688–700 (1985).
[12]T. Stanton, D. Chu, and P. Wiebe, "Acoustic scattering characteristics of several zooplankton groups," ICES J. Mar. Sci. **53**, 289–295 (1996).
[13]T. Stanton, D. Chu, and P. Wiebe, "Sound scattering by several zooplankton groups. II. Scattering models," J. Acoust. Soc. Am. **103**, 236–253 (1998).
[14]T. Stanton and D. Chu, "Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: Euphausiids and copepods," ICES J. Mar. Sci. **57**, 793–807 (2000).
[15]D. Reeder and T. Stanton, "Acoustic scattering by axisymmetric finite-length bodies: An extension of a two-dimensional conformal mapping method," J. Acoust. Soc. Am. **116**, 729–746 (2004).
[16]D. McGehee, R. O'Driscoll, and L. Traykovski, "Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz," Deep-Sea Res., Part II **45**, 1273–1294 (1998).
[17]A. Lavery, T. Stanton, D. McGehee, and D. Z.Chu, "Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton," J. Acoust. Soc. Am. **111**, 1197–1210 (2002).
[18]D. E. McGehee, M. Benfield, D. V. Holliday, and C. Greenlaw, "Advanced multifrequency inversion methods for classifying acoustic scatterers," http://zooplankton.lsu.edu/scattering_models/MultifreqInverseMethods.html.
[19]M. Azimi-Sadjadi, D. Yao, Q. Huang, and G. Dobeck, "Underwater target classification using wavelet packets and neural networks," IEEE Trans. Neural Netw. **11**, 784–794 (2000).
[20]N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Bucaro, and G. Dobeck, "Class-based target identification with multiaspect scattering data," IEEE J. Ocean. Eng. **28**, 271–282 (2003).
[21]S. Ji, X. Liao, and L. Carin, "Adaptive multiaspect target classification and detection with hidden Markov models," IEEE Sens. J. **5**, 1035–1042 (2005).
[22]M. Robinson, M. Azimi-Sadjadi, and J. Salazar, "Multi-aspect target discrimination using hidden Markov models and neural networks," IEEE Trans. Neural Netw. **16**, 447–459 (2005).
[23]P. Bharadwaj, P. Runkle, L. Carin, J. Berrie, and J. Hughes, "Multiaspect classification of airborne targets via physics-based HMMs and matching pursuits," IEEE Trans. Aerosp. Electron. Syst. **37**, 595–606 (2001).
[24]B. Pei and M. Bao, "Multi-aspect radar target recognition method. Based on scattering centers and HMMs," IEEE Trans. Aerosp. Electron. Syst. **41**, 1067–1074 (2005).
[25]Y. Dong, P. Runkle, L. Carin, R. Damarla, A. Sullivan, M. Ressler, and J. Sichina, "Multi-aspect detection of surface and shallow-buried unexploded ordnance via ultra-wideband synthetic aperture radar," IEEE Trans. Geosci. Remote Sens. **39**, 1259–1270 (2001).
[26]P. Runkle, P. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov

models for multiaspect target classification," IEEE Trans. Signal Process. **47**, 2035–2040 (1999).
[27]D. Li, M. Azimi-Sadjadi, and M. Robinson, "Comparison of different classification algorithms for underwater target discrimination," IEEE Trans. Neural Netw. **15**, 189–194 (2004).
[28]A. Fleminger, J. D. Isaacs, and J. G. Wyllie, "Zooplankton biomass measurements from calcofi cruises of July 1955 to 1959 and remarks on comparison with results from October, January, and April cruises of 1955 to 1959," Technical Rep., CalCOFI Atlas No. 21, 1974.
[29]P. Morse and K. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, 1968).
[30]J. Jaffe, E. Reuss, and G. Chandran, "Ftv: A sonar for tracking macrozooplankton in three dimensions," Deep-Sea Res., Part I **42**, 1495–1512 (1995).
[31]A. Genin, J. S. Jaffe, R. Reef, C. Richter, and P. J. S. Franks, "Swimming against the flow: A mechanism of zooplankton aggregation," Science **308**, 860–862 (2005).
[32]S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall, Upper Saddle River, NJ, 1998), Vol. **1**.
[33]T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing* (Prentice Hall, Upper Saddle River, NJ, 2000).
[34]R. Duda, P. Hart, and D. G. Stork, *Pattern Classification* (Wiley Interscience, New York, 2000).
[35]C. M. Bishop, "Neural networks and their applications," Rev. Sci. Instrum. **65**, 1803–1932 (1994).
[36]I. Nabney, *Netlab: Algorithms for Pattern Recognition* (Springer, New York, 2001).
[37]M. D. Ohman and B. E. Lavaniegos, "Comparative zooplankton sampling efficiency of a ring net and bongo net with comments on pooling of subsamples," Technical Rep., CalCOFI Rep., Vol. **43**, 2002.
[38]E. Brinton and J. G. Wyllie, "Distributional atlas of euphausiid growth stages off southern California, 1953–1956," Technical Rep., CalCOFI Atlas, No. 24, 1976.

Time reversal imaging for sensor networks with optimal compensation in time

Grégoire Derveaux[a)]
INRIA Domaine de Voluceau BP105, 78153 Le Chesnay, Cedex France

George Papanicolaou
Department of Mathematics, Stanford University, Stanford, California 94305

Chrysoula Tsogka
Department of Mathematics, University of Chicago, Chicago, Illinois 60637

(Received 4 July 2006; revised 13 November 2006; accepted 11 January 2007)

Using extensive numerical simulations, several distributed sensor imaging algorithms for localized damage in a structure are analyzed. Given a configuration of ultrasonic transducers, a full response matrix for the healthy structure is assumed known. It is used as a basis for comparison with the response matrix that is recorded when there is damage. Numerical simulations are done with the wave equation in two dimensions. The healthy structure contains many scatterers. The aim is to image point-like defects with several regularly distributed sensors. Because of the complexity of the environment, the recorded traces have a lot of delay spread and travel time migration does not work so well. Instead, the traces are back propagated numerically assuming that there is some knowledge of the background. Since the time at which the back propagated field will focus on the defects is unknown, the Shannon entropy or the bounded variation norm of the image is computed and the time where it is minimal is picked. This imaging method performs well because it produces a tight image near the location of the defects at the time of refocusing. When there are several defects, the singular value decomposition of the response matrix is also carried out. [DOI: 10.1121/1.2536888]

PACS number(s): 43.60.Pt, 43.60.Gk, 43.60.Tj [PEB] Pages: 2071–2085

I. INTRODUCTION

Signals recorded by sensors placed in a structure can be used to monitor its integrity and to detect the appearance of defects. In ultrasonic nondestructive testing the sensors are often small, isotropic transducers that operate in broadband regimes.

The regular monitoring of a structure generates huge amounts of data which are largely redundant and difficult to use. Detection can be done, in principle, by comparing the response of the structure in its normal state with that recorded when defects are present. The literature for this problem goes back some 30 years.[1,2] To what extent can these responses be used to also image the defects? Imaging the location and shape of the defects is a much more complex problem that has received a lot of attention when arrays are used.[3] Imaging with distributed sensors in structural health monitoring applications is considered in Refs. 4–7. We consider this question in this paper using time-reversal imaging methods. Numerical back propagation of the recorded signals will focus them near the defects, which behave like weak secondary sources. However, we do not know at what time during the back propagation process this focusing will occur since the location of the defects is not known. We propose here an algorithm for optimally stopping the back propagation by using an entropy or bounded variation norm for the image which are sparsity norms. Indeed at refocusing time, the information contained in the field is localized and it is natural to use norms that characterize that sparsity. When several small defects are present we image using the singular value decomposition of the response matrix together with the optimally stopped back propagation. We carry out extensive numerical simulations in order to assess the effectiveness of this algorithm. We find that back propagation with optimal stopping works well, especially when the Green's function for the structure is known. Travel time migration imaging assumes that the background is homogeneous. It is less computationally demanding, but it does not work as well because it does not use information about the background. We therefore expect that the difference between our method and travel time migration imaging will become more important as the complexity of the background incresases. Full wave migration in the known background is computationally very demanding and therefore not competitive.

One important difficulty for the data analysis is due to the complexity of the propagation of ultrasound in thin composite structures. Lamb waves propagating in a thin elastic structure are dispersive. Also, structures like an airplane or a bridge contain many objects like stiffeners or rivets. The propagating waves will be scattered at all these objects and the recorded signals will have long codas. In this paper, we are interested in imaging in such hetereogeneous media. We do not consider dispersive effects. Time reversal with Lamb waves for an active source problem has been investigated in

[a)]Author to whom correspondence should be addressed. Electronic mail: gregoire.derveaux@inria.fr

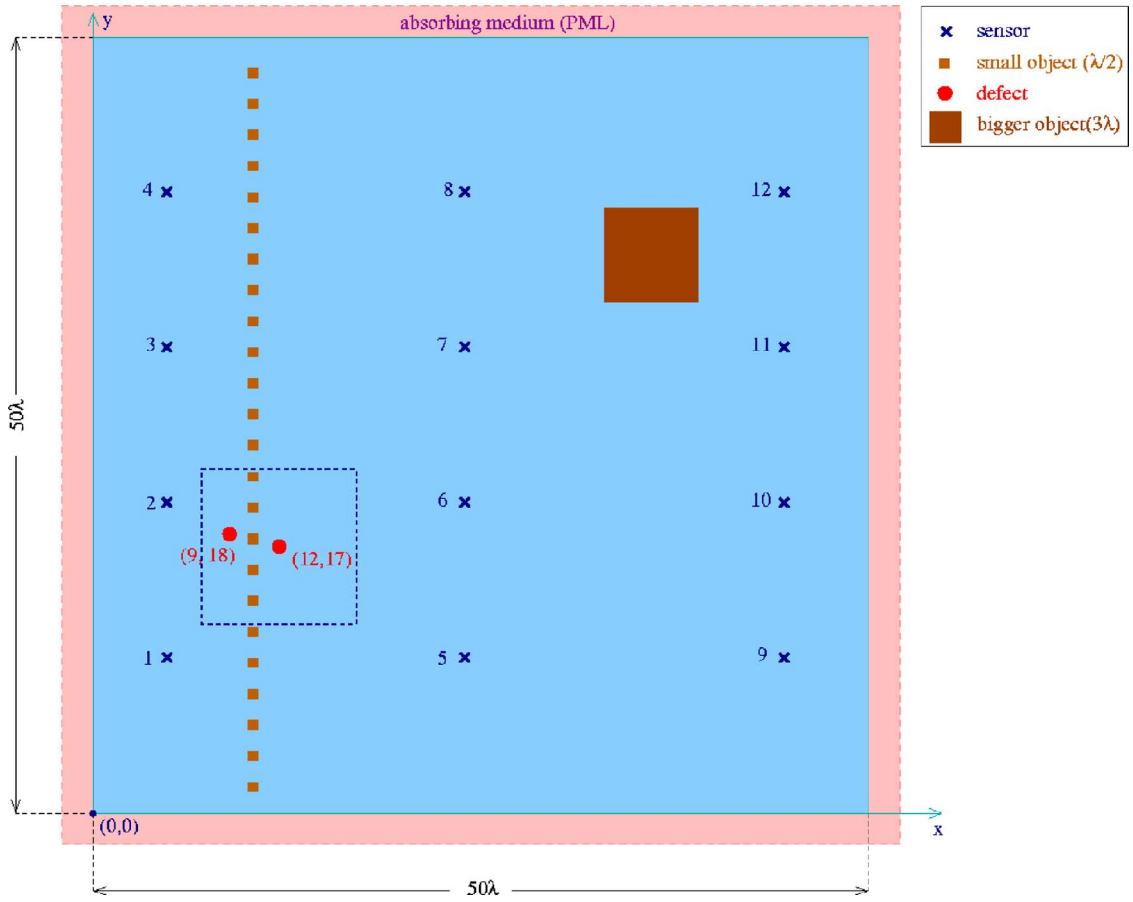


FIG. 1. (Color online) Description of the structure. It is a plate of size 50λ by 50λ on which are placed a fixed object of size 3λ and 25 smaller objects of size λ/2 λ′=1 cm is the central wavelength of the probing pulse used by the sensors. All dimensions are given in units of λ. The wave speed is taken to be c_0=5000 m s^{-1}. For numerical simulations the domain is surrounded by perfectly matched layers (PML) in order to simulate propagation in the free space. We want to image two point-like defects shown by the red dots. All images in this paper show the imaging function in a square of size 10λ centered on the defect located at (12, 17), with a grid resolution of λ/4. This imaging square is shown with dashed lines.

Ref. 8. Time reversal imaging with dispersive waves is an important issue that will be addressed in a later study. There are several idealizations in this approach to distributed sensor imaging that need to be pointed out. Regarding the imaging algorithm, the main one is the assumption that the wave propagation properties of the structure, its Green's function, are known. This is a reasonable one for nondestructive testing or structural health monitoring because a lot is known about the structure. Another one is the use we make of the difference of signals with and without defects, which is difficult to do in practice and requires high signal to noise ratios. Regarding the numerical simulations, they are done in two dimensions with the wave equation, without dispersion.

This paper is organized as follows. In the next section we present the numerical setup used in this study. In Sec. III we consider the travel time migration algorithm and show that it does not perform well with the data that we use. In Sec. IV, we present the time reversal algorithm with optimal stopping. The numerical simulations confirm the expected good performance of this algorithm as well as its reliability.

II. DISTRIBUTED SENSOR FRAMEWORK FOR NUMERICAL SIMULATIONS

In order to assess the effectiveness of distributed sensor imaging algorithms we have carried out extensive numerical simulations with the wave equation in two dimensions. We do not consider the dispersive effects of Lamb wave propagation. This model is thus valid in the low frequency range for a thin plate. It describes the propagation of the first symmetric mode *S*0, assuming that there is no conversion from *S*0 mode to *A*0 mode.

We consider the structure shown in Fig. 1. It is a domain of size 50λ by 50λ, where λ=1 cm is the central wavelength of the probing pulse used by the sensors. All dimensions in this paper are given in units of λ. The wave speed is taken to be c_0=5000 m s^{-1}, which is typically the speed of the lowest symmetric propagating Lamb mode in a 1-mm-thick aluminum plate. On this plate structure we place a fixed object of size 3λ and 25 smaller objects of size λ/2 which are on the same line. The latter could represent a line of rivets, for example. We simulate propagation in an infinite plate, as if we were considering a small part of a bigger structure. Reflections from boundaries provide more information at the sensors and make time reversal imaging more robust, as explained in Sec. IV. Imaging in an infinite region is therefore an important case to consider in some detail. There is no intrinsic absorption in the structure, so the only cause of dissipation is the outgoing radiation.

The defects we want to image are two identical point-like objects, located on the plate at (12,17) and (9,18). They

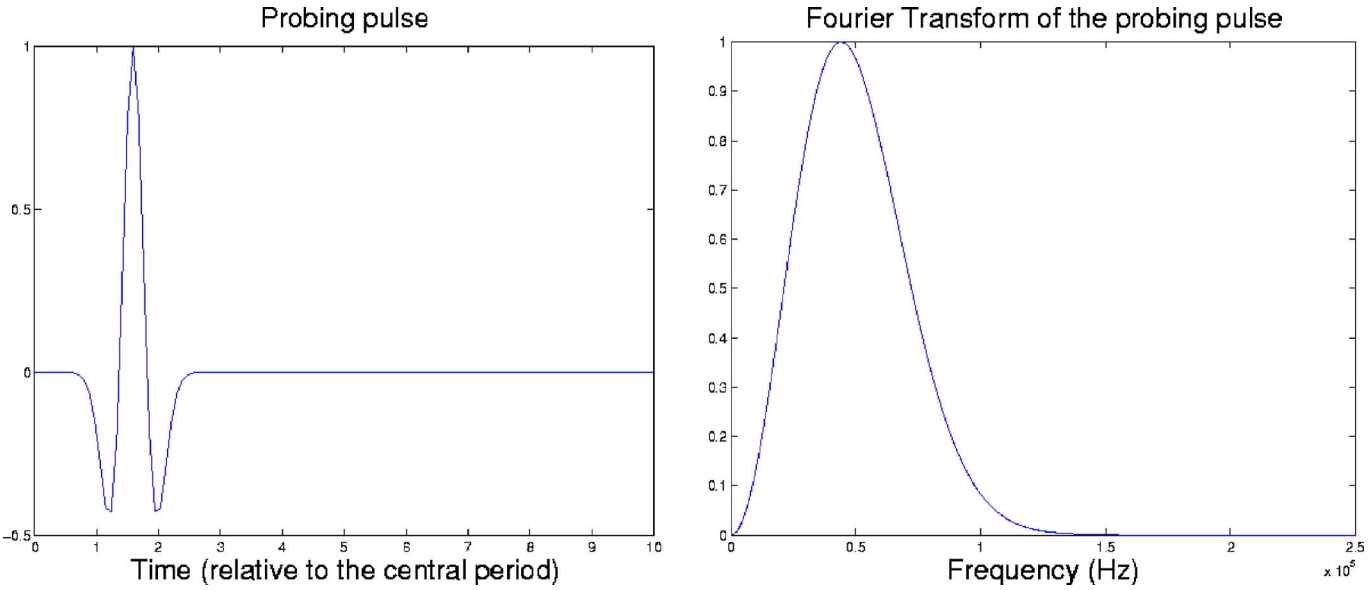


FIG. 2. (Color online) The probing pulse used by the sensors in the time domain (left) and its Fourier transform (right). It is a second derivative of a Gaussian whose central frequency is 500 kHz.

are thus approximately 3λ apart. All the features and the defects in our computations are perfect reflectors with Dirichlet boundary conditions. They could also be penetrable objects, that is, heterogeneities of the background with finite propagation speed. With point-like defects this does not affect the imaging algorithms much.

The structure is illuminated with a small number of sensors regularly distributed. There are $N=12$ of them that are placed in 4 rows of three sensors each. The rows of sensors are separated by 10λ and in each row the sensors are 20λ apart. The location of the sensors is denoted x_p, $1 \leqslant p \leqslant N$. They are point-like and isotropic, and capable of both emitting a pulse into the medium and recording the vibration at their location. The probing pulse that we use in the computations is the second derivative of a Gaussian given by

$$f(t) = (2[\alpha(t-t_0)]^2 - 1)\exp(-[\alpha(t-t_0)]^2), \tag{1}$$

where $\alpha = \pi\nu$, and t_0 is a translation of the time origin. It is shown in Fig. 2 along with its Fourier transform. Its central frequency is $\nu = 500$ kHz and with background velocity $c_0 = 5000$ m s^{-1} the central wavelength is $\lambda = 1$ cm. The frequency band at −6 dB is approximately [220 kHz, 850 kHz], which gives a 130% relative bandwidth.

The wave equation in two dimensions is solved with a numerical method based on the discretization of the mixed velocity-pressure formulation for acoustics. For the spatial discretization we use a finite element method which is compatible with mass lumping,[9,10] that is, which leads to a diagonal mass matrix so that explicit time discretization schemes can be used. For the time discretization we use an explicit second order centered finite difference scheme. In the simulations the point-like defects are modeled by small squares whose side is given by the space step of the grid, namely λ/32. The infinite medium is simulated by embedding the computational domain into a perfectly matched absorbing layer.[11]

For the given distribution of sensors, the response matrix of the healthy structure is computed in the time domain. Each sensor $p=1\ldots N$ emits a pulse into the structure and the echoes are measured at all sensors $q=1\ldots N$. This response matrix is denoted $P^0(t)=[P^0_{pq}(t)]_{p,q=1\ldots N}$. We call it the base line. It is symmetric because of reciprocity. Each column of P^0 corresponds to a different illumination of the structure: the pth column are the signals or traces measured at all sensors when sensor p is firing. The response matrix of the damaged structure is computed with the same configuration of sensors and is denoted by $P^d(t)$. The difference between the damaged and healthy response matrices is denoted by $P_{pq}(t)=P^d_{pq}(t)-P^0_{pq}(t)$, $p,q=1\ldots N$. Henceforth we call $P(t)$ the *response matrix* (of the damaged structure). In this difference matrix the direct arrivals of emitted pulses and reflections coming from the scatterers in the healthy structure have been removed. The matrix $P(t)$ contains therefore the backscattered echos coming from the defects and from multiple scattering between them and also with the scatterers in the healthy structure. Since the sensors are close to each other in array imaging, the direct arrivals are in the early part of the received signals. They can therefore be removed by cutting off that part of the signal, as the region to be imaged is at some distance from the array. However, in distributed sensor imaging the direct arrivals cannot be removed by a simple cutoff. This is one reason why knowledge of the response of the healthy structure is needed when imaging with distributed sensors.

The sixth column of each of the matrices $P^0(t)$, $P^d(t)$, and $P(t)$ is shown from left to right in Fig. 3. The signals are the vibrations recorded as functions of time at one of the 12 sensors, when sensor 6 is firing. All signals shown are normalized so that their maximum is 1. The direct arrival of the probing pulse is clearly visible on the traces for the healthy structure at sensors 4–12, because they have a direct line of sight with sensor 6. The smaller vibrations arriving at later

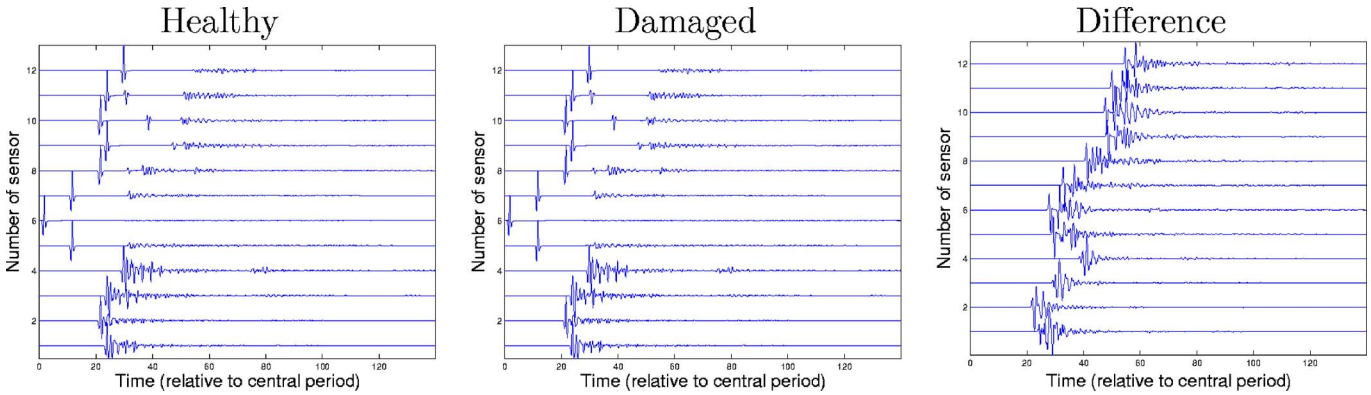


FIG. 3. (Color online) Traces recorded at all 12 sensors when sensor 6 is firing. From left to right: Traces in the healthy structure, in the damaged structure and difference between them. All signals shown here are normalized so that their maximum is 1. The amplitude of the difference traces on the right is approximately 100 times smaller than the amplitude of the healthy or damaged traces. *X* axis: time, *Y* axis: number of sensor.

times are from the multiple scattering with the reflectors in the healthy structure. The sensors 1–4 are located behind the line of rivets so there is no line of sight with sensor 6. Therefore, there is no clear direct arrival of the probing pulse but rather a long coda that comes from the multiple reflections with the rivets. This signal coda is called the delay spread. The differences between $P^0(t)$ and $P^d(t)$ can hardly be seen by looking at them separately because the reflections coming from the defects are very small compared to the direct arrivals. The amplitude of the signals coming from the defects [in $P(t)$] is approximately 100 times smaller than that of the direct arrivals [in $P^0(t)$ and $P^d(t)$]. The difference traces contain signals coming only from the defects and it is quite clear that the direct arrivals have been removed. However, there is also no clear arrival time coming from the defects. This is because the healthy structure around the defects has other scatterers that generate delay spread, which depends significantly on the illumination. We also note that with distributed sensors the trace peaks do not form hyperbolas as is in array imaging. It is therefore not possible to get a rough estimate of the location of the defects from a quick glance at the data, as is often the case in array imaging.

Signal to noise ratio is an issue that is not addressed in this paper. It is assumed to be very high. Therefore the presence of defects can be detected if at least one singular value of the Fourier transform of $P(t)$ is above some threshold. Our purpose is to go well beyond this step, to an algorithm that images the defects.

III. TRAVEL TIME IMAGING

Perhaps the simplest way to image with distributed sensors is by triangulation. The main difficulty in implementing triangulation is getting a reliable estimate of arrival times from the traces of the response matrix $P(t)$. This difficulty can arise from the dispersive nature of Lamb waves, as is discussed in Ref. 7, or from multiple scattering that generates large delay spread in the traces, as described in the previous section.

Travel time imaging, or travel time migration, or Kirchhoff migration is an important imaging algorithm that is based on travel time computations. It is different from basic triangulation because it does not require the estimation of arrival times from the traces. It is used extensively in seismic array imaging[3,12] and elsewhere. Several variants of it have also been used in structural health monitoring.[4–6] The main idea in travel time migration is to compute the value of an imaging functional of the data at each "search point" y^S in the region that we want to image. With the traces recorded at the N receivers at $(x_q)_{1\leqslant q\leqslant N}$ when the sensor at x_p is firing, we compute for each y^S

$$I_p^{KM}(y^S) = \sum_{q=1}^{N} P_{pq}[\tau_p(y^S) + \tau_q(y^S)], \tag{2}$$

where $\tau_p(y^S) = |x_p - y^S| / c_0$ is the travel time from x_p to y^S and c_0 is the background propagation speed. That this is an imaging algorithm can be seen as follows. The travel time $\tau_p(y^S) + \tau_q(y^S)$ is the time for the wave to go from the source at x_p to the search point y^S and then from y^S to the receiver at x_q. If y^S is near a defect location then the trace $P_{pq}(t)$ will have a peak at that time. The imaging functional $I_p^{KM}(y^S)$ coherently sums these peak values for the different sources and receivers, producing a local maximum or minimum. However, if y^S is far from the location of a defect, then the traces will be added incoherently and $|I_p^{KM}(y^S)|$ will be small.

The imaging functional $I_p^{KM}(y^S)$ is an approximation to the least squares solution of the linearized inverse scattering problem,[3,12] and its resolution gets better as the number of sensors increases. The resolution theory for travel time migration is well understood for sensor arrays in a homogeneous or smooth background.[3,13,14] It is well known, for example, that the range resolution, that is, the resolution in the direction orthogonal to the array, is controlled by the bandwidth B of the probing pulse and is given by c_0/B. The cross range resolution, that is, the resolution in the direction parallel to the array, is limited by the aperture of the array a and the distance L between the array and the defect, and it is given by $\lambda L/a$, where λ is the central wavelength. There does not appear to be a resolution theory for travel time migration with distributed sensors while range and cross range are no longer relevant terms. It is expected that the resolution is limited by the bandwidth of the probing pulse and by the uniformity of distribution of the sensors around the defect. Imaging with travel time migration improves, in general, if a large number of sensors is used.

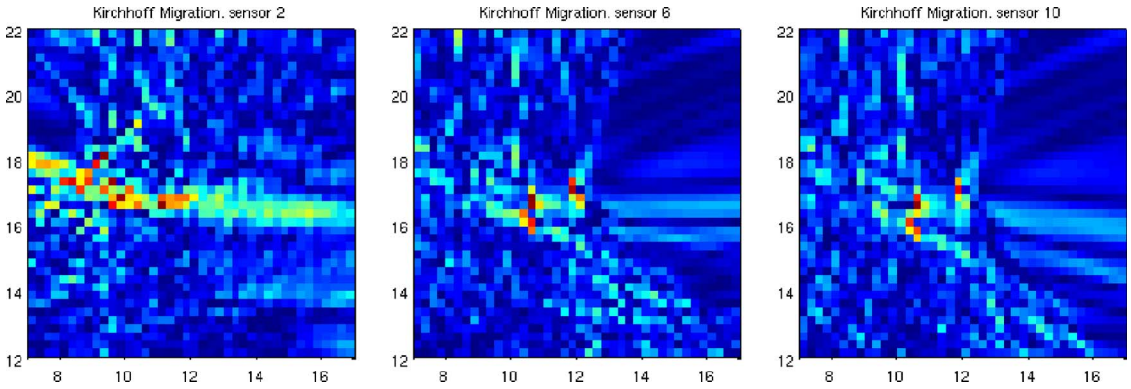


FIG. 4. (Color online) Travel time or Kirchhoff migration images in a square domain of size 10λ centered at (12, 17). Each figure corresponds to a different illumination of the structure. From left to right, illumination with sensors 2, 6, and 10. The image functions are normalized to be between zero and one, and the color scale in the plots is linear.

The three images obtained with the travel time migration algorithm in our numerical simulations are shown in Fig. 4, when the structure is illuminated with the second row of sensors, that is, when sensors 2, 6, and 10 are probing. The migration images obtained with the other rows are very similar.

All images in the figures in this paper show the imaging function in a square of size 10λ centered on the defect located at (12, 17), with a grid resolution of λ/4. This imaging square is shown with dashed lines in Fig. 1. The second defect is in the upper left part of this image domain. The image functions are normalized to be between zero and one, and the color scale in the plots is linear.

Because the sensor data in our simulations have a lot of delay spread, travel time migration does not work as well. The three images in Fig. 4 differ from each other and depend sensitively on which sensor is illuminating. It is not possible to rely on any particular illumination more than another. However, this result can be improved by accumulating all the images as is shown in Sec. IV D.

The difference traces have a lot of delay spread because of multiple scattering between the defects and the reflectors that are in the background. In travel time migration the background is assumed to be homogeneous and multiple reflections between the defects are neglected, which is the Born approximation. We will next introduce an imaging method that uses knowledge of the background. It gives more reliable results that are stable when different sensors illuminate.

IV. TIME REVERSAL WITH OPTIMAL STOPPING

A. Physical time reversal

In physical time reversal sensor arrays focus energy on sources with resolution that improves when there is multiple scattering.[15–19] The signal emitted by a source is received by the sensor array, it is time reversed and then reemitted into the medium. The waves propagate back toward the source and focus around it. The refocusing location is not known in this process but the time of refocusing is known if we know at what time the source started to emit. Refocusing occurs both in space and time.[15] The spatial resolution of the focusing is better when there is a lot of multiple scattering[19,20] because the complex medium effectively enhances the size of the sensor array, and the quality and stability of the refocusing improves when the bandwidth of the pulse emitted by the source is large. Physical time reversal provides therefore an efficient way to focus energy on a defect[16] or for communications.[17,21]

B. Numerical time reversal for imaging

Time reversal can also be used for imaging sources. In this case, the traces recorded at the sensors are back propagated *numerically* in an idealized medium, since the actual medium is not known in detail, in general. An image of the location of the sources is obtained by taking a snapshot of the back propagated field at the refocusing time. This procedure can be applied both with active sources and with passive scatterers. Indeed, back propagation of the recorded traces with travel times is the migration algorithm of the previous section.

For a passive defect, there are in principle several refocusing times, due to the multiple scattering between the defect and the structure. We neglect here those higher reflections that are weaker and consider thus only the strongest refocusing event.

In structural health monitoring it is reasonable to assume that we have some knowledge of the healthy structure, up to some level of detail. We will assume here that the background is known, meaning that the traces can be back propagated in the healthy structure shown in Fig. 1.

For each illumination of the structure the difference traces recorded at each sensor are time reversed and back propagated numerically in the medium. Let $u_p(y,t)$ denote the field at time t and point y that is back propagated when the traces of the pth column of the response matrix P are used. Then $u_p(y,t)$ is the solution of the partial differential equation

$$\begin{cases} \dfrac{\partial^2 u_p}{\partial t^2}(y,t) - c_0^2 \Delta u_p(y,t) = \displaystyle\sum_{q=1}^{N} \delta_{(y=x_q)} P_{pq}(T-t) \quad \text{in } \mathbb{R}^2 \setminus \Omega, \\ u_p(y,t) = 0, \quad \text{on } \delta\Omega, \end{cases} \tag{3}$$

where Ω is the set of all reflectors in the healthy structure, as shown in Fig. 1, and $\delta\Omega$ denotes their boundary. Here $\delta(y$

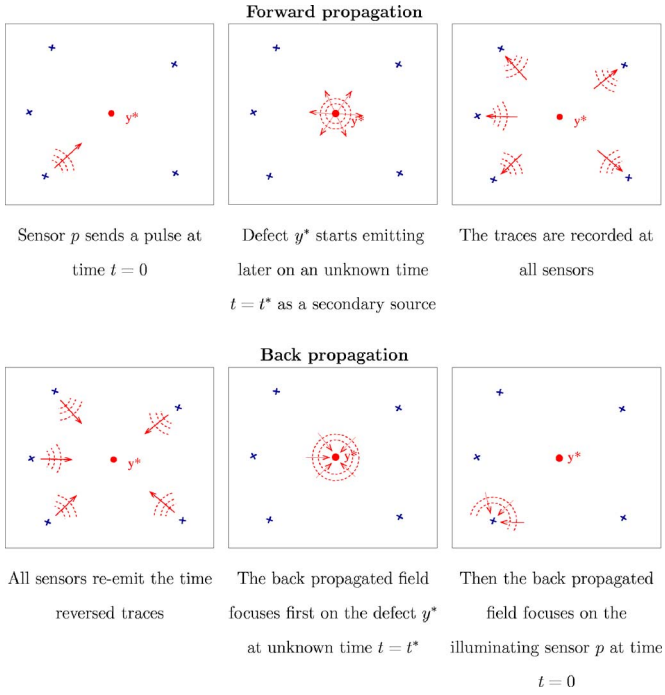


FIG. 5. (Color online) Schematic of echo-mode time reversal explaining why the refocusing time is not known.

$=x_q$) is the Dirac function at x_q and $(0,T)$ is the recording time interval. This equation is solved numerically with the finite element method discussed in Sec. II.

We want to obtain an image of the defects by taking a snapshot of the back propagated field $u_p(y,t)$ at the time it refocuses on them. The problem with this approach is that the refocusing time is *not known*, as explained schematically in Fig. 5. This is a major difference between active source imaging by time reversal and the echo mode imaging. In echo mode the wave emitted by the probing sensor must first reach the defects before they can act as secondary sources. Since the location of the defects is not known, the time t^* at which they start emitting is not known. The back propagated field will first focus on some defect at time t^*. If we continue back propagating it will focus on the emitting sensor, which is the actual source, at time 0, but we are obviously not interested in this. Therefore, we must consider ways to determine the refocusing time t^*.

We want to distinguish between back propagated fields that are spread out from those that are more focused. A simple way to do this is to pick the time at which the amplitude of the field is maximal because at that time the signals coming from all the sensors are superposed constructively. This does not work because of the decrease in amplitude with distance from the emitting source. It is not possible to compensate for this when the sensors are distributed because the defect might be anywhere, near or far from any sensor. The situation here is different from that encountered with arrays. If the defect is far enough from the array then the sensor-to-defect distance is approximately the same for all sensors and an amplitude correction could be considered.

For distributed sensors that are more or less uniformly distributed around the defects, the back propagated field is coming toward them from every direction. It will focus locally in time and it is spread around the defects both before and after the refocusing time. A way to characterize focused images is to measure them with norms that are small in that case and large otherwise. Norms that penalize images with a lot of fluctuations, a lot of speckles or many geometrical features, are called *sparsity norms* or sparsity measures. They are widely used in image processing in particular in denoising or in data compression.[22–24] They work well with distributed sensors, as can be seen in Fig. 6, because the information comes from every direction and the norms are lower at the refocusing time. They do not work so well with back propagation from an array because in that case the field is coming mostly from one direction only.

We consider here two sparsity norms:

- The Shannon entropy, $S[u_p(.,t)]$, which is a measure of the information needed to encode a pixelized image.
- The bounded variation norm, $BV[u_p(.,t)]$, which is an L^1 sparsity norm that tends to penalize images that have a lot fluctuations.

The entropy is a norm which characterizes the sparsity globally, as it treats all points independently, while the BV norm is rather local, as it measures local variations. In our images, those two norms give about the same result because we consider point-like objects for which the difference between local and global behavior plays no role.

Let $u_{ij}(t), i,j=1\ldots N_d$ denote the space-discretized version of the continuous field $u_p(y,t)$ at time t on a square grid with spatial step h containing N_d points in each direction. We denote by $y_{ij}=y_0+ih\boldsymbol{e_x}+jh\boldsymbol{e_y}$, $i,j=1\ldots N$, the discretization points, where y_0 denotes the lower left corner of this square grid and $(\boldsymbol{e_x},\boldsymbol{e_y})$ are the coordinate vectors. We define $u_{ij}(t)=u_p(y_{ij},t)$, $i,j=1\ldots N_d$. In the images shown in this paper, the square grid contains 41 points in each direction and the space step is $\lambda/4$. We use a square grid for simplicity.

1. Shannon entropy

Shannon's definition[25] of the entropy of a pixelized image u_{ij} is a measure of the sparsity of the histogram of the gray levels of the image. For a given number of gray levels N_c we introduce a linear gray level scale (c_k), $k=0\ldots N_c$, ranging from the minimum of u_{ij} to its maximum. The histogram of gray levels of the image is defined by counting the number of pixels contained in each gray level set

$$h_k=\sum_{i,j}\mathbf{1}_{\{|c_k,c_{k+1}|\}}(u_{ij}),\quad k=0\ldots N_c-1. \tag{4}$$

Here $\mathbf{1}_A$ is the characteristic function of a set A. Clearly, $\Sigma_{k=0}^{N_c}h_k=N_d^2$ and thus (h_k/N_d^2) is the probability distribution of gray levels of the image for a given number N_c. The Shannon entropy of the image is the Boltzmann entropy of that probability distribution, defined by

$$S(u_{ij})=-\sum_{k=0}^{N_c-1}\left(\frac{h_k}{N_d^2}\right)\log_2\left(\frac{h_k}{N_d^2}\right). \tag{5}$$

In this paper the number of gray levels is $N_c=256$. The results are not sensitive to N_c unless it is very small, such as $N_c=2$.

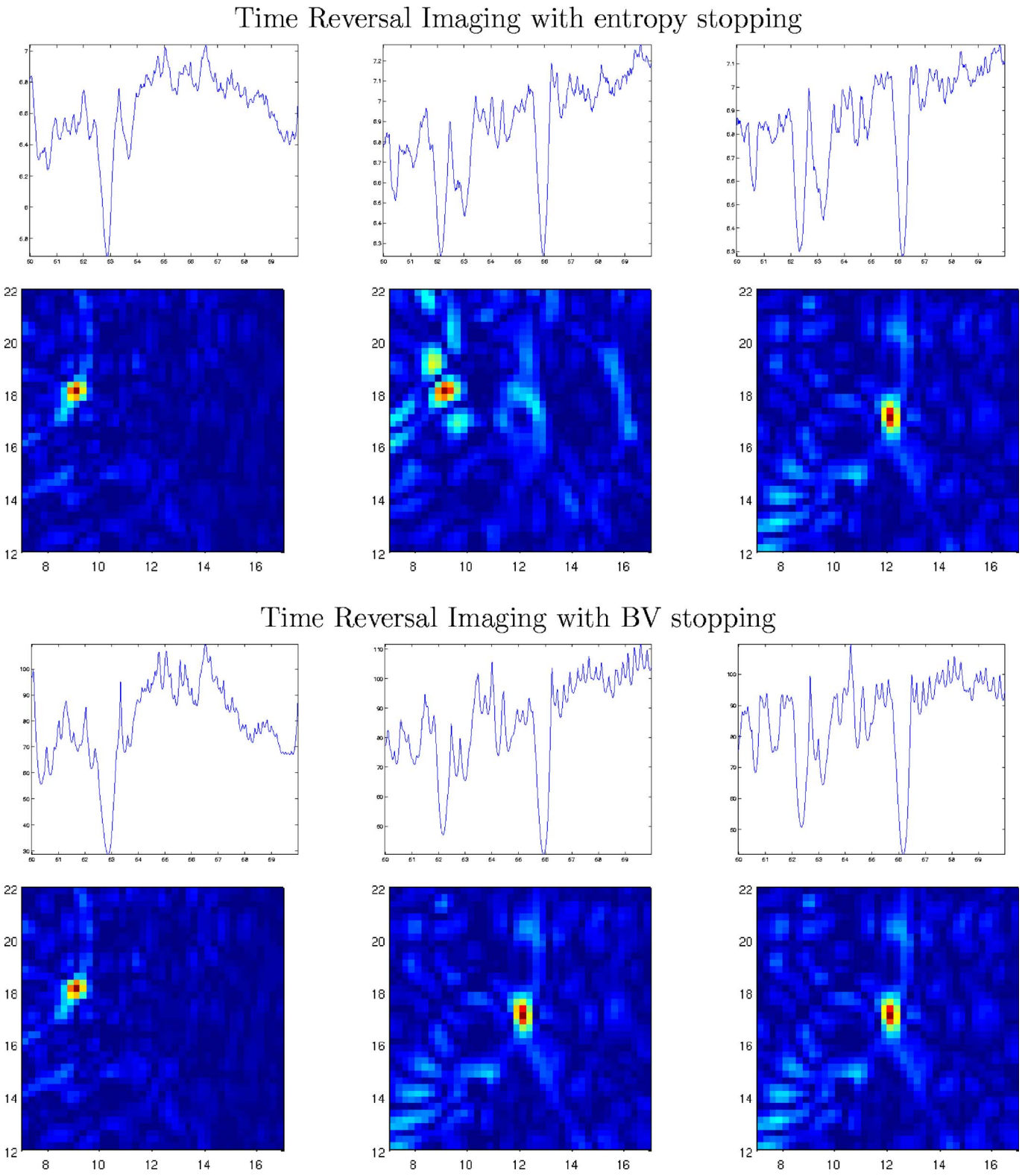


FIG. 6. (Color online) Top two rows: Entropy vs time of the back propagated field in a square domain of size 10λ centered at (12, 17) and snapshot of the back propagated field at the time where entropy is minimum. Each figure corresponds to a different illumination of the structure. From left to right, illumination with sensors 2, 6, and 10. Bottom two rows: Same results with *BV* norm.

The entropy quantifies the amount of information needed to encode an image and is often given in bits per pixel. It is used in image compression[26] and for other applications in image analysis.[27,28] It penalizes back propagated fields that have a lot of speckles. The definition of entropy in Eq. (5) treats all points in the image independently. Therefore, the entropy of an image whose pixels have been shuffled around has exactly the same entropy as the original.

2. BV norm

The bounded variation norm[23] of a regular function $u(x)$ defined in a domain Ω is given by

$$BV(u) = \int_{\Omega} (|u(x)| + |\nabla u(x)|)dx. \quad (6)$$

For a pixelized image u_{ij} defined on a grid with spatial step h the BV norm is given by

$$BV(u_{ij}) = h^2 \sum_{i,j} (|\tilde{u}_{ij}| + |\nabla_h \tilde{u}_{ij}|), \quad (7)$$

where $|\nabla_h \tilde{u}_{ij}|$ is a finite difference approximation of the gradient of $\tilde{u}_{ij}$. We let $\tilde{u}_{ij} = u_{ij} / \max_{ij}(|u_{ij}|)$ be the normalized version of the image u_{ij}. As already noted above, the amplitude of the field at the time of refocusing depends on the distance between the defect and the sensors. So it is necessary to normalize the image before taking its BV norm so as to avoid dependence on field amplitudes. Note that this normalization is intrinsically made with entropy since $S(u_{ij}) = S(\alpha u_{ij})$ for all $\alpha > 0$. The BV norm penalizes images that have a lot of oscillations, because it has the gradient in it. It also penalizes images that are spread out diffusely, and the L^1 norm plays an important role in this. The BV norm is used widely in image denoising because it preserves sharp features.[29,30]

3. Time reversal imaging with optimal stopping

The imaging algorithm we use is the following:

1. For the pth column of the response matrix P, compute numerically the wave field $u_p(y,t)$ defined by Eq. (3).
2. Compute the sparsity norm (Shannon entropy or BV norm) of the field $u_{ij}(t) = u_p(y_{ij},t)$ in the imaging domain (y_{ij}) as a function of time.
3. Pick the time at which it is minimal, denote it t^*.
4. Plot the image $u_{ij}(t^*)$.

4. Results of numerical simulations

The entropy and BV norm versus time are plotted in Fig. 6 for the three illuminations from the second row of sensors (sensors 2, 6, and 10). The back propagated wave fields are also shown at the optimal time. These three illuminations are typical and the results obtained with the other rows of sensors are similar. The time plots are shown in a time window that is zoomed, near the focusing events. The results are good because they give rather clean images of the defects. There are few speckles, the focusing spot is smooth, the defects are at the right locations, and there are no ghosts. The two norms, Shannon entropy and BV give comparable results. They both pick a stopping time close to the refocusing time on one of the two defects. The optimal stopping times picked by the two norms differ by at most one or two time steps.

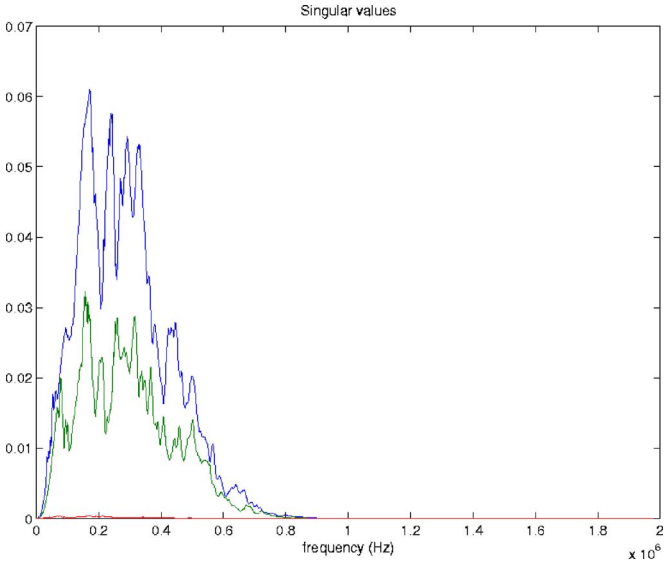


FIG. 7. (Color online) The first three singular values of the response matrix $\hat{P}(\omega)$ versus frequency. There are clearly two distinct leading singular values at each frequency in the frequency band, which correspond to the two point-like defects.

However, since only one refocusing time is picked with this technique, one cannot get an image for each defect at the same time. This method images the strongest defect as it is perceived by the sensors for a given illumination of the structure. For example, the method picks the defect located at (9, 18) when sensor 2 is illuminating, and there is only one minimum, which means that only one defect is detectable. This may be explained roughly by noting that the defect (9, 18) and sensor 2 are located on the same side of the line of rivets. On the other hand, the strength of the defects is roughly the same when they are illuminated with sensors 6 or 10. There are two clear minima that have almost the same value, both with the entropy and the BV norm. The minima are focusing times on each of the two defects. Therefore, it is not possible to select one minimum rather than the other. This is an illumination issue that is best dealt with the singular value decomposition (SVD) that allows for selective imaging of each defect. It is discussed in the next section.

C. Separation of the defects by singular value decomposition

1. SVD of the response matrix in the frequency domain

The relation between the singular vectors of the response matrix $P(t)$ and the scatterers has been analyzed extensively.[31–33] Each localized defect can be associated with a singular vector of the response matrix, except for a few degenerate cases. It is called the DORT method, which is the French acronym for "Decomposition of the Time Reversal Operator." The SVD is a way of finding the optimal illumination of a defect,[34,35] that is, the one that generates the strongest received signals at the sensors. Selective time reversal focusing using the SVD of the response matrix has been successfully used theoretically and experimentally,[31–33] as well as for imaging in random media.[36,37]

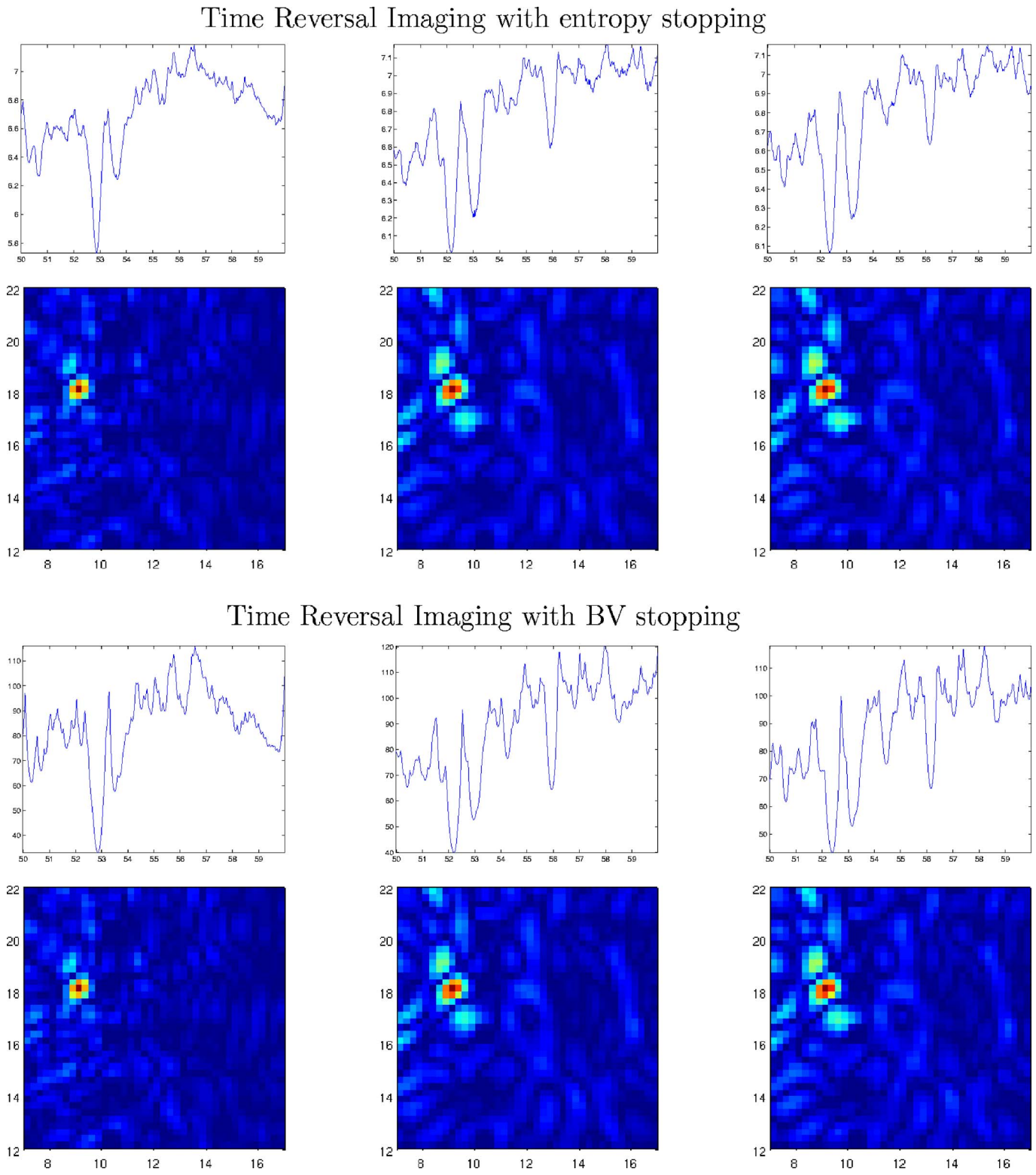


FIG. 8. (Color online) Same as Fig. 6 but with back propagating traces projected on the first singular vector.

One rather direct application of the SVD of the response matrix is estimating the number of localized defects.[15] The number of nonzero, leading singular values is an estimate of the number of localized defects. It is denoted by M. This is seen very well in our numerical simulations. The first three singular values versus frequency are shown in Fig. 7. There are clearly two distinct leading eigenvalues at each frequency inside the bandwidth of the probing pulse, which correspond to the two point-like defects. These two singular values are very well separated over the frequency band. This could not have been anticipated since the defects are identical. The curves of singular values versus frequency will, in general, cross each other.

Localized defects are said to be well resolved (or well separated) if the illuminating vectors associated with them are orthogonal. The elements of the illuminating vectors are

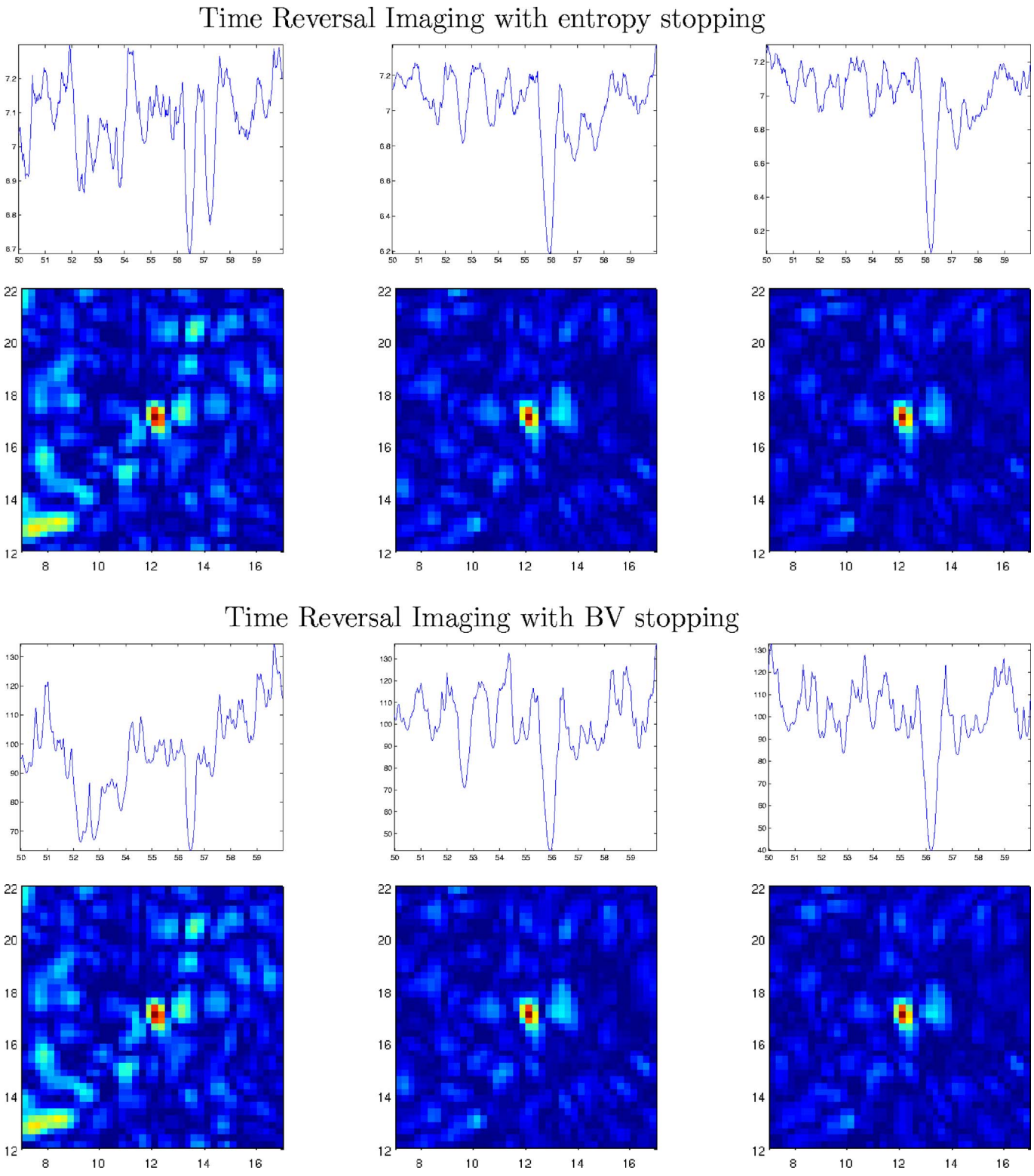


FIG. 9. (Color online) Same as Fig. 6, but with back propagating traces projected on the second singular vector.

the Green's functions from the sensors to the defects.[37] These vectors are also right singular vectors in this case. Of course, a larger number of sensors helps in resolving defects so illuminating vectors are more likely to be orthogonal in that case. We consider imaging with time reversal and the SVD in the next section.

2. Imaging the defects using the traces projected on each singular vector

We can say that, *in principle*, the singular value decomposition transforms an echo mode problem into an active source problem. This is because at least for well separated defects the singular vectors are also illuminating vectors to

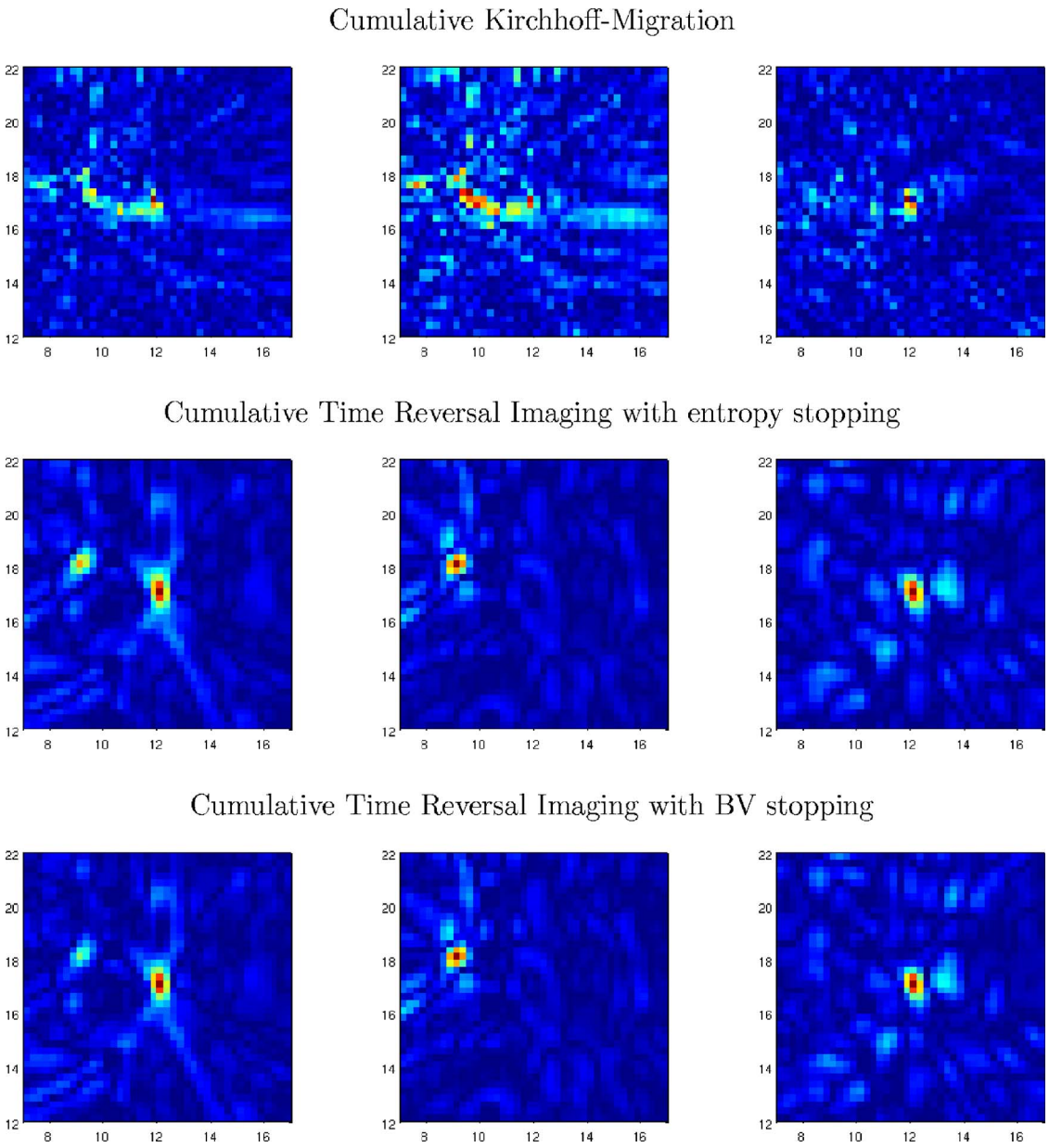


FIG. 10. (Color online) Cumulative images obtained by summation over each illumination as in Eq. (9). Left: with full traces, middle: With traces projected on the first singular vector, right: With traces projected on the second singular vector.

the unknown defect locations. However, because they carry an arbitrary, frequency-dependent phase, the singular vectors look *incoherent* in the time domain. In order to get rid of this arbitrary phase, we first project the response matrix on the space spanned by each singular vector $\hat{U}_k^H(\omega), k=1\ldots M$. We now image the defects with the response matrix $P_k(t), k=1\ldots M$, obtained by projection of the full response matrix onto each leading singular vector.[37] The pth column of the Fourier transform of $P_k(t)$ is given by

$$\hat{P}_k^{(p)}(\omega) = [\hat{U}_k^H(\omega)\hat{P}^{(p)}(\omega)]\hat{U}_k(\omega), \quad p = 1, \ldots, N. \tag{8}$$

Here $\hat{P}^{(p)}(\omega)$ denotes the pth column of Fourier transform of the full response matrix $\hat{P}(\omega)$. Because of the orthogonality of the singular vectors, this projection removes the reflections coming from the other defects. We can think of $P_k(t)$ as the response matrix of the distributed sensors when only the kth localized defect is present.

Since the phases of the projected response matrices are preserved, any algorithm that can be used for processing the original response matrix $P(t)$ can also be used with the projected matrices, without any change. For example, this can be done with travel time migration or with time reversal imaging and optimal stopping. By analogy with the columns of $P(t)$, we refer to the columns of $P_k(t)$ as responses from illumination by the sensor labeled with the column index.

Time reversal images obtained by optimal stopping using the traces projected on the first and second singular vectors are shown in Figs. 8 and 9, respectively. As before, only results obtained with the illuminations corresponding to the

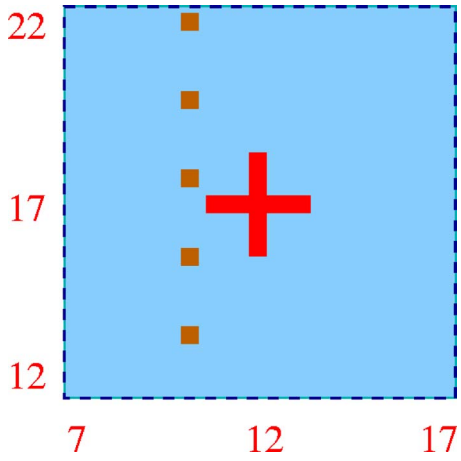


FIG. 11. (Color online) Schematic of the extended defect. Zoom in the square of size 10λ centered at the point (12, 17). The defect is a cross whose four sides are of size λ times λ/2, so that its overall size is 2.5λ.

second row of sensors (2, 6, and 10) are shown, for they are typical. The *BV* norm and entropy versus time are plotted above each image.

First, it is clear that all the images obtained with the first singular vector focus on the defect at (9, 18) and all those obtained with the second singular vector focus on the other defect. This illustrates well the stability of this algorithm. For the images obtained with data projected on the first singular vector there is now only one minimum for all 12 illuminations. So the ambiguity that was noted in the previous section has disappeared. This remark holds also for images obtained with the data projected on the second singular vector. Moreover, in this case the illumination of the defect at (12, 17) from sensor 2 is now possible. This is an illustration of the ability of the singular value decomposition to provide an optimal illumination that will focus selectively on one particular localized defect.

D. Cumulative images

All the algorithms presented above give an image for each illumination of the medium. The images can be enhanced by averaging over all illuminations. More precisely, if $I_p(y^S)$ denotes the image with the pth column of the response matrix with either travel time migration, time reversal imaging with entropy stopping or time reversal imaging with *BV* stopping, we form the following cumulative image:

$$I(y^S) = \frac{1}{N}\sum_{p=1}^{N} \frac{I_p(y^S)}{\max_{y^S}|I_p(y^S)|}. \tag{9}$$

We have normalized the traces before averaging in Eq. (9) so as not to mask the information provided by the remote sensors with the strong image provided by the near ones. A weighted average with an optimal selection of weights could be used here as well and is currently being investigated.

We have computed cumulative images obtained with the traces of the original response matrix as well as with the traces projected on the first and second singular vectors. Results are shown in Fig. 10. Note that the averages shown are computed with all illuminations and not only the three illuminations that were shown previously. As expected, these images have fewer speckles. Even with averaging, travel time migration does not work as well. It provides the right focusing spots but there are still strong speckles. Note also that projection on the singular vectors does not improve the performance of this algorithm.

The time reversal algorithm takes advantage of the multiple scattering and of knowledge of the Green's function of the healthy structure so it works well with both entropy stopping and *BV* stopping. When used with the original traces the defects are imaged as if they have different strengths, depending on how they are perceived by the distributed sensors. The projection onto the singular vectors of the response matrix improves the resolution of the defects.

V. IMAGING OF AN EXTENDED DEFECT

a. Formulation of the problem. We consider now the imaging of a spatially extended defect, rather than two point-like defects. The defect has the shape of a cross as shown in Fig. 11. Each of the four sides of the cross is of size λ times λ/2, so that its overall size is 2.5λ. Its shape is not convex, which makes it more difficult to image. Its size has been chosen to be larger that the resolution limit so that it may be possible to image its different features. As in the previous computations, the defect is modeled as a perfect reflector using Dirichlet boundary conditions. The healthy structure is the one shown in Fig. 1. Both the response matrix of the

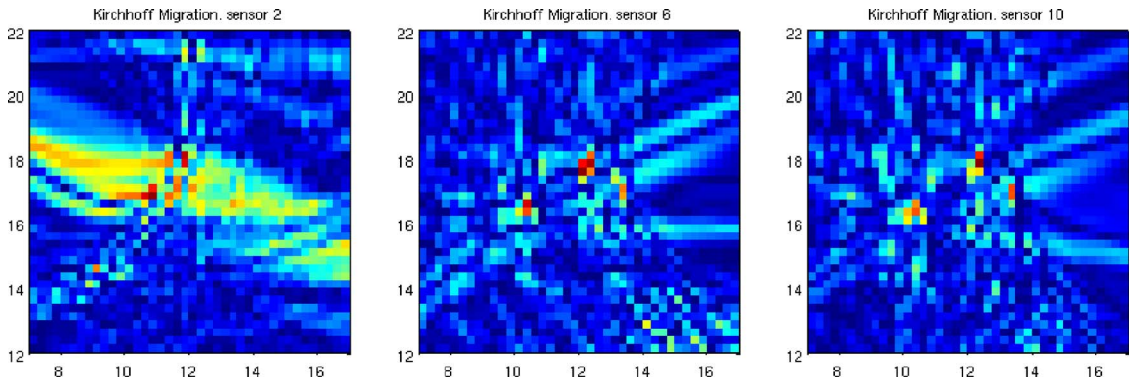


FIG. 12. (Color online) Travel time or Kirchhoff Migration in a square domain of size 10λ centered at (12, 17) when the damage is the cross depicted in Fig. 11. Each figure corresponds to a different illumination of the structure. From left to right, illumination with sensors 2, 6, and 10.

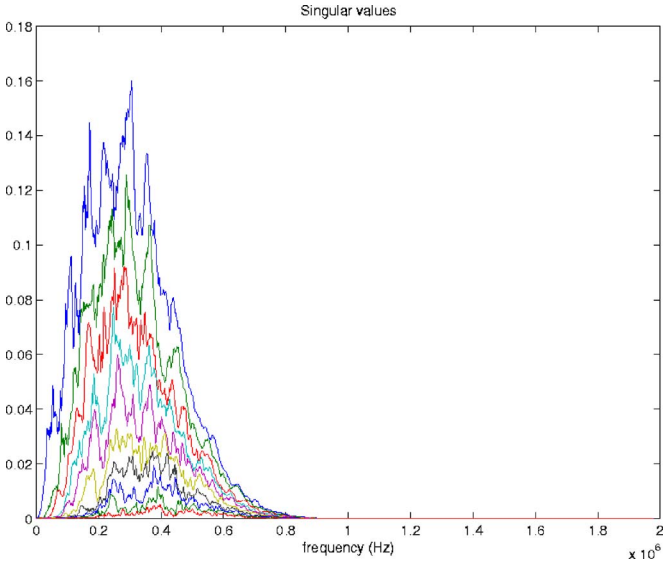


FIG. 13. (Color online) The 12 singular values of the response matrix $\hat{P}(\omega)$ vs frequency when the defect is the cross depicted in Fig. 11. The number of singular values of the response matrix is not related in a simple manner to the defect in the structure.

healthy and of the damaged structure are computed. We want to image the shape of the defect using the difference traces.

b. Travel time migration. The results obtained using travel time migration are shown in Fig. 12 for three different illuminations of the medium (the second row of sensors 2, 6, and 10). As in the case of two point-like defects, the results are very unstable with respect to the illumination. This is due to the multiple scattering between the defects and the scatterers that are present in the healthy structure, which is not taken into account in travel time migration.

c. Singular value decomposition. In the case of an extended defect the number of leading singular values is not related to it in a simple manner.[38–40] The 12 singular values of the response matrix are plotted as functions of frequency in Fig. 13.

d. Time-reversal algorithm. The images obtained with the time reversal algorithm described in Sec. IV for three different illuminations of the structure (2, 6, and 10) are shown in Fig. 14. For simplicity, only the results obtained using the *BV* norm are shown. The images we get using the entropy stopping are similar. Even for an extended defect, this algorithm gives an image of one part of the object, the one that has the strongest reflection, depending on the illumination. However, it gives an image of the back propagated field at only one time. Therefore, we cannot expect to get an image of the defect with only one illumination because the back propagated field does not surround it at one particular time. If the illumination is coming from the left of the defect, then the image tends to show the left tip of the cross, as can be seen in the left image in Fig. 14. If the illumination is coming from the right, then the right tip of the cross can be seen in the right image in Fig. 14.

A rough overall image of the damage can be obtained by summing over the images with different illumination, as in Eq. (9). The cumulative images obtained for the three algorithms (travel time migration, time reversal with entropy stop, and time reversal with *BV* stop) are shown at the top of

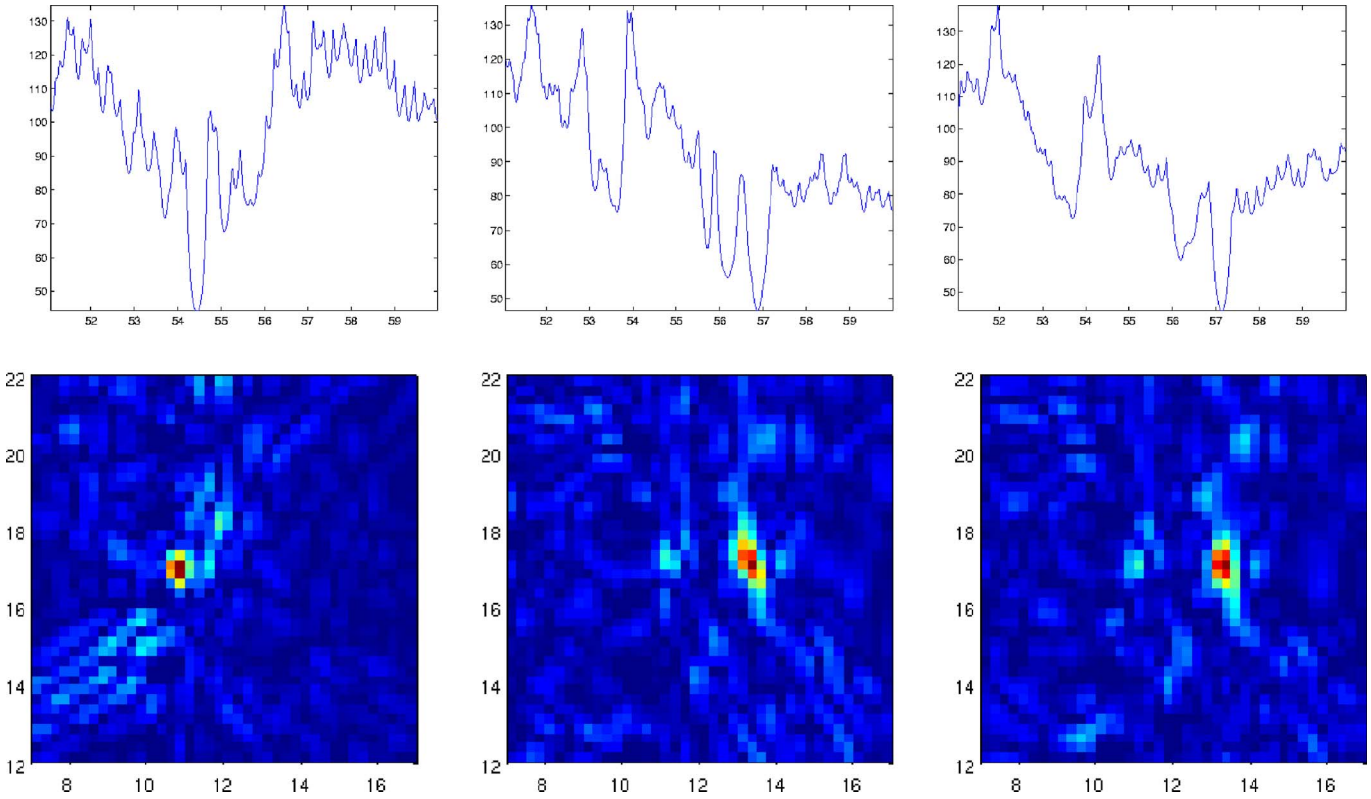


FIG. 14. (Color online) *BV* norm vs time of the back propagated field in a square domain of size 10λ centered at (12, 17) and snapshot of the back propagated field at time where *BV* norm is minimum. The damage is the cross presented on Fig. 11. Each figure corresponds to a different illumination of the structure. From left to right, illumination with sensors 2, 6, and 10.

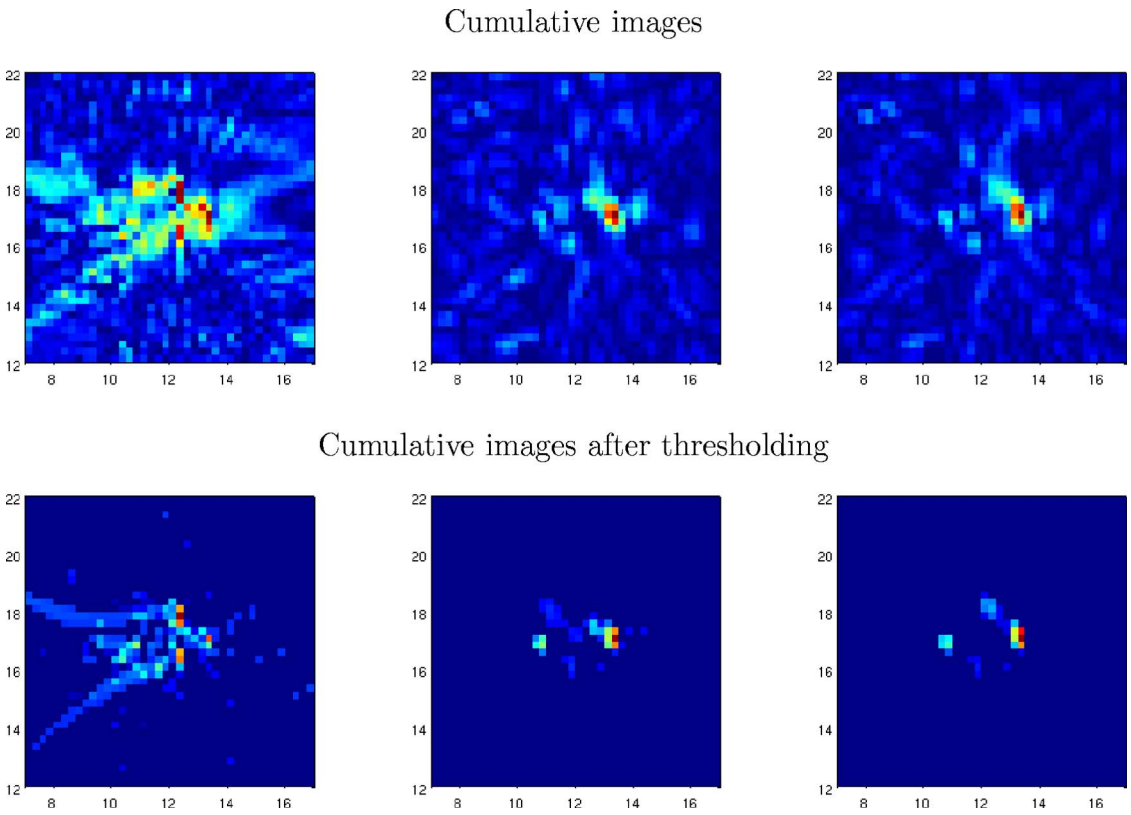


FIG. 15. (Color online) Imaging of the cross shaped defect depicted in Fig. 11. Left: Kirchhoff-Migration, middle: Time-reversal imaging with entropy stopping, right: Time-reversal imaging with *BV* stopping, top: Cumulative images obtained by summation over each illumination as in Eq. (9). Botttom: Cumulative images after thresholding as discussed in Eq. (10).

Fig. 15. The image obtained by simple summation over each illumination shows only the strongest edge of the cross. The other parts of the cross do not appear because they cancel out along with the speckles. We can get around this problem if we first threshold the image obtained for each illumination and then sum over the illuminations. More precisely we compute

$$\tilde{I}_p(y^S) = \begin{cases} I_p(y^S), & \text{if } |I_p(y^S)| \geqslant \alpha \max_{y^S} |I_p(y^S)| \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

where $\alpha \in [0,1]$ is a thresholding parameter. We then form

$$\tilde{I}(y^S) = \frac{1}{N} \sum_{p=1}^{N} \tilde{I}_p(y^S).$$

The images obtained by thresholding (with $\alpha = 0.6$) and summation are shown at the bottom of Fig. 15. The image obtained with travel time migration is definitely better than the images obtained for single illuminations. One can even guess the shape of the right side of the cross. However, there are still speckles and the size of the defect is overestimated. The images obtained with the time reversal algorithm are more stable and have fewer speckles. Even if they do not provide a clear contour of the defect, some important features can be seen and its size is rather well estimated.

VI. SUMMARY AND CONCLUSIONS

Imaging with distributed sensors is different from imaging with arrays mainly because we need to know the response matrix of the healthy structure in order to remove direct arrivals and other strong scattering from the background. There are many important issues that need to be addressed in order to deal effectively with noise in the data and with small scale inhomogeneities in the structure, which are not considered here.

We have presented here a detailed numerical study of several algorithms for distributed sensor imaging in the context of structural health monitoring. When the propagation characteristics of the healthy structure are known, as we assume, then time reversal imaging with optimal stopping, introduced here, gives good images of localized defects. When we also use the singular value decomposition, then the time reversal images improve. Time reversal imaging with distributed sensors also gives rough but stable images for extended defects.

ACKNOWLEDGMENTS

The work of G. Derveaux, G. Papanicolaou, and C. Tsogka was partially supported by the Office of Naval Re-

search N00014-02-1-0088, by the National Science Foundation DMS-0354674-001 and CMS-0451213, and by DARPA/ARO 02-SC-ARO-1067-MOD 1.

[1]F. Gustafsson, *Adaptive Filtering and Change Detection* (Wiley, New York, 2000).
[2]M. Basseville, "Lessons learned from the theory and practice of change detection," In *Proceedings of the 5th International Workshop on Structural Health Monitoring*, Stanford, CA, 929–936 (2005).
[3]N. Bleistein, J. K. Cohen, and J. W. Stockwell Jr., *Mathematics of Multidimensional Seismic Imaging, Migration and Inversion* (Springer, New York, 2001).
[4]C. H. Wang, J. T. Rose, and F.-K. Chang, "A synthetic time-reversal imaging method for structural health monitoring," Smart Mater. Struct. **13**, 415–423 (2004).
[5]L. Wang and F. G. Yuan, "Damage identification in a composite plate using prestack reverse-time migration technique," Struct. Health Monit. **4**, 195–211 (2005).
[6]J.-B. Ihn, F.-K. Chang, J. Huang, and M. Derriso, "Diagnostic imaging technique for structural health monitoring" in *Proceedings of the 2nd International Workshop on Structural Health Monitoring*, Stanford, CA (2003).
[7]M. Lemistre and D. Balageas, "Structural health monitoring system based on diffracted lamb wave analysis by multiresolution processing," Smart Mater. Struct. **10**, 504–511 (2001).
[8]R. E. Ing and M. Fink, "Time-reversed lamb waves," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 2192–2197 (1998).
[9]E. Bécache, P. Joly, and C. Tsogka, "An analysis of new mixed finite elements for the approximation of wave propagation problems," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **37**, 1053–1084 (2000).
[10]E. Bécache, P. Joly, and C. Tsogka, "Etude d'un nouvel élément finin mixte permettant la condensation de masse (construction and analysis of a new mixed finite element allowing mass lumping)," C. R. Acad. Sci., Ser. I: Math. **324**, 1281–1286 (1997).
[11]J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[12]Jon F. Claerbout, *Imaging the Earth Interior* (Blackwell, Palo Alto, CA, 1985).
[13]G. Beylkin and R. Burridge, "Linearized inverse scattering problems in acoustics and elasticity," Wave Motion **12**, 15–52 (1990).
[14]L. Borcea, G. Papanicolaou, and C. Tsogka, "Interferometric array imaging in clutter," Inverse Probl. **21**, 1419–1460 (2005).
[15]M. Fink and C. Prada, "Acoustic time-reversal mirrors," Inverse Probl. **17**, R1–R38 (2001).
[16]M. Fink, "Time reversed acoustics," Phys. Today **50**, 34–40 (1997).
[17]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[18]D. R. Dowling and D. R. Jackson, "Narrow band performance of phase conjugate arrays in dynamic random media," J. Acoust. Soc. Am. **91**, 3257–3277 (1992).
[19]P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," J. Acoust. Soc. Am. **111**, 238–248 (2002).
[20]A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).
[21]G. Montaldo, G. Lerosey, A. Derode, A. Tourin, J. de Rosny, and M. Fink, "Telecommunication in a disordered environment with iterative time reversal," Waves Random Media **14**, 287–302 (2004).
[22]Stéphane Mallat, *A Wavelet Tour of Signal Processing* (Academic, San Diego, 1998).
[23]T. F. Chan and J. Shen, *Image Processing and Analysis* (Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2005).
[24]J. L. Starck and F. Murtagh, *Astronomical Image and Data Analysis* (Springer, New York, 2006).
[25]C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423 (1948).
[26]S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989).
[27]J. Sporring and J. Weickert, "Information measures in scale-spaces," IEEE Trans. Inf. Theory, **45**, 1051–1058 (1999).
[28]J. Grazzini, A. Turiel, and H. Yahia, "Presegmentation of high-resolution satellite images with a multifractal reconstruction scheme based on an entropy criterium," in *IEEE International Conference on Image Processing ICIP*, Genova, Italy, September 11–14 (2005).
[29]L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D **60**, 259–268 (1992).
[30]Y.-H. R. Tsai and S. Osher, "Total variation and level set based methods in image science," Acta Numerica **1**, 1–61 (2005).
[31]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).
[32]C. Prada, S. Manneville, D. Spoliansky, and M. Fink, "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2076 (1996).
[33]C. Prada and J. L. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," J. Acoust. Soc. Am. **114**, 235–243 (2003).
[34]G. Montaldo, M. Tanter, and M. Fink, "Real time inverse filter focusing by iterative time reversal," J. Acoust. Soc. Am. **112**, 2446–2446 (2002).
[35]M. Cheney, D. Isaacson, and M. Lassas, "Optimal acoustic measurements," SIAM J. Appl. Math. **61**, 1628–1647 (2001).
[36]J. G. Berryman, L. Borcea, G. C. Papanicolaou, and C. Tsogka, "Statistically stable ultrasonic imaging in random media," J. Acoust. Soc. Am. **112**, 1509–1522 (2002).
[37]L. Borcea, G. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," Inverse Probl. **18**, 1247–1279 (2002).
[38]H. Zhao, "Analysis of the response matrix for an extended target," SIAM J. Appl. Math. **64**, 725–745 (2004).
[39]D. H. Chambers and A. K. Gautesen, "Time reversal for a single spherical scatterer," J. Acoust. Soc. Am. **109**, 2616–2624 (2001).
[40]S. Hou, K. Solna, and H. Zhao, "Imaging of location and geometry for extended targets using the response matrix," J. Comput. Phys., **199**, 317–338 (2004).

Reconstruction of source distributions from sound pressures measured over discontinuous regions: Multipatch holography and interpolation

Moohyung Lee[a)]
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

J. Stuart Bolton[b)]
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

(Received 26 June 2006; revised 4 November 2006; accepted 10 January 2007)

In the present work, a method of alternating orthogonal projections is described in the context of near-field acoustical holography; it allows missing (or "not measured") data to be recovered, thus relieving the strictness of measurement requirements related to the use of the discrete Fourier transform. The method described here provides the detailed foundation for the patch holography procedure that has previously been introduced to mitigate finite measurement aperture effects by allowing the sound field to be iteratively extended beyond the measurement aperture. It is also shown that the latter iterative algorithm can be used regardless of the spatial distribution of measured data: i.e., patches can be discontinuous. Numerical simulations performed by using a synthetic sound field created by a point-driven, simply supported plate were used to demonstrate the latter point. In particular, a multipatch holography procedure is described that allows a source distribution to be reconstructed from the hologram pressure measured over multiple, unconnected patches. It is finally shown that a related approach allows spatial resolution enhancement by interpolation between measured points. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2536675]

PACS number(s): 43.60.Sx, 43.60.Pt [EGW] Pages: 2086–2096

I. INTRODUCTION

The use of discrete Fourier transform (DFT)-based, near-field acoustical holography (NAH)[1,2] for sound field visualization is attractive since it is efficient in terms of computation time; it also allows a sound field to be decomposed into wave number components, thus providing physically meaningful information about the sound field (by distinguishing between propagating and evanescent wave components, for example). However, the application of DFT-based NAH is sometimes limited due to the requirement that a sound field should be sampled with a uniform spacing on a surface of constant coordinate in a separable geometry (e.g., planar, cylindrical, and spherical coordinates), and that the hologram surface should extend into a sufficiently large region to avoid the inherent windowing problem related to the use of the DFT.

Various methods have been introduced to address spurious effects resulting from the undue truncation of the sound field. These methods allow NAH measurements to be made over a limited region of interest, thus reducing measurement effort. In statistically optimized NAH (SONAH),[3–5] a plane-to-plane propagation in the spatial domain is performed by evaluating a two-dimensional convolution with a propagation kernel. Since no Fourier transforms are involved in this procedure, the effects of sound field truncation are avoided. In the Helmholtz equation least-squares (HELS) procedure,[6] the sound field is expressed in terms of an orthonormal expansion of spheroidal functions that satisfy the Helmholtz equation, and the Helmholtz equation is solved directly in a least-squares sense. This method was further developed into the combined HELS (CHELS) procedure[7] that combined the advantages of HELS and boundary element method (BEM)-based NAH. In the so-called method of superposition, the sound field is approximated as the superposition of fields produced by a number of simple sources.[8,9] The latter method can be used either to enlarge a finite measurement aperture (extrapolation) or to fill in gaps within the measurement aperture (interpolation). In a wavelet-based method, the sound field is decomposed by using multiresolution analysis (MRA),[10] and spatial filtering is then performed in a selective way before the wave number spectrum is calculated.[11] In an iterative method, referred to as patch NAH (or patch holography), the sound field measured over a finite measurement aperture is extrapolated into the region exterior to the measurement patch by an iterative smoothing process, thus increasing the effective size of a measurement aperture.[12–17]

In the present work, the patch holography procedure is studied in detail, and its applications are generalized. The patch holography procedure comprises two parts: i.e., a missing data recovery (or data restoration) procedure and a holographic projection procedure. Since the second procedure is

a)Author to whom correspondence should be addressed. Electronic mail: leemoohy@ecn.purdue.edu
b)Electronic mail: bolton@purdue.edu

 0001-4966/2007/121(4)/2086/11/$23.00

the same as that employed in usual NAH applications, the focus of the present work is on the first procedure.

The subject of missing data recovery has been studied extensively in the field of image processing. The patch holography procedure is based on the use of a well-known method in that field, referred to as the Papoulis-Gerchberg algorithm (PGa),[18,19] that was originally considered for extrapolation problems. In the present work, it is shown that the latter iterative algorithm can be derived from the method of alternating orthogonal projections,[20] and that its convergence can be established so long as signals satisfy a certain condition regardless of the spatial distribution of measured data; thus, its application can be expanded, for example, to interpolation problems.[21] Also, the conditions required for the success of the procedure are defined here and concerns related to practical implementation are described.

This article is organized in the following manner. In Sec. II, the theoretical background of the procedure described here is provided, and, in Sec. III, numerical simulation results are presented to demonstrate the applicability of the procedure both to sound field reconstruction from the hologram pressure measured over multiple patch regions, referred to here as "multipatch holography," and to spatial resolution enhancement. Finally, conclusions are presented in Sec. IV.

II. FUNDAMENTALS OF SIGNAL RESTORATION

A. Preliminaries

Consider square integrable functions in a Hilbert space, $\mathcal{H}$, in which a norm is defined by $\|f\|=\sqrt{\langle f,f\rangle}$ where $\langle\cdot,\cdot\rangle$ represents the inner product of functions. Since "square integrable" means that the integral of the square of a function's absolute value is finite, the latter norm necessarily has a finite value.

Every function f in $\mathcal{H}$ can be decomposed uniquely as

$$f=g+h \tag{1}$$

where $g\in\mathcal{P}$ and $h\in\mathcal{P}^{\perp}$, the orthogonal complement of $\mathcal{P}$. Here, $\mathcal{P}$ is a closed subspace of $\mathcal{H}$, and $\mathcal{P}^{\perp}$ is defined by

$$\mathcal{P}^{\perp}=\{h\in\mathcal{H}:\langle g,h\rangle=0\ \ \forall\, g\in\mathcal{P}\}. \tag{2}$$

Thus, $\mathcal{H}$ is the internal Hilbert direct sum of $\mathcal{P}$ and $\mathcal{P}^{\perp}$ (i.e., $\mathcal{H}=\mathcal{P}\oplus\mathcal{P}^{\perp}$). In the latter case, g and h are expressed in terms of the orthogonal projection operators projecting onto $\mathcal{P}$ and $\mathcal{P}^{\perp}$ as $g=Pf$ and $h=Qf$, respectively. The orthogonal projection operator is a self-adjoint linear operator (i.e., $P=P^{*}$ and $Q=Q^{*}$) on $\mathcal{H}$ of norm $\leqslant 1$ with the properties that $P^2=P$ and $Q^2=Q=1-P$.

B. Method of alternating orthogonal projections (Ref. 20)

Let P_a, Q_a, P_b, and Q_b represent the orthogonal projection operators projecting onto $\mathcal{P}_a$, $\mathcal{P}_a^{\perp}$, $\mathcal{P}_b$, and $\mathcal{P}_b^{\perp}$, respectively. Missing data recovery represents a task that reconstructs f when only its projection $g=P_af$ onto the known subspace $\mathcal{P}_a$ is given. An arbitrary signal cannot always be recovered from partially known information, of course. Therefore, an additional constraint must be imposed on the nature of the signals to enable the latter task. Suppose that f belongs to the known subspace $\mathcal{P}_b$: i.e.,

$$f=P_bf, \tag{3}$$

then

$$\begin{aligned} g=P_af=P_aP_bf&=(1-Q_a)P_bf=P_bf-Q_aP_bf\\ &=f-Q_aP_bf. \end{aligned} \tag{4}$$

From Eq. (4), f can be uniquely determined from $g=P_af$ if the inverse of $A=P_aP_b=(1-Q_a)P_b$ exists. It was shown by Youla that $\mathcal{P}_b$ and $\mathcal{P}_a^{\perp}$ should have only the zero signal in common for solutions to be unique: i.e.,

$$\mathcal{P}_b\cap\mathcal{P}_a^{\perp}=\{\phi\} \tag{5}$$

where $\cap$ and ϕ denote the intersection and a zero signal, respectively. The latter can also be described by the condition that the homogeneous equation, $Af=0$, has only a trivial solution.

Rearrangement of the last result in Eq. (4) gives

$$f=Q_aP_bf+g, \tag{6}$$

and a recursive relation for recovering f is obtained by using a method of successive approximations for finding the fixed point that satisfies the latter relation: i.e.,

$$f^{(k+1)}=Q_aP_bf^{(k)}+g,\quad k=1,2,\ldots,\quad f^{(1)}=g. \tag{7}$$

Equation (7) is well-posed if

$$\|Q_aP_b\|<1, \tag{8}$$

and the solution of Eq. (6) can then be rewritten explicitly, by using the relation $\Sigma_{n=0}^{\infty}ar^n=a/(1-r)$, as

$$f=\frac{g}{1-Q_aP_b}=\sum_{n=0}^{\infty}(Q_aP_b)^ng. \tag{9}$$

The convergence of Eq. (7) can be established if the condition given in Eq. (5) is also satisfied, which can be proven as follows. After some manipulation, the approximated solution is given by

$$\begin{aligned} f^{(k)}&=\sum_{n=0}^{k-1}(Q_aP_b)^ng=\sum_{n=0}^{k-1}(Q_aP_b)^n(1-Q_aP_b)f\\ &=f-(Q_aP_b)^kf. \end{aligned} \tag{10}$$

According to von Neumann's alternating projection theorem,[22] for every $f\in\mathcal{H}$

$$\lim_{k\to\infty}(Q_aP_b)^kf=f_c \tag{11}$$

where f_c is the projection of f onto the closed subspace, $\mathcal{P}_b\cap\mathcal{P}_a^{\perp}$. If the condition shown in Eq. (5) is valid, then $f_c=0$, and, as a result,

$$\lim_{k\to\infty}f^{(k)}=f-\lim_{k\to\infty}(Q_aP_b)^kf=f. \tag{12}$$

Therefore, the solution of Eq. (7) converges to the desired one provided that the two conditions [i.e., Eqs. (5) and (8)] are satisfied.

The latter convergence is monotonically increasing (i.e., $f^{(k)}$ approaches f from below in its norm), which can be shown by examining the error norm defined by $\|E^{(k)}\|=\|f^{(k)}-f\|$. The following relation is obtained by combining Eqs. (6) and (7):

$$f^{(k+1)}-f=Q_aP_b(f^{(k)}-f), \quad k=1,2,\ldots . \tag{13}$$

It is apparent that $\|E^{(k+1)}\|<\|E^{(k)}\|$ since $\|E^{(k+1)}\|/\|E^{(k)}\|=\|Q_aP_b\|<1$.

In Eq. (7), $f^{(1)}$ is chosen to be g. In fact, the choice of $f^{(1)}$ also has an impact on convergence, which will be discussed here. From Eq. (13), the following relation is finally obtained:

$$f^{(k)}-f=(Q_aP_b)^{k-1}(f^{(1)}-f). \tag{14}$$

When $f^{(1)}=g=(1-Q_aP_b)f$, Eq. (14) is identical to Eq. (10). According to von Neumann's alternating projection theorem,[22] when applied to the composite orthogonal projection operator, the error norm converges to zero if and only if $f^{(1)}-f\in\mathcal{P}_b$: i.e.,

$$\lim_{k\to\infty}\|E^{(k)}\|=\lim_{k\to\infty}\|f^{(k)}-f\|=0, \quad \text{if and only if } f^{(1)}-f\in\mathcal{P}_b. \tag{15}$$

Note that f is already restricted so that $f\in\mathcal{P}_b$, and thus any choice of $f^{(1)}$ is possible provided that $f^{(1)}\in\mathcal{P}_b$. The simplest choice is $f^{(1)}=0$, which results in $f^{(2)}=g$. For this reason, g is usually taken to be the initial approximation by skipping the first iteration with a zero signal.

A method of alternating orthogonal projections is described by "orthogonal" projections onto subspaces, as shown above. More generally, this method can be described as a special case of "projections onto convex sets (POCS)."[23]

C. Choice of the orthogonal projection operators

Equation (7) represents the general relation for recovering missing data from partially known information. Thus, particular orthogonal projection operators can be defined depending on the nature of the problem when Eq. (7) is implemented. In this subsection, the two orthogonal projection operators associated with patch holography applications will be introduced.

1. Sampling operator

The first operator defined here is the sampling operator in the spatial domain. In patch holography applications, the known function g corresponds to the sound pressure truncated (or windowed) by the finite measurement aperture(s), $p_m=p_m(\vec{r})$, where $\vec{r}=(r_1,r_2)$ represents the two-dimensional position vector on the hologram surface. The partially measured pressure, p_m, can be expressed in terms of the sound pressure over the complete region, $p=p(\vec{r})$: i.e.,

$$p_m(\vec{r})=\begin{cases} p(\vec{r}), & \text{when } \vec{r}\in\Gamma_m \\ 0, & \text{when } \vec{r}\in\Gamma_m^{\perp} \end{cases} \tag{16}$$

where Γ_m and $\Gamma_m^{\perp}$ denote the region where measurements are performed and its complement (i.e., the not-measured region), respectively. Equation (16) can be expressed in terms of the sampling operator, $D=D(\vec{r})$, as

$$p_m=Dp \tag{17}$$

where

$$D(\vec{r})=\begin{cases} 1, & \text{when } \vec{r}\in\Gamma_m \\ 0, & \text{when } \vec{r}\in\Gamma_m^{\perp} \end{cases} . \tag{18}$$

From the properties of the sampling operator, it is known that the sampling operator is a self-adjoint linear operator and that $D^2p=Dp$ for all $p\in\mathcal{H}$. Therefore, the sampling operator is an orthogonal projection operator and can be used in place of P_a appearing in Sec. II B.

Further, p can be decomposed in terms of the sampling operator as a form of Eq. (1): i.e.,

$$p=Dp+(1-D)p=p_m+p_m^{\perp}. \tag{19}$$

In Eq. (19), $p_m^{\perp}=(1-D)p$ is the projection of p onto $\Gamma_m^{\perp}$ (i.e., the missing part of the hologram pressure), and it is apparent that $\Gamma_m^{\perp}$ is the orthogonal complement of Γ_m since the inner product between p_m and $p_m^{\perp}$ is zero. Note that the properties of the sampling operator as an orthogonal projection operator are maintained regardless of the spatial distribution of the measured data.

2. Bandlimiting operator

The second operator is related to the additional constraint regarding the nature of the signals. Signals in many practical cases satisfy a certain constraint rather than being completely arbitrary in nature. In particular, among various types of signals, bandlimited signals are frequently observed: the implications of the latter constraint are explored here.

When the sampling operator is defined, p is decomposed in the spatial domain as shown in Eq. (19): but p can also be decomposed by projecting it onto the Fourier domain. Let F and F^{-1} denote the forward and inverse two-dimensional Fourier transform operators, respectively, and let $\bar{P}=\bar{P}(\vec{k})$ represent the wave number spectrum of p where $\vec{k}=(k_1,k_2)$ represents the two-dimensional wave number vector. Then, p is represented by the inverse Fourier transform of the wave number spectrum, and can be decomposed as

$$p=F^{-1}(\bar{P}(\vec{k}))=F^{-1}(\bar{P}(\vec{k})_{\in\Omega_B}+\bar{P}(\vec{k})_{\in\Omega_B^{\perp}})=p_B+p_B^{\perp} \tag{20}$$

where Ω_B denotes the region in k-space supported by $a_1\leqslant k_1\leqslant b_1$ and $a_2\leqslant k_2\leqslant b_2$, and $\Omega_B^{\perp}$ denotes its complement: i.e., in a planar case, for example,

$$p_B=F^{-1}(\bar{P}(\vec{k})_{\in\Omega_B})=\frac{1}{4\pi^2}\int_{a_y}^{b_y}\int_{a_x}^{b_x}\bar{P}(k_x,k_y)e^{j(k_xx+k_yy)}dk_x\,dk_y, \tag{21}$$

and p_B and $p_B^{\perp}$ are, respectively, the orthogonal projections of p onto Ω_B and $\Omega_B^{\perp}$. Since p_B and $p_B^{\perp}$ comprise, respectively, the wave number components inside and outside the region, Ω_B, it is clear that the inner product between p_B and $p_B^{\perp}$ is zero. Equation (20) can also be expressed in terms of an orthogonal projection operator: i.e.,

$$p = p_B + p_B^{\perp} = F^{-1}LFp + F^{-1}(1-L)Fp = Bp + (1-B)p \tag{22}$$

where $L=L(\vec{k})$ is a $\vec{k}$-space low-pass filter defined by

$$L(\vec{k}) = \begin{cases} 1, & \text{when } \vec{k} \in \Omega_B \\ 0, & \text{when } \vec{k} \in \Omega_B^{\perp} \end{cases}, \tag{23}$$

and $B=F^{-1}LF$ is referred to as the bandlimiting operator. Note here that the shape of the k-space low-pass filter is made rectangular rather than circular since signals may have different bandwidths in the two k-space directions.

Suppose that p is bandlimited in k-space to Ω_B; then, p satisfies the condition that $p=Bp$ (thus, B corresponds to the quantity P_b discussed in Sec. II B). Since high wave number, evanescent sound field components decay quickly, the sound pressure measured on the hologram surface usually (at least weakly) satisfies the latter bandlimitedness condition.

D. Iterative algorithm for generalized patch holography

By combining the two orthogonal projection operators defined above with Eq. (7), the final form of the iterative relation for recovering a missing part of a bandlimited signal is obtained:[24] i.e.,

$$p^{(k+1)} = (1-D)Bp^{(k)} + p_m, \quad k = 1,2,\ldots, \quad p^{(1)} = p_m. \tag{24}$$

Equation (24) corresponds to the Papoulis-Gerchberg algorithm (PGa)[18,19] in two dimensions. Since continuous functions cannot be both bandlimited and space-limited at the same time, the closed subspace onto which p is projected by $(1-D)B$ contains only the zero signal, thus satisfying the convergence condition given in Eq. (5).

Equation (24) represents a "smooth-and-replace" procedure. The iteration starts with the initial pressure that is prepared by filling the region where data is not available with zeros. The bandlimiting operation that removes the high wave number components of $p^{(k)}$ resulting from the sharp transition from the measured data to zero is then performed to obtain the smoothed pressure (i.e., $Bp^{(k)}$). Finally, the input pressure for the next iteration, $p^{(k+1)}$, is created by replacing the smoothed pressure over the measurement region with the originally measured pressure. In this way, the pressure in the region where data is not available gradually approaches the true value in that region while the pressure over the measurement region is maintained at its original value.

In practice, Eq. (24) is implemented numerically by using a sampled set of data; thus, it is convenient to rewrite it in matrix-vector form: i.e.,

$$\mathbf{p}^{(k+1)} = (\mathbf{I}-\mathbf{D})\mathbf{B}\mathbf{p}^{(k)} + \mathbf{p}_m = (\mathbf{I}-\mathbf{D})\mathbf{F}^{-1}\mathbf{L}\mathbf{F}\mathbf{p}^{(k)} + \mathbf{p}_m,$$
$$k = 1,2,\ldots, \quad \mathbf{p}^{(1)} = \mathbf{p}_m, \tag{25}$$

where $\mathbf{I}$ represents the identity matrix. As for Eq. (8), the convergence of the latter successive approximations can be checked by examining the eigenvalues of $\mathbf{T}=(\mathbf{I}-\mathbf{D})\mathbf{B}$, referred to as the iteration matrix. It is well known that such iterative methods converge if the absolute value of the largest eigenvalue of the iteration matrix is strictly smaller than unity.

In the case of continuous signals in an infinite domain, the convergence to the desired solution over a complete region is well-established in the absence of noise, as proven earlier. In the case of discrete signals, however, a unique solution does not exist due to the effects of the artificial truncation of an infinite domain and the discretization of continuous functions.[25] In addition, since the inclusion of measurement noise is inevitable in practical cases,[26] the bandlimitedness assumption is not likely to be strictly valid. Due to the effects just mentioned, convergence can only be achieved to a limited extent, and, in the worst case, solutions may not converge. The latter problem can, however, be addressed by choosing the proper termination condition for the sequence of iterations:[19] i.e., the iteration is terminated when

$$d^{(k)} = \|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}\| \leqslant \varepsilon \tag{26}$$

where ε is an *ad hoc* factor having a small value. Thus, the result is a solution that minimizes an error norm in a given situation, with the result that the region in which accurate recovery is ensured is limited, which is the situation usually observed in extrapolation problems as seen in many previous patch holography applications. Due to the latter fact, the source distribution, which is reconstructed by a holographic projection of the pressure restored by using the latter procedure, may not be identified accurately over the complete region. Even in the latter case, however, the data recovery procedure reduces the spurious effects associated with an incomplete measurement, thus allowing a significant improvement in reconstruction results, at least in the region directly under the measurement patch.[17]

It follows from the properties of the sampling operator that there is no restriction on the spatial distribution of measured data, so the algorithm described here can be used in applications other than the extrapolation of hologram pressure measured over a single patch. However, the convergence rate depends intimately on the distribution of the known samples.[24] Let λ denote the eigenvalues of $\mathbf{T}^{\mathrm{H}}\mathbf{T}=\mathbf{B}(\mathbf{I}-\mathbf{D})\mathbf{B}$ where the superscript H denotes the conjugate transpose operator. A smaller λ results in a faster convergence rate, and the convergence rate of the given iteration matrix changes within a range determined by $\lambda_{\max}$ and $\lambda_{\min}$ (the smallest nonzero eigenvalue): i.e., the best convergence rate can be achieved when the region where energy is concentrated (i.e., near a peak) is included in the sampling set, which corresponds to $\lambda_{\min}$, and, on the contrary, the worst case occurs when energy is concentrated in the complement of the sampled set. When measurement locations are well distributed (e.g., in interpolation problems), $\lambda_{\max}$ is relatively small compared to unity, and the difference between $\lambda_{\max}$ and $\lambda_{\min}$ is usually small. Thus, the convergence in interpolation problems is relatively fast, and the selection of the sampling set does not have a significant impact on the convergence rate, which is to be expected since peak locations can usually be included in any choice of a well-distributed sampling set. In comparison, when the missing data lie in a large contiguous region (e.g., in extrapolation problems), $\lambda_{\max}$ is usually close to unity, and the difference between

λ_{max} and λ_{min} is large. Thus, convergence tends to be slow in this case, and its rate is sensitive to the choice of sampling locations. Therefore, when Eq. (25) is applied to extrapolation problems, measurement locations should be selected with care so that as many peaks as possible are included in the measured data whenever *a priori* knowledge of peak locations is available. Also, the convergence rate and the region where accurate recovery is possible depend on the sampling density defined by the ratio of the number of measured data points to the total number of data points. Needless to say, a higher sampling density broadens the possible region of accurate recovery and allows for faster convergence.

To ensure the success of this procedure, the cutoff wave numbers of the bandlimiting operator must be chosen appropriately. In most cases, the k-space bandwidth of a signal is not known *a priori*, which is, however, not critical in practical implementation since its selection within a sensible range usually yields results with reasonable accuracy. The latter point will be discussed here in detail.

Let $\mathbf{p}$ be supported by $\Omega^{(0)}=[a_1,b_1]\times[a_2,b_2]$, let $\Omega^{(+)}$ and $\Omega^{(-)}$ represent the k-space regions having larger and smaller apertures than $\Omega^{(0)}$, respectively, and let $\Omega=[k_{1cn},k_{1cp}]\times[k_{2cn},k_{2cp}]$ be the low-pass band of the bandlimiting operator. In Fig. 1, the various k-space regions just defined are shown schematically: the regions are shown in one dimension for illustrative purpose, and the solid and dotted lines depict the wave number spectra of the complete and truncated pressures, respectively. With respect to the various choices of the k-space region for defining the cutoff of the bandlimiting operator, $\mathbf{p}$ is expressed by

$$\mathbf{p}=\begin{cases}\mathbf{B}^{(0)}\mathbf{p} & \text{when } \Omega=\Omega^{(0)}\\ \mathbf{B}^{(+)}\mathbf{p} & \text{when } \Omega=\Omega^{(+)}\\ \mathbf{B}^{(-)}\mathbf{p}+(\mathbf{B}^{(0)}-\mathbf{B}^{(-)})\mathbf{p} & \text{when } \Omega=\Omega^{(-)}.\end{cases} \tag{27}$$

Evidently, the best choice is $\Omega=\Omega^{(0)}$. When the cutoff is chosen to be larger than the actual bandwidth (i.e., $\Omega=\Omega^{(+)}$), results are affected by the effects of the wave number components (which are associated with an incomplete measurement and noise) included in the region between $\Omega^{(0)}$ and $\Omega^{(+)}$. However, the latter effect is not significant since $\mathbf{p}$ belongs to the subspace defined by the k-space regions, $\Omega^{(+)}$, thus satisfying the condition shown in Eq. (3), and since the amount of energy included in those wave number components is relatively small. Thus, the upper limit can be a value that can filter out a sufficient amount of the high wave number components considered to result from the windowing effect (e.g., a_h and b_h in Fig. 1). When the cutoff is chosen to be smaller than the bandwidth of a signal (i.e., $\Omega=\Omega^{(-)}$), the wave number components lying in the region between $\Omega^{(-)}$ and $\Omega^{(0)}$ [i.e., $(\mathbf{B}^{(0)}-\mathbf{B}^{(-)})\mathbf{p}$ that corresponds to the shaded region in Fig. 1] are filtered out, thus causing an aliasing error. Since an aliasing error may cause results to diverge, the lower bound of the possible cutoff should be set relatively carefully so that the wave number components in that region contain a negligible amount of energy. In practice, the cutoff wave number can be chosen to be a value larger than the highest wave number of the truncated pressure having significant amplitude (e.g., a_l and b_l in Fig. 1). The cutoff

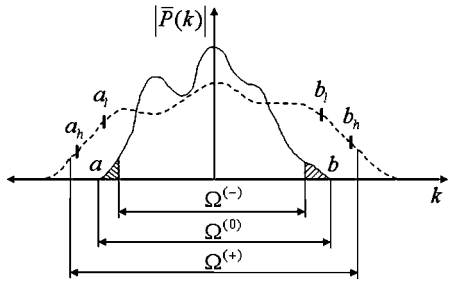


FIG. 1. Typical shapes of the wave number spectra estimated from the complete (solid line) and truncated (dotted line) pressures, and various low-pass bands of the bandlimiting operator (a_l and b_l, and a_h and b_h represent, respectively, the lower and upper bound of the possible choices of cutoff values).

chosen in the latter way is usually safe since the wave number spectrum of the complete pressure is confined to a smaller region than that of the truncated pressure, thus allowing a significant aliasing error to be avoided. The procedure just described can be applied to the two k-space directions independently.

The procedure described above provides insight into how to choose the cutoff wave numbers of the bandlimiting operator. Since the degree to which data can be restored is inevitably limited by various effects, it can be said, in a conservative sense, that the application of patch holography is usually restricted to the reconstruction of a source distribution in the region(s) directly under the measurement patch(es). Arguably, the latter fact gives relative flexibility to the choice of the cutoff values, as described above.

Recently, regularization has been incorporated to determine the bandlimiting operator without *a priori* knowledge of the bandwidth of a signal.[13,16] In that approach, a k-space low-pass filter is constructed by regularization at each iteration, and, in the best case, the cutoff of the latter filter can approach its optimal value from above as the iteration proceeds: i.e., at the early stages of the iteration, the cutoff provided by regularization is typically far larger than the bandwidth of the signal since the wave number spectrum of the truncated pressure contains high wave number components beyond the actual bandwidth, and the wave number spectrum of the pressure resulting after each iteration is confined to a progressively smaller region compared with that obtained at the previous iteration, thus resulting in a lower cutoff in the following iteration, and so on.

However, there still exist concerns related to the use of the k-space low-pass filter provided by the latter regularization-based approach (which will be referred to as the regularization filter hereafter). First, strictly speaking, the bandlimiting operator constructed by using the regularization filter does not fall into the category of an orthogonal projection operator because of the regularization filter's tapered shape, which is, however, not a significant concern in practice so long as the regularization filter has a steep slope. Second, the bi-axisymmetric characteristic in k-space of the regularization filter cannot deal with all bandlimited signals appropriately: i.e., it is more appropriate that the shape of the k-space low-pass filter should be rectangular (or elliptic) depending on the bandwidths in the two k-space directions, and

that its center be located at wave numbers other than the zero wave number (i.e., the origin of k-space). However, a regularization filter does not have the latter capabilities. Finally, the convergence of the cutoff provided by a regularization-based method to the bandwidth of a signal is not always ensured. Recently, an iterative algorithm for determining a regularization filter was proposed.[16] In the latter work, it was suggested that it may be possible to obtain a regularization filter similar to the regularization filter for the complete pressure, and that the resulting filter can be used in a data restoration procedure. For the latter approach to be strictly valid, it should be possible for the complete pressure to be fully recovered from the truncated pressure so that the regularization filter obtained at each iteration can finally approach that of the complete pressure as the iteration proceeds: the latter cannot, however, generally be achieved, especially when the size of the measurement aperture is far smaller than that of the full aperture. More serious problems occur, for example, when the dominant wave number components are concentrated in the central part of radiation region (e.g., the sound field radiated by an oscillating piston at a high frequency). In the latter case, the resulting cutoff is far larger than the desired value since the smallest possible cutoff given by a regularization procedure is the acoustic wave number, $k=\omega/c$, at a given frequency (see Fig. 2 in Ref. 27 and the related discussion). Nevertheless, except in some cases, the latter algorithm may still be useful since the cutoff of the resulting regularization filter is likely to be within a range of the possible cutoffs in many practical cases. Thus, when regularization is to be used for constructing the bandlimiting operator, it may be a good idea to check whether the cutoff provided by regularization is given in the range determined by a strategy for selecting the cutoff as described earlier.

To conclude, the restored hologram pressure obtained by using the data restoration procedure described in this section can then be used as an input to the holographic projection procedure.[2] Since the effective size of the measurement aperture can be increased by extending a truncated sound field smoothly into a region exterior to the measurement aperture, the reconstruction error resulting from a sharp transition of a sound field from finite values to zero can be reduced.

III. NUMERICAL SIMULATION RESULTS

A. Simulation description

In this section, numerical simulation results obtained by using a model comprising a point-driven, simply supported steel plate within an infinite baffle are presented to demonstrate the applicability of the algorithm described above to multipatch holography and spatial resolution enhancement.

First, the velocity distribution on the steel plate was created by

$$\dot{w}(x,y,\omega)=\frac{j\omega}{\rho h_t}\sum_{m=1}^{10}\sum_{n=1}^{10}\frac{\Phi_{mn}(x_0,y_0)\Phi_{mn}(x,y)}{\omega^2-\omega_{mn}^2} \tag{28}$$

where Φ_{mn} and ω_{mn} represent, respectively, the normal modes and the natural frequencies: i.e.,

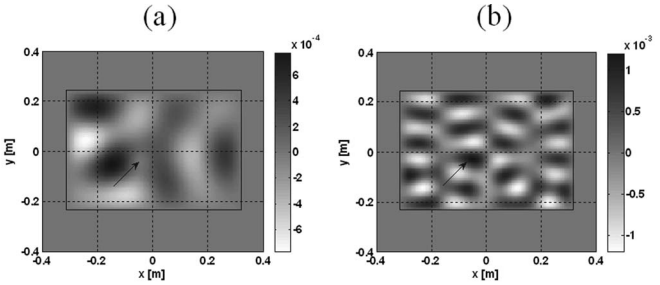


FIG. 2. Velocity distributions on the source surface (the rectangle and arrow represent the region occupied by the steel plate and the location where the point force was applied, respectively): (a) at 1 kHz; (b) at 3.8 kHz.

$$\Phi_{mn}=\frac{2}{\sqrt{L_xL_y}}\sin\left(\frac{m\pi x}{L_x}\right)\sin\left(\frac{n\pi y}{L_y}\right), \tag{29}$$

$$\omega_{mn}=\sqrt{\frac{Eh_t^2}{12\rho(1-\nu^2)}}\left[\left(\frac{m\pi}{L_x}\right)^2+\left(\frac{n\pi}{L_y}\right)^2\right]. \tag{30}$$

In this simulation, the dimension of the rectangular steel plate was $64(L_x)\times48(L_y)$ cm with a thickness, h_t, of 5 mm, the force was applied at $(x_0,y_0)=(-5.5,-4)$ cm, the Young's modulus, E, was 2×10^{11} Pa, the Poisson's ratio, ν, was 0.28, the density, ρ, was 7860 kg/m^3, and the natural frequency of the highest mode included in this simulation, $f_{10,10}=\omega_{10,10}/(2\pi)$, was approximately 8 kHz.

The sound pressure on the planar hologram surface at a height of 3 cm was then calculated by the holographic forward-projection of the surface velocity. To do this, the surface velocity was sampled with a 0.5-cm lattice spacing in the x- and y-directions, and a sufficiently large number of zeros were added to simulate a baffle and to avoid wrap-around error. The total number of samples in each direction after zero padding was 512 (thus, the complete region extends between $x,y=\pm1.28$ m), and random noise (SNR $=40$ dB) was added to the calculated hologram pressure. In Figs. 2 and 3, respectively, the velocity distributions on the source surface and the sound pressure on the hologram surface are presented at 1 and 3.8 kHz. Results are plotted within the region between $x,y=\pm0.4$ m.

B. Multipatch holography

The extrapolation of the hologram pressure measured over a single patch has been studied previously.[12–17] In this

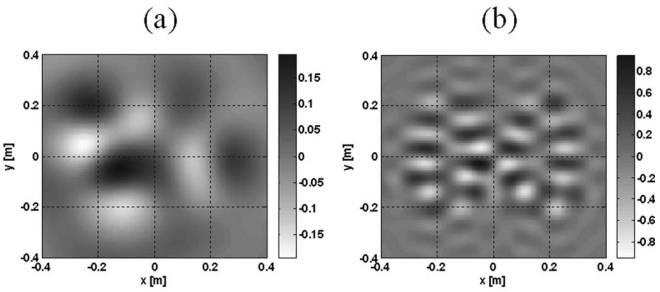


FIG. 3. Sound pressures on the hologram surface at a height of 3 cm (the real parts are plotted): (a) at 1 kHz; (b) at 3.8 kHz.

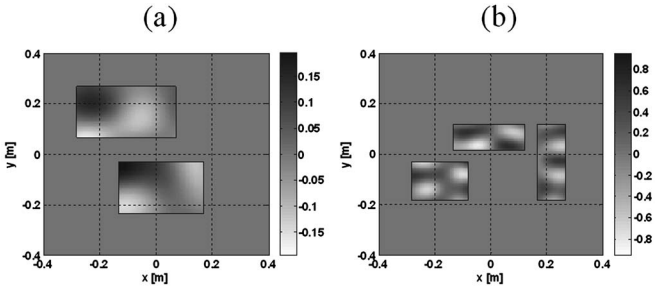


FIG. 4. Sound pressures on the hologram surface truncated by the measurement patches (the real parts are plotted): (a) at 1 kHz; (b) at 3.8 kHz.

simulation, the measurement patches comprised two or three unconnected subpatches at 1 and 3.8 kHz, respectively (see Fig. 4).

In Figs. 5 and 6, the wave number spectra of the pressures measured over the complete region and truncated by the measurement patches, respectively, are shown. As seen in Fig. 5, the wave number spectra of the complete pressures were confined in the regions defined by approximately $k_x, k_y = \pm 40\ \mathrm{m}^{-1}$ at 1 kHz and $k_x, k_y = \pm 80\ \mathrm{m}^{-1}$ at 3 kHz. However, the exact values of the latter wave numbers cannot be known from the wave number spectrum of the truncated pressure due to a leakage of the low wave number components into the high wave number region beyond the bandwidth of the complete pressure (see Fig. 6). Based on the strategy described in Sec. II D, it is expected that the choice of the cutoff wave numbers in both directions between, say, 35 m^{-1} (i.e., a value that does not filter out the wave number components having significant amplitudes) and 70 m^{-1} (i.e., a value that filters out a sufficient amount of the high wave number components resulting from the sharp transition of a sound field) at 1 kHz and between 75 and 100 m^{-1} at 3.8 kHz can yield results with a reasonable accuracy.

To demonstrate the effect of the choice of the cutoff on the performance of the data restoration procedure, the results obtained by using five k-space low-pass filters having different cutoff values are compared at 1 kHz. In all five cases, the four cutoff wave numbers that define the k-space low-pass filter were set to be the same in their absolute values (i.e., $|k_{xcn}| = |k_{xcp}| = |k_{ycn}| = |k_{ycp}| = k_c$, thus resulting in square filters), and the values of k_c were 25, 35, 40, 50, and 70 m^{-1}. In Fig. 7, the percentage root-mean-square errors of the extended pressures (i.e., $100 \times \|\mathbf{p}^{(k)} - \mathbf{p}\| / \|\mathbf{p}\|$) evaluated over the complete hologram surface are plotted with respect to the number of iterations. When the cutoff of the bandlimiting

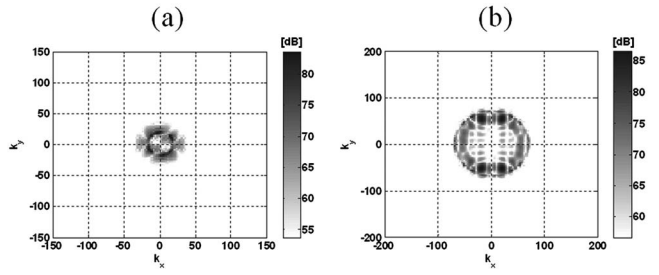


FIG. 5. Wave number spectra of the hologram pressures over the complete region: (a) at 1 kHz; (b) at 3.8 kHz.

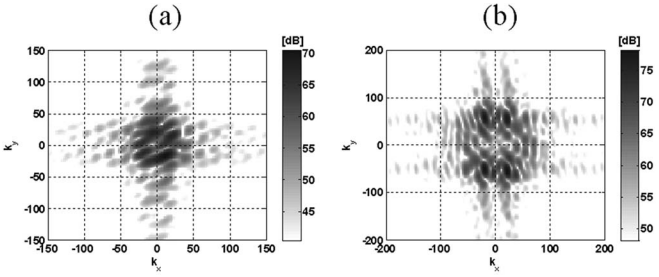


FIG. 6. Wave number spectra of the hologram pressures truncated by the measurement patches: (a) at 1 kHz; (b) at 3.8 kHz.

operator was chosen to be equal to (i.e., 40 m^{-1}) or larger than (i.e., 50 and 70 m^{-1}) the actual bandwidth of the complete pressure, it can be seen that the error decreased monotonically as the number of iterations increased in all cases. However, when the cutoff was made progressively larger, the convergence rate decreased [see Fig. 7(a)]. When the cutoff was chosen to be 35 m^{-1} (i.e., a value smaller than the actual bandwidth, but for which the amount of energy truncated by the corresponding bandlimiting operator was small), the error was bounded, but it did not decrease gradually. Thus, the latter value is the smallest possible choice of the cutoff at 1 kHz. In contrast, when the cutoff was set to 25 m^{-1}, it can be seen that the error actually increased as the number of iterations increased [see Fig. 7(b)]. The latter results support the strategy outlined above for determining the cutoff of the bandlimiting operator.

In a real implementation, the termination of the iteration is determined by examining the difference between the pressures obtained in two consecutive iterations [i.e., Eq. (26)]. In Fig. 8, the values of $d^{(k)} = \|\mathbf{p}^{(k+1)} - \mathbf{p}^{(k)}\|$ obtained when the various values of the cutoff were used are plotted with respect to the number of iterations. It can be seen that the values in all the cases decreased as the number of iterations increased. Attention should be paid to the result in the case when the cutoff was set to 25 m^{-1}, in particular, since in that case the solution diverged while at the same time $d^{(k)}$ progressively decreased. Thus, it should be kept in mind when implementing Eq. (26) that the value of $d^{(k)}$ represents just the degree to which the pressure changes after each iteration rather than the degree of convergence to the desired solution. In all cases, it can be seen that the latter value decreased at a very slow rate after a certain number of iterations (say, 300 iterations). As mentioned earlier, the extent of the region where accurate recovery is possible is largely "predetermined" by the combined effects of various factors. Recovery in the latter region is usually achieved at an early stage of the iteration (i.e., when $d^{(k)}$ is decreasing rapidly). Additional iterations rarely contribute to broadening the region of accurate recovery but mostly to reducing the error norm by increasing the levels of the pressure at locations remote from the patches: i.e., the pressure recovered at the latter locations does not strictly converge to the desired solution, or at least does so extremely slowly. Therefore, it may be useful to examine the rate of decrease of $d^{(k)}$ to determine the termination point: i.e., the optimal (in a practical sense and possibly not "best") number of iterations for termination can be determined, first, by finding the number of iterations at

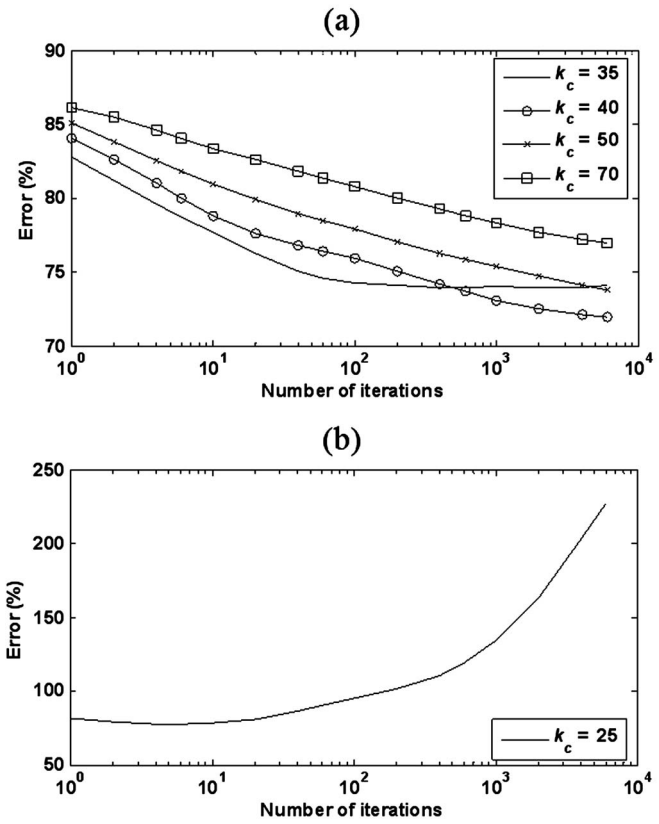


FIG. 7. A comparison of the pressure recovery errors with respect to the number of iterations (in the multipatch measurement case at 1 kHz): (a) when the cutoff, k_c was 35, 40, 50, and 70 m^{-1}; (b) when the cutoff, k_c was 25 m^{-1}. Errors were evaluated over the complete hologram region.

which the rate of decrease slows rapidly (the latter point corresponds graphically to the corner of a $d^{(k)}$-curve and can be found numerically by locating the peak of the second derivative of $d^{(k)}$). Then, the iteration can be terminated after some additional number of iterations for safety.

In Figs. 9 and 10, the extended pressures are shown at 1 and 3.8 kHz, respectively. In Fig. 9, in particular, the extended pressures obtained by using the various cutoff wave numbers with which convergence was achieved are compared. Based on the observation just mentioned above, the iteration was terminated after 400 iterations. In all cases, it can be seen that the pressures at the locations near the measurement patches were recovered to a reasonable degree, but that convergence could not be achieved over the complete region, as discussed above. The pressure recovered at 3.8 kHz showed the same result (see Fig. 10). The latter pressure was obtained after 600 iterations when the cutoff was set to be 80 m^{-1}, which was considered to be the optimal value at this frequency.

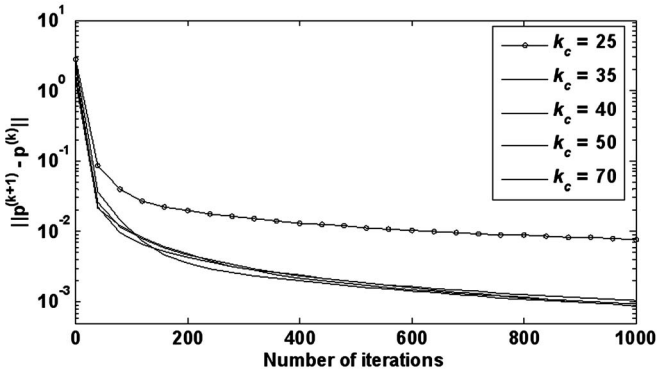


FIG. 8. A comparison of $\|\mathbf{p}^{(k+1)}-\mathbf{p}^{(k)}\|$ with respect to the number of iterations for the various cutoff wave numbers.

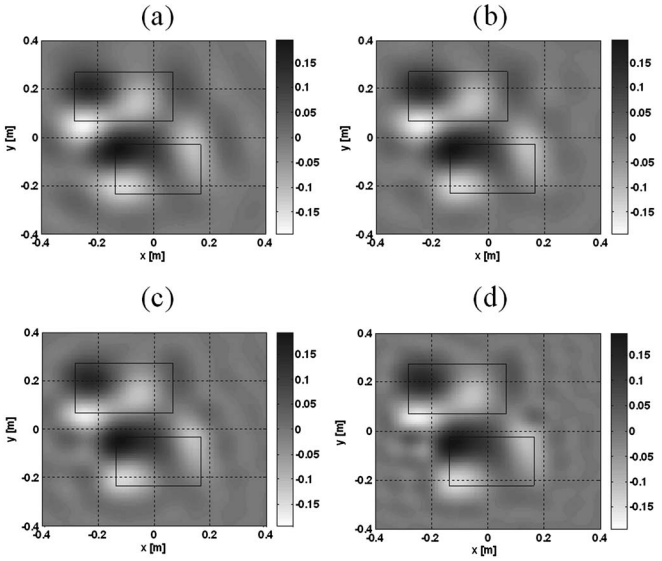


FIG. 9. Extended hologram pressures at 1 kHz (after 400 iterations): (a) obtained when k_c=35 m^{-1}; (b) obtained when k_c=40 m^{-1}; (c) obtained when k_c=50 m^{-1}; (d) obtained when k_c=70 m^{-1}.

In Tables I and II, the percentage root-mean-square errors of the surface velocities reconstructed by using various hologram pressures are compared at 1 and 3.8 kHz, respectively. Three pressures were used to reconstruct the surface velocity: the pressure measured over the complete hologram surface, the pressure truncated by the measurement patches, and the pressure extended by the procedure described here. Error was evaluated over two regions: the entire region of the plate and the regions directly under the measurement patches. Compared with the errors obtained when the truncated pressures were used, it can be seen that the errors were reduced significantly when the extended pressures were used. Also, since the pressures could only be extended in a limited region, errors evaluated over the entire region of the plate were larger than those evaluated only over the regions directly under the measurement patches. As already reported, a relatively larger amount of error is distributed near the regions corresponding to the edges of the measurement patches compared with that in their central region.[17] Thus, better re-

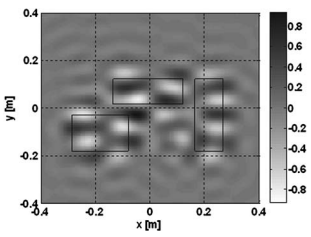


FIG. 10. Extended hologram pressures at 3.8 kHz (obtained after 600 iterations when k_c=80 m^{-1}).

TABLE I. A comparison of the percentage root-mean-square errors of the surface velocities reconstructed by using various hologram pressures at 1 kHz.

		Evaluated over the entire region	Evaluated over the region directly under the patches
When the complete pressure was used		13.0	13.8
When the truncated pressure was used		119.5	104.9
When the extended pressure was used (after 400 iterations)	$k_c=35$ m^{-1}	35.9	19.4
	$k_c=40$ m^{-1}	49.2	24.8
	$k_c=50$ m^{-1}	53.0	34.8
	$k_c=70$ m^{-1}	67.2	48.2

TABLE II. A comparison of the percentage root-mean-square errors of the surface velocities reconstructed by using various hologram pressures at 3.8 kHz.

		Evaluated over the entire region	Evaluated over the region directly under the patches
When the complete pressure was used		8.9	4.7
When the truncated pressure was used		92.4	83.4
When the extended pressure was used (after 600 iterations)	$k_c=80$ m^{-1}	51.9	10.7

sults can be obtained by making measurements over a region slightly larger than that of interest. Table I also provides a comparison of errors corresponding to the various choices of cutoff values. In all cases, a significant improvement was observed, but the use of the cutoff values close to the actual bandwidth gave better results.

C. Spatial resolution enhancement

It may also be of interest to enhance the spatial resolution of a measurement by interpolating the sound field between measured points: i.e., it is assumed that the data between the measured points is "missing." The initial pressure is prepared by inserting zeros (whose number is determined by the spatial resolution to be achieved) on uniform grids between the measured points (see Fig. 11 where the squares and circles depict the locations where data is sampled and zeros are inserted, respectively). To simulate a measurement on a coarse grid, the hologram pressures shown in Fig. 3 were sampled with an increment of 4 cm in both directions [see Figs. 12(a) and 13(a)]. In Figs. 12–16, $\times c$ denotes that the spatial resolution was enhanced by a factor of c. It can be seen in Figs. 12 and 13 that spatial resolution was enhanced successfully. It was also noted that the results converged far faster than in the extrapolation cases shown above (e.g., convergence was achieved after only 20 iterations or so when spatial resolution was doubled). In fact, this example is the case in which the fastest convergence rate can be achieved.

In the latter case, selecting the cutoff wave numbers of the bandlimiting operator is straightforward since the bandwidth of a signal is a property that does not change depending on the sampling rate. In Fig. 14, the wave number spectra of the zero-inserted hologram pressures are shown at 3.8 kHz [compare with Fig. 5(b)]. It can be seen that a replica of the wave number spectrum before inserting zeros appears periodically without a change in shape. Thus, the exact bandwidth of a signal can be "known" from the wave number spectrum either of the original or zero-inserted pressure in this case, and the cutoff wave numbers should be chosen accordingly.

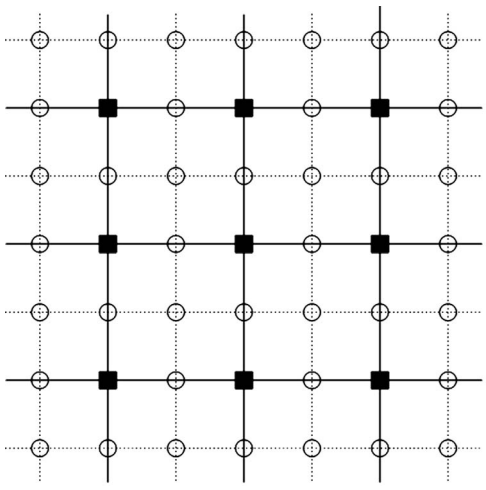


FIG. 11. The preparation of an initial pressure for spatial resolution enhancement (the example is shown for the case when the resolution is to be doubled, and the squares and circles depict the locations where data is sampled and zeros are inserted, respectively).

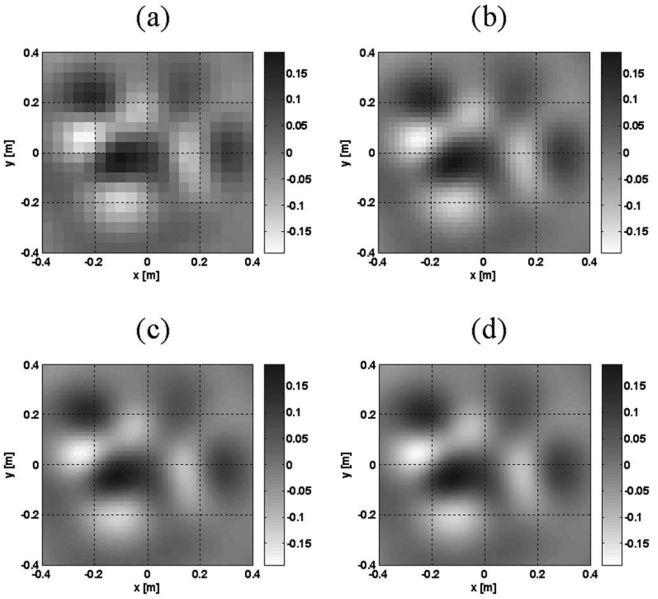


FIG. 12. Spatial resolution enhancement results at 1 kHz: (a) measured with $\Delta x, \Delta y=4$ cm; (b) enhanced ($\times 2$, after 20 iterations); (c) enhanced ($\times 4$, after 75 iterations); (d) enhanced ($\times 8$, after 320 iterations).

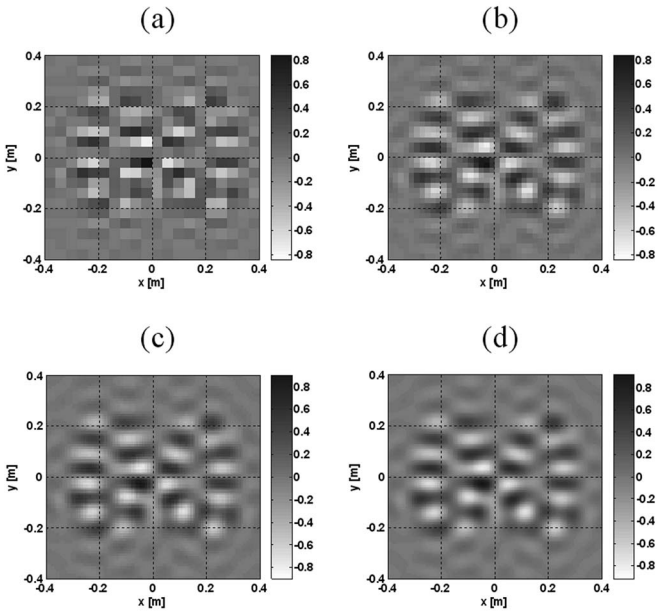


FIG. 13. Spatial resolution enhancement results at 3.8 kHz: (a) measured with $\Delta x, \Delta y = 4$ cm; (b) enhanced ($\times 2$, after 22 iterations); (c) enhanced ($\times 4$, after 80 iterations); (d) enhanced ($\times 8$, after 320 iterations).

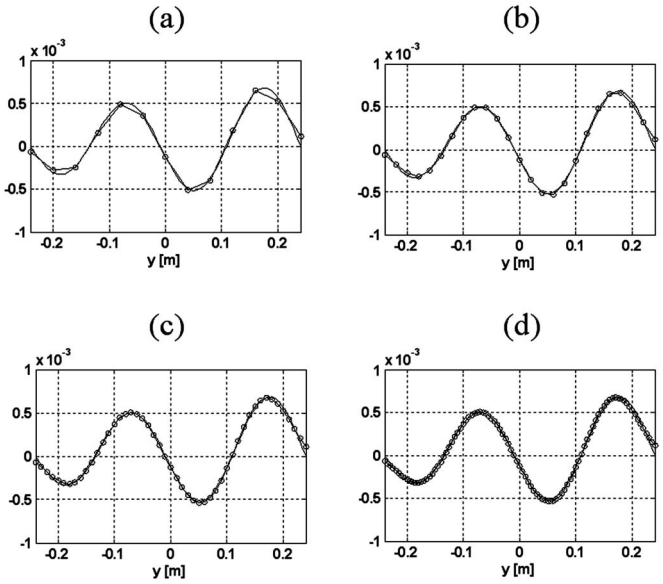


FIG. 15. A comparison of the surface velocities reconstructed by using the measured and resolution-enhanced hologram pressures with the exact one at 1 kHz (the results are compared at $x = -20$ cm): (a) measured with $\Delta x, \Delta y = 4$ cm; (b) enhanced ($\times 2$); (c) enhanced ($\times 4$); (d) enhanced ($\times 8$).

In Figs. 15 and 16, the surface velocities reconstructed by using the original and resolution-enhanced hologram pressures are compared to the exact ones. It can be seen that the benefit of spatial resolution enhancement is more apparent at 3.8 kHz since the number of samples per wavelength at 1 kHz was already sufficient. According to the sampling theories, two samples per wavelength are required to avoid aliasing, but it is well known that a larger number of samples per wavelength are necessary to describe the signal shape and locate its peaks correctly.

IV. CONCLUSIONS

In the present work, a method of alternating orthogonal projections has been described, and it was shown that an iterative algorithm for recovering the hologram pressure at the locations where measurements are not performed can be derived from it by defining the sampling and bandlimiting operators as the orthogonal projection operators. From the properties of the sampling operator, it is known that there is no restriction to the distribution of measurement locations. Thus, the procedure described here can be applied to hologram pressures measured over arbitrary locations (superimposed on uniform grids if the discrete Fourier transform is to be used). As assumed when the iterative algorithm was derived, it is important for signals to be bandlimited in k-space for this procedure to be successful, and the cutoff wave numbers of the bandlimiting operator should be chosen properly: approaches and concerns related to the latter task were described, but the existing approaches still need to be further developed. Also, the practical limitation of the procedure resulting from the effects of the artificial truncation of an infinite domain and the discretization of continuous functions was discussed: i.e., when combined with the effect of measurement noise, the latter effect causes the region where ac-

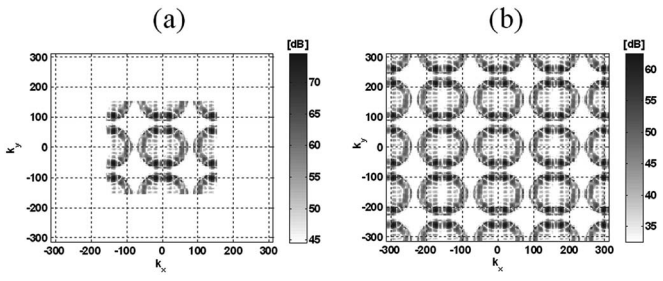


FIG. 14. Wave number spectra of the zero-inserted pressures at 3.8 kHz: (a) zero-inserted ($\times 2$); (b) zero-inserted ($\times 4$).

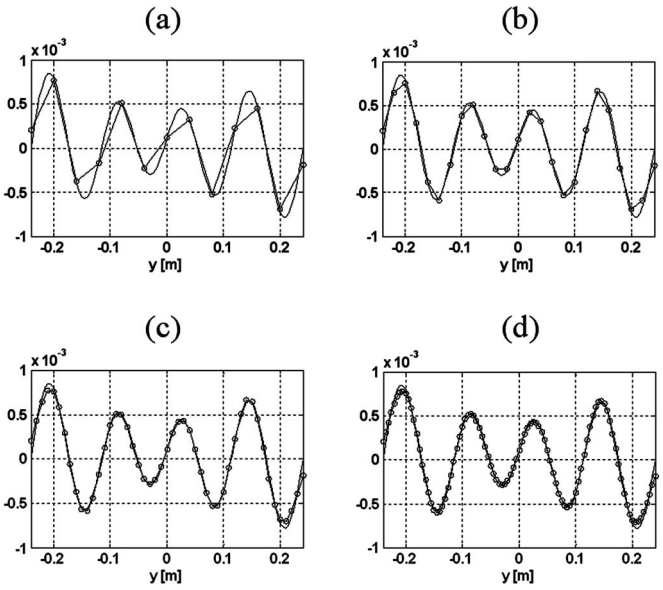


FIG. 16. A comparison of the surface velocities reconstructed by using the measured and resolution-enhanced hologram pressures with the exact one at 3.8 kHz (the results are compared at $x = -24$ cm): (a) measured with $\Delta x, \Delta y = 4$ cm; (b) enhanced ($\times 2$); (c) enhanced ($\times 4$); (d) enhanced ($\times 8$).

curate recovery can be achieved to be limited, which is the case usually observed in extrapolation problems. Numerical simulation results obtained in two cases were presented to demonstrate the various applications of the procedure: i.e., the extrapolation of the hologram pressure measured over multiple, distinct patches, and spatial resolution enhancement by interpolation between measured points.

[1]E. G. Williams, J. D. Maynard, and E. J. Skudrzyk, "Sound source reconstructions using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).

[2]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).

[3]R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," in Proceedings of 6th International Congress on Sound and Vibration (1999).

[4]J. Hald, "Patch near-field acoustical holography using a new statistically optimal method," in Proceedings of INTER-NOISE 2003, No. 975 (2003).

[5]Y.-T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized nearfield acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2364 (2005).

[6]Z. Wang and S. F. Wu, "Helmholtz equation—least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).

[7]S. F. Wu and X. Zhao, "Combined Helmholtz equation—least-squares method for reconstructing acoustic radiation from arbitrarily shaped objects," J. Acoust. Soc. Am. **112**, 179–188 (2002).

[8]A. Sarkissian, "Extension of measurement surface in near-field acoustic holography," J. Acoust. Soc. Am. **115**, 1593–1596 (2004).

[9]A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography," J. Acoust. Soc. Am. **118**, 671–678 (2005).

[10]S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989).

[11]J.-H. Thomas and J.-C. Pascal, "Wavelet preprocessing for lessening truncation effects in nearfield acoustical holography," J. Acoust. Soc. Am. **118**, 851–860 (2005).

[12]K. Saijyou and S. Yoshikawa, "Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography," J. Acoust. Soc. Am. **110**, 2007–2023 (2001).

[13]E. G. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).

[14]E. G. Williams, B. H. Houston, and P. C. Herdic, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).

[15]K. Saijyou and H. Uchida, "Data extrapolation method for boundary element method-based near-field acoustical holography," J. Acoust. Soc. Am. **115**, 785–796 (2004).

[16]K. Saijyou, "Regularization method for the application of *K*-space data extrapolation to near-field acoustical holography," J. Acoust. Soc. Am. **116**, 396–404 (2004).

[17]M. Lee and J. S. Bolton, "Patch near-field acoustical holography in cylindrical geometry," J. Acoust. Soc. Am. **118**, 3721–3732 (2005).

[18]R. W. Gerchberg, "Super-resolution through energy reduction," Opt. Acta **21**, 709–720 (1974).

[19]A. Papoulis, "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **22**, 735–742 (1975).

[20]D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," IEEE Trans. Circuits Syst. **25**, 694–702 (1978).

[21]P. J. S. G. Ferreira, "Interpolation and the discrete Papoulis-Gerchberg algorithm," IEEE Trans. Signal Proc. **42**, 2596–2606 (1994).

[22]J. von Neumann, *Functional Operators, Vol. II: The Geometry of Orthogonal Spaces* (Princeton University Press, Princeton, 1950).

[23]D. C. Youla and H. Webb, "Image restoration by the method of convex projections. I. Theory," IEEE Trans. Med. Imaging **1**, 81–94 (1982).

[24]P. J. S. G. Ferreira, "Iterative and noniterative recovery of missing samples for 1-D bandlimited signals," Chap. 5 in *Nonuniform Sampling: Theory and Practice*, edited by F. A. Marvasti (Plenum, New York, 2001).

[25]M. H. Hayes and R. W. Schafer, "On the bandlimited extrapolation of discrete signals," Intl. Conf. Acoustics, Speech, and Signal Proc., Boston, **8**, 1450–1453 (1983).

[26]J. L. C. Sanz and T. S. Huang, "Some aspects of band-limited signal extrapolation: Models, discrete approximations, and noise," IEEE Trans. Acoust., Speech, Signal Proc. **31**, 1492–1501 (1983).

[27]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).

Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions

Radha Kalluri
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114 and Speech and Hearing figureBioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138

Christopher A. Shera[a)]
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114; Speech and Hearing Bioscience and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138; and Department of Otology & Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 15 August 2006; revised 29 December 2006; accepted 29 December 2006)

Otoacoustic emissions (OAEs) evoked by broadband clicks and by single tones are widely regarded as originating via different mechanisms within the cochlea. Whereas the properties of stimulus-frequency OAEs (SFOAEs) evoked by tones are consistent with an origin via linear mechanisms involving coherent wave scattering by preexisting perturbations in the mechanics, OAEs evoked by broadband clicks (CEOAEs) have been suggested to originate via nonlinear interactions among the different frequency components of the stimulus (e.g., intermodulation distortion). The experiments reported here test for bandwidth-dependent differences in mechanisms of OAE generation. Click-evoked and stimulus-frequency OAE input/output transfer functions were obtained and compared as a function of stimulus frequency and intensity. At low and moderate intensities human CEOAE and SFOAE transfer functions are nearly identical. When stimulus intensity is measured in "bandwidth-compensated" sound-pressure level (cSPL), CEOAE and SFOAE transfer functions have equivalent growth functions at fixed frequency and equivalent spectral characteristics at fixed intensity. This equivalence suggests that CEOAEs and SFOAEs are generated by the same mechanism. Although CEOAEs and SFOAEs are known by different names because of the different stimuli used to evoke them, the two OAE "types" are evidently best understood as members of the same emission family. [DOI: 10.1121/1.2435981]

PACS number(s): 43.64.Jb, 43.64.Bt, 43.64.Kc, 43.58.Ry [BLM] Pages: 2097–2110

I. INTRODUCTION

The stimuli used to evoke otoacoustic emissions (OAEs) range from the spectrally dense (e.g., the broadband clicks used to elicit click-evoked OAEs) to the spectrally sparse (e.g., the single pure tones used to evoke stimulus-frequency OAEs). Although linear reflection models of OAE generation (e.g., Zweig and Shera, 1995; Talmadge *et al.*, 1998) predict that both click-evoked and stimulus-frequency otoacoustic emissions (CEOAEs and SFOAEs) originate via essentially linear mechanisms (i.e., wave reflection off preexisting mechanical perturbations), other models imply that differences in the spectrum of the evoking stimulus result in differences in the mechanisms of OAE generation. Nobili *et al.* (2003a), for example, use model simulations to argue that the mechanisms responsible for CEOAE generation are both inherently nonlinear and fundamentally different from those responsible for generating SFOAEs. In Nobili *et al.*'s simulations, CEOAEs result from spatially complex "residual oscillations" of the basilar membrane that trace their origin to spectral irregularities in middle-ear transmission (see also Nobili, 2000; Nobili *et al.*, 2003b). Based on OAE measurements in guinea pig, Yates and Withnell (1999b) also posit a distinction between OAEs evoked by narrow- and broadband stimuli. They argue that although SFOAEs may originate from the independent "emission channels" predicted by linear reflection models, CEOAEs are essentially broadband distortion-product emissions (broadband DPOAEs). In this view, CEOAEs arise not from independent channels but from intermodulation distortion sources induced as a consequence of nonlinear interactions among the multiple frequency components of the broadband click stimulus (see also Withnell and Yates, 1998; Yates and Withnell, 1999a; Carvalho *et al.*, 2003).

The work reported here was motivated by these basic disagreements about the influence of stimulus spectrum on mechanisms of OAE generation. Our goal was to determine the relationship between the OAEs evoked by stimuli with the most dissimilar temporal and spectral structure (i.e., CEOAEs and SFOAEs). Interpretation of the experiments assumes that differences in OAE spectral characteristics im-

a)Electronic mail: shera@epl.meei.harvard.edu

ply differences in OAE generating mechanisms, and conversely. Similar logic has been used to distinguish "reflection-" and "distortion-source" OAEs. Whereas reflection-source OAEs (e.g., SFOAEs at low levels) have a rapidly varying phase and a slowly varying amplitude occasionally punctuated by sharp notches, distortion source OAEs (e.g., DPOAEs evoked at fixed, near-optimal primary-frequency ratios) have an almost constant amplitude and phase. These differences in OAE spectral characteristics are taken as indicative of fundamental differences in their mechanisms of generation (e.g., Kemp and Brown, 1983; Shera and Guinan, 1999).

Despite its fundamental importance, only a handful of studies have addressed the comparison between CEOAEs and SFOAEs. Although Zwicker and Schloth (1984) measured tone- and click-evoked frequency responses in the same human subject, the uncertain reliability of their tone-evoked data precludes any compelling comparison. Unlike the tone-evoked responses observed in subsequent studies, the "synchronous-evoked" OAEs reported by Zwicker and Schloth appear inconsistent with an origin in a causal system (Shera and Zweig, 1993). Furthermore, the emission data for the two stimulus types are presented in different ways: Although the CEOAE data represent the emission alone, the tone-evoked data represent the combined pressure of the stimulus and the emission. In what appears to be the only other study to explicitly address the issue, Prieve *et al.* (1996) found that the OAEs evoked by clicks and by tone bursts have similar intensity dependence, consistent with a common mechanism of generation. Unfortunately, the bandwidths of their tone bursts were not all that narrow (they typically spanned an octave or more), and their data do not allow a comparison of spectral structure or phase.

The experiments reported here examine the effect of stimulus bandwidth on OAE generation mechanisms by measuring and appropriately comparing the emissions evoked by wideband clicks (CEOAEs) with those evoked by tones (SFOAEs). Comparisons are made across stimulus frequency and intensity in the same human subjects.

II. MEASUREMENT METHODS

Measurements were made in one ear of each of four ($n=4$) normal-hearing human subjects who were comfortably seated in a sound-isolated chamber. All procedures were approved by human studies committees at the Massachusetts Eye and Ear Infirmary and the Massachusetts Institute of Technology.

Stimuli were digitally generated and recorded using implementations of standard OAE measurement protocols on the Mimosa Acoustics measurement system. The measurement system consists of a DSP-board (CAC Bullet) installed in a laptop computer, an Etymōtic Research ER10c probe system, and two software programs—one for measuring CEOAEs (T2001 v3.1.3) and another for measuring SFOAEs (SF2003 v2.1.18).

Signals were delivered and recorded in the ear canal. In-the-ear calibrations were made before each measurement. Stimuli were digitally generated using a fixed sampling rate of 48 kHz and data buffer lengths of 4096 samples, resulting in a frequency resolution of approximately 12 Hz. Potential artifacts were detected in real time by computing the difference between the current data buffer and an artifact-free reference buffer. The current data buffer was discarded whenever the rms value of the difference waveform exceeded a subject-specific criterion. Accepted data buffers were added to the averaging buffer. Continual replacement of the reference buffer minimized the effects of slowly varying drifts in the baseline.

We briefly outline the procedures for measuring each type of OAE below. Interested readers can consult Mimosa Acoustics technical documentation for more detailed descriptions of the measurement system (see also Lapsley-Miller *et al.*, 2004a,b).

A. Measuring CEOAEs

CEOAEs were evoked using broadband clicks (0.5–5 kHz) ranging in intensity from 35 to 80 dB pSPL (peak-equivalent SPL). To enable comparisons with SFOAEs, which are evoked using iso-intensity pure tones, the click waveform was adjusted using the in-the-ear calibration data to produce a flat-spectrum microphone signal. Responses were averaged across 500–4000 repetitions, depending on the stimulus level. Noise floors for the measurements typically ranged from −25 to −33 dB SPL.

A typical response waveform is shown in Fig. 1. The large pulse is the acoustic click, the smaller, more temporally dispersed portion of the waveform is the CEOAE. CEOAEs were extracted from the ear-canal pressure waveform by using either the linear-windowing or the nonlinear-residual method. The following sections describe each method in turn. Our standard-protocol used a click repetition period T_a of approximately 26 ms (1253 samples). As a check for possible efferent effects (e.g., Guinan *et al.*, 2003), we varied the interstimulus time from roughly 20 ms up to 100 ms but found no significant dependence on repetition period.

1. The linear windowing method

In the linear windowing paradigm (e.g., Kemp, 1978) the stimulus and emission, $p_{ST}(t)$ and $p_{CE}(t)$, are extracted from the total ear-canal pressure, $p(t)$, by applying stimulus and emission windows, $w_s(t)$ and $w_e(t)$:

$$p_{ST}(t) = w_s(t)p(t), \tag{1}$$

$$p_{CE}(t) = w_e(t)p(t). \tag{2}$$

The stimulus and emission spectra are then computed by taking the 4096-point discrete Fourier transform $\mathcal{F}\{\cdot\}$ of zero-padded waveforms $p_{CE}(t)$ and $p_{ST}(t)$:

$$P_{ST}(f) = \mathcal{F}\{p_{ST}(t)\}, \tag{3}$$

$$P_{CE}(f) = \mathcal{F}\{p_{CE}(t)\}. \tag{4}$$

The input/output CEOAE transfer function $T_{CE}(f;A)$ is defined as the ratio of the two spectra:

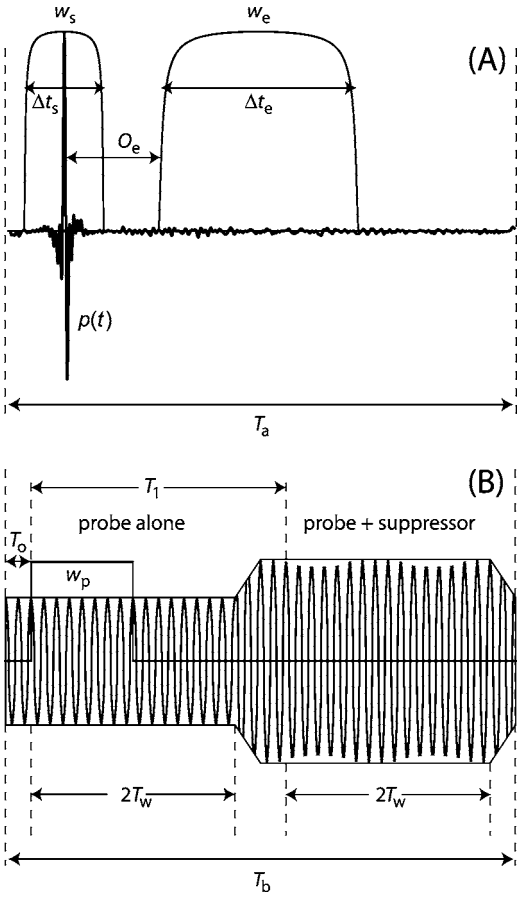


FIG. 1. Schematic diagrams of the measurement paradigms. (A) For CEOAEs, the stimulus and emission are measured using the linear windowing technique by applying recursive-exponential windows w_s and w_e to the ear-canal pressure, $p(t)$. The windows' center positions, t_s and t_e, and widths, Δt_s and Δt_e, are chosen to optimize the separation between the stimulus and emission. (B) For SFOAEs, the stimulus and emission are measured using the interleaved suppression technique. The emission is computed as the Fourier component at the probe frequency by taking the complex difference between the probe-alone and probe-suppressor segments of the ear-canal pressure. The probe-alone and probe-suppressor waveforms are extracted using rectangular windows $w_p^{(n)}$ and $w_{ps}^{(n)}$; only $w_p^{(0)}$ is shown in the figure. The Fourier analysis buffer (duration T_w) contains an integral number of cycles of both probe and suppressor.

$$T_{CE}(f;A) = \frac{P_{CE}(f;A)}{P_{ST}(f)}, \tag{5}$$

where $A \equiv |P_{ST}|$ is the stimulus amplitude. Although we refer to the ratio as a transfer function, $T_{CE}(f;A)$ depends on the stimulus amplitude and is therefore more correctly known as a "describing" function (e.g., Krylov and Bogolyubov, 1947; Gelb and Vander Velde, 1968).

For the windowing technique to work, the stimulus click must be sufficiently localized in time so that the end of the stimulus does not significantly overlap with the early components of the emission. Unless otherwise noted, the clicks used in these experiments were bandlimited from 0.5 to 5 kHz—the broadest flat spectrum click without notches that the measurement system was able to generate. Interference between stimulus and emission can be further reduced by the proper choice of windows. We used tenth-order recursive-exponential windows $w_{rex}(t;\Delta t)$ (Shera and Zweig, 1993;

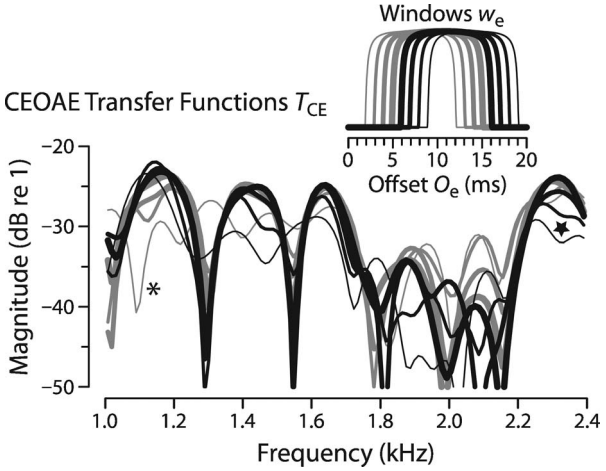


FIG. 2. Dependence of CEOAE spectra on window offset, O_e. The inset illustrates the windows $w_e(t-t_e;\Delta t_e)$ with poststimulus offsets ranging from 2 to 8 ms. The spectral structure of the CEOAE transfer function varies with window offset. For window offsets between 5 and 7 ms, CEOAE transfer functions are almost independent of O_e (thick lines). For shorter offsets (<5 ms) the transfer functions manifest additional spectral structure (*), presumably due to interference-like interactions between the stimulus and the emission. For offsets greater than 7 ms the short latency, high-frequency components of the CEOAE degrade (⋆).

Kalluri and Shera, 2001) with time offsets and widths chosen to reduce interactions between the stimulus and emission. Thus,

$$w_s(t) = w_{rex}(t - t_s;\Delta t_s), \tag{6}$$

$$w_e(t) = w_{rex}(t - t_e;\Delta t_e), \tag{7}$$

with standard offsets $\{t_s, t_e\}=\{0, 10\}$ ms and widths $\{\Delta t_s, \Delta t_e\}=\{5, 10\}$ ms. All offsets are relative to the center of the stimulus click at $t=0$. The recursive-exponential window is defined in footnote 10 of Kalluri and Shera (2001).

The location of the emission analysis window must be carefully chosen to reduce interference caused by interactions between the stimulus and the early components of the emission. The window $w_e(t)$ begins at time $O_e=t_e-\Delta t_e/2$ after the click (see Fig. 1). To determine the "optimal" window offset, we varied O_e until small shifts had negligible effects on the transfer function within the frequency range of interest (1–4 kHz). Offsets smaller than about 5 ms or larger than 7 ms produced significant changes in the magnitude of the transfer function (see Fig. 2). Except where noted, we adopted the value $O_e=5$ ms for all the results shown here. Because CEOAEs are dispersed in time, with high frequency components arriving before the low frequency components, the optimal window for the 1–4-kHz region will not be optimal for emissions in other frequency bands.

2. The nonlinear residual method

The nonlinear residual method is an alternate and generally more popular procedure for measuring CEOAEs. In this method CEOAEs are extracted by exploiting the nonlinear compressive growth of the emissions in conjunction with the linear growth of the stimulus. Three identical clicks are

followed by a fourth click that is three times larger but of the opposite polarity. The CEOAE estimate is the average of the four responses.

Unlike the linear windowing technique, in which short-latency components of the emission are typically eliminated, the nonlinear residual method separates the emission from the stimulus without removing the early arriving components of the emission. However, to avoid confusion, reduce potential artifacts due to system distortion, and enable direct comparison between the two CEOAE techniques, we apply the standard emission window, $w_e(t)$, to the nonlinear-derived emission as well. Therefore, just as in the linear technique, early arriving components of the emission are eliminated.

B. Measuring SFOAEs

We measured the SFOAE pressure, $P_{SF}(f)$, using a variant of the suppression method (Shera and Guinan, 1999; Kalluri and Shera, 2001). As illustrated in Fig. 1, the emission is obtained as the complex difference between the ear-canal pressure at the probe frequency (f) measured first with the probe tone alone and then in the presence of a more intense (55 dB SPL) suppressor tone at a nearby frequency, f_s, roughly 47 Hz below the probe frequency (Fig. 1). The suppressor was presented in interleaved time segments to minimize possible artifactual contamination from time-varying drifts in the base signal. To reduce spurious contamination by earphone distortion, the probe and suppressor were generated using separate sounds sources.

The probe-alone waveform, $p_p(t)$, and probe-suppressor waveform, $p_{ps}(t)$, are obtained from the measured ear-canal pressure, $p(t)$, by averaging over two subsegments extracted using windows $w_p(t)$ and $w_{ps}(t)$:

$$p_p(t) = \frac{1}{2}\sum_{n=0}^{1} w_p^{(n)}(t)p(t), \tag{8}$$

$$p_{ps}(t) = \frac{1}{2}\sum_{n=0}^{1} w_{ps}^{(n)}(t)p(t). \tag{9}$$

In this case, the windows are rectangular boxcars of width T_w:

$$w_p^{(n)}(t) = w_{box}(t - T_0 - nT_w; T_w); \tag{10}$$

$$w_{ps}^{(n)}(t) = w_{box}(t - T_0 - T_1 - nT_w; T_w), \tag{11}$$

where

$$w_{box}(t;\Delta t) = \begin{cases} 1 & \text{for } 0 \le t \le \Delta t, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The window offsets T_0 and T_1 were chosen to allow the system time to return to steady state after switching the suppressor tone on or off. The window duration T_w equals that of the Fourier analysis buffer. Stimulus frequencies were chosen so that the analysis buffer (of duration $T_w = N\Delta t$, where N is the buffer size and Δt is the reciprocal of the sampling rate) always contained an integral number of cycles of both probe and suppressor. For the measurements reported here $\{T_0, T_1, T_b\} = \{\frac{1}{4}, \frac{5}{2}, 5\}T_w$. The SFOAE pressure is computed as

$$P_{SF}(f) = \mathcal{F}\{p_p(t)\} - \mathcal{F}\{p_{ps}(t)\}e^{i2\pi f T_1}, \tag{13}$$

where $\mathcal{F}\{\cdot\}$ indicates the 4096-point discrete Fourier transform at the probe frequency, f. The stimulus pressure is extracted from the probe-suppressor segment:

$$P_{ST}(f) = \mathcal{F}\{p_{ps}(t)\}e^{i2\pi f T_1}. \tag{14}$$

By analogy with Eq. (5) for $T_{CE}(f;A)$, the transfer function $T_{SF}(f;A)$ is defined as the ratio of probe-frequency spectral components

$$T_{SF}(f;A) = \frac{P_{SF}(f;A)}{P_{ST}(f)}, \tag{15}$$

where we have now explicitly indicated the dependence on stimulus amplitude ($A \equiv |P_{ST}|$). We measured $T_{SF}(f;A)$ with a frequency resolution of approximately 23 Hz using probe-tone levels ranging from approximately 10 to 40 dB SPL. We typically employed 32 averages at the highest probe level and 128 averages at the lowest.

III. EXPERIMENTAL COMPLICATIONS

Before describing our main results, we first address two measurement issues that complicate the comparison between $T_{CE}(f;A)$ and $T_{SF}(f;A)$. The first pertains to differences between the two different CEOAE measurement methods. The second deals with complications arising from synchronized spontaneous otoacoustic emissions (SSOAE).

A. CEOAE transfer functions from linear and nonlinear methods

Figure 3 compares the CEOAE transfer functions $T_{CE}(f;A)$ measured using the linear-windowing and nonlinear-residual methods. We denote transfer functions measured using the two methods by $T_{CE}(f;A)$ and $T_{CE}^{NL}(f;A)$, respectively. For brevity, we show measurements from one subject; similar results were obtained in all.

Although both the linear-windowing and nonlinear-residual techniques yield qualitatively similar values of $T_{CE}(f;A)$ at high stimulus levels, CEOAEs at low levels can only be extracted using the linear technique. As stimulus levels are decreased from 80 to 60 dB pSPL, the magnitudes of both $T_{CE}(f;A)$ and $T_{CE}^{NL}(f;A)$ increase. At these levels $T_{CE}(f;A)$ and $T_{CE}^{NL}(f;A)$ have similar peaks, notches, and phase behaviors. This similarity in behavior does not carry through to the lowest levels. As stimulus levels are further reduced, the magnitude of $T_{CE}(f;A)$ continues to grow and eventually becomes nearly independent of level. By contrast, $T_{CE}^{NL}(f;A)$ reaches a maximum value and then falls quickly into the noise floor. The combination of results—near level independence of $T_{CE}(f;A)$ and the rapid fall of $T_{CE}^{NL}(f;A)$ at low stimulus levels—suggests that CEOAEs grow almost linearly at the lowest stimulus levels. Note, however, that by using the nonlinear-derived method, Withnell and McKinley (2005) found short-latency CEOAE components in guinea pigs that appear to result from nonlinear mechanisms within the cochlea. When measured using the linear-windowing

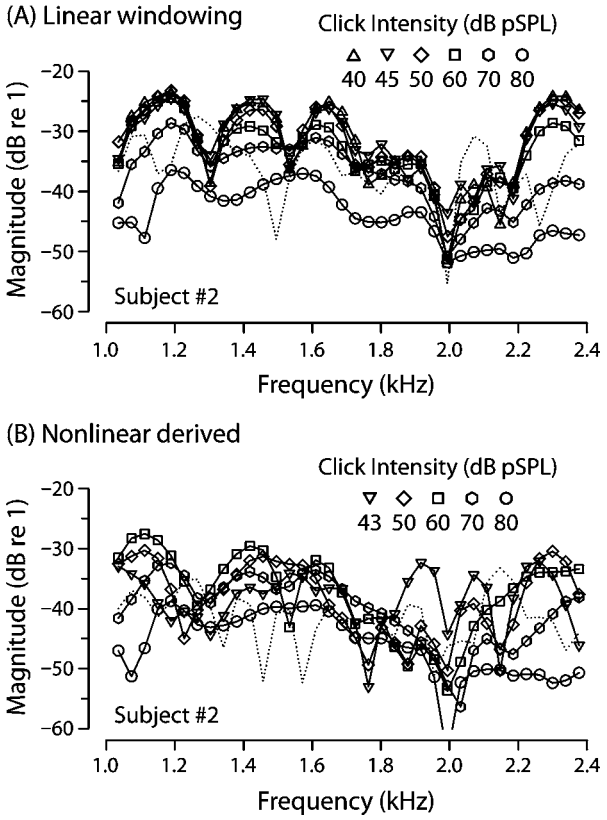


FIG. 3. Linear-windowed and nonlinear-derived CEOAE transfer functions. Panels (A) and (B) show $T_{\mathrm{CE}}(f;A)$ and $T_{\mathrm{CE}}^{\mathrm{NL}}(f;A)$, respectively, at click intensities ranging from 40 to 80 dB pSPL. The two techniques yield qualitatively similar results at high stimulus levels (70–80 dB pSPL). At lower intensities, however, they diverge: Whereas the nonlinear-derived $T_{\mathrm{CE}}^{\mathrm{NL}}(f;A)$ ultimately falls into the noise, the linear-windowed $T_{\mathrm{CE}}(f;A)$ continues to increase until it becomes independent of intensity. The transfer-function noise floors shown here (dotted lines) were measured at the lowest click intensities. Because they are scaled by the stimulus spectrum, transfer-function noise floors are much lower at higher stimulus levels.

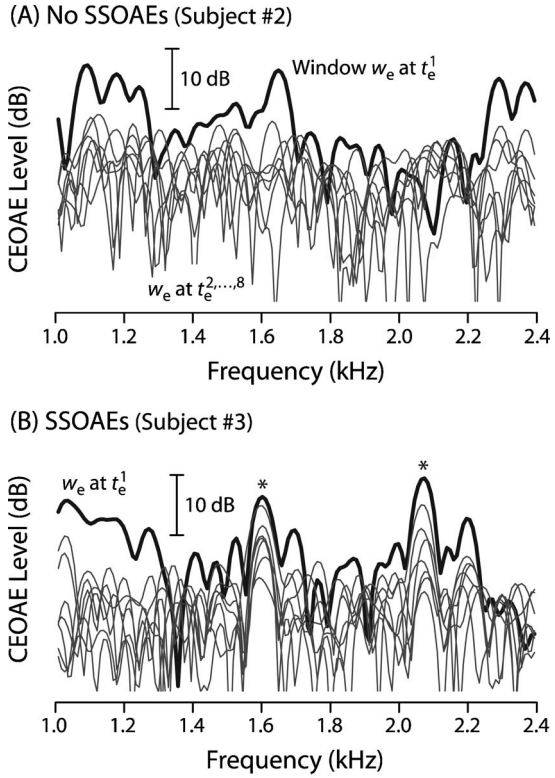


FIG. 4. Sustained activity in the CEOAE spectrum identifies synchronized SOAEs. SSOAEs were identified using the linear-windowing technique with an interstimulus time of 100 ms. Eight 20-ms analysis windows (with center offsets t_e^1 through t_e^8 ranging from 15 to 85 ms after the stimulus) were applied to the measured ear-canal pressure. Panel (A) shows the corresponding CEOAE spectra in a subject without measurable SSOAEs. Only the spectrum of the response during the first window (t_e^1, thick line) contains significant emission energy; responses from all subsequent windows (t_e^2 through t_e^8; thin lines) are small by comparison. Panel (B) shows the spectra in a subject with SSOAEs. A response at SSOAE frequencies appears as a spectral peak in all windows (*).

protocol, these short latency components would typically be obscured by the stimulus. Since our measurements had a residual short-latency stimulus artifact due to earphone nonlinearities (e.g., Kapadia *et al.*, 2005), we cannot rule out the possibility that human CEOAEs also contain small short-latency nonlinear components buried beneath the stimulus artifact. Because the nonlinear technique cannot be used to measure $T_{\mathrm{CE}}(f;A)$ at the lowest stimulus levels, all subsequent CEOAE measurements presented in this paper were made using the linear protocol.

B. Synchronized spontaneous otoacoustic emissions

Some of our subjects had synchronized spontaneous otoacoustic emissions (SSOAEs). SSOAEs are long-lasting transient responses that are not always identifiable by conventional SOAE searches, in which no external stimulus is presented. SSOAEs can, however, be detected when they are evoked by or synchronized to an applied stimulus, in this case the click used to evoke CEOAEs.

We measured SSOAEs using a variant of the standard linear-windowing technique for measuring CEOAEs. The variant employed an interclick time of 100 ms rather than the standard 20 ms used in the CEOAE measurements. To detect SSOAEs we then computed and compared the response spectra, $P_{\mathrm{CE}}^i(f)$, in eight partially overlapping analysis windows centered at different post-stimulus times ($t_e^i = 5 + 10i$ ms):

$$P_{\mathrm{CE}}^i(f) = \mathcal{F}\{p(t)w(t - t_e^i;\Delta t)\}, \quad i = 1,2,\ldots,8, \tag{16}$$

where the nominal window duration Δt is 20 ms. Figure 4 shows the spectra for the eight windowed segments in two subjects. The dark thick line gives the spectrum of the response during the first window, centered at 15 ms after the click. The narrow lines show the spectra measured during subsequent windows. In two of the four subjects, significant response energy occurs only within the first 20 ms, and we considered these subjects to have unmeasurable SSOAEs. In the remaining two subjects, some peaks in the spectrum [e.g., those identified by asterisks (*) in Fig. 4] disappear more slowly over the eight windows. We identified these long-lasting transient responses to the click as SSOAEs.

The existence of SSOAEs complicates the measurement of $T_{\mathrm{CE}}(f;A)$, and to a lesser extent $T_{\mathrm{SF}}(f;A)$, in at least two ways. First, SSOAEs make it more difficult to determine the stimulus spectrum [i.e., the denominator in Eq. (5)]. In sub-

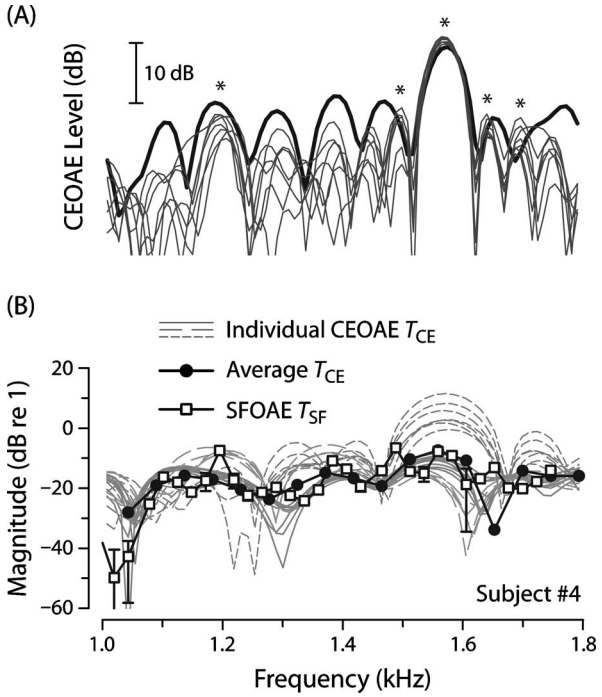


FIG. 5. Averaging out variability due to SSOAEs. Panel (A) shows the results of the SSOAE identification procedure (see Fig. 4) in a subject with strong SSOAEs. The frequencies of significant SSOAEs are marked by asterisks (*). Panel (B) shows the magnitude of individual CEOAE transfer functions (thin gray lines) made at 40 dB pSPL. Gray lines of different styles show transfer functions measured during different days. The solid circles (•) give the magnitude of the complex ensemble average of the multiple measurements. The open squares (□) show the SFOAE transfer function at a comparable stimulus level in the low-level linear regime (20 dB SPL).

jects with long-lasting SSOAEs, the response in the stimulus window is contaminated by responses that have not fully decayed by the time the next stimulus presentation occurs. Contamination by SSOAEs typically creates spurious ripples in the measured stimulus spectrum. To reduce errors in the computation of the transfer function, we estimate the stimulus spectrum at low stimulus levels (i.e., at 40–70 dB pSPL) by appropriately rescaling the stimulus spectrum measured at high levels (i.e., at 80 dB pSPL). This rescaling procedure reduces the error because the relative influence of SSOAE ripples is smallest at high levels.

Second, SSOAEs increase variability in the measurements, as shown in Fig. 5. The gray lines in the bottom panel show individual measurements of $T_{CE}(f;A)$ made during several sessions on two different days. Note the increased variability near SSOAE frequencies, indicated by asterisks in the top panel. The variability presumably reflects an instability in the relative phase with which the stimulus initiates or synchronizes the SSOAE. For example, sometimes the SSOAE seems to add to the CEOAE; at other times it appears to subtract. To reduce this variability in subjects with SSOAE we calculate the (complex) average of measurements made during multiple sessions at each frequency (black circle in the figure) before comparing with measurements of $T_{SF}(f;A)$. We find that matches between $T_{CE}(f;A)$ and $T_{SF}(f;A)$ are generally improved significantly by this ensemble averaging.

IV. COMPARISON OF SFOAE AND CEOAE TRANSFER FUNCTIONS

Figure 6 shows measurements of $T_{CE}(f;A)$ and $T_{SF}(f;A)$ versus frequency in a subject lacking the complications introduced by the existence of SSOAEs. Error bars on the magnitude represent the standard deviation of the mean.[1] The figure demonstrates that $T_{CE}(f;A)$ and $T_{SF}(f;A)$ have qualitatively similar spectral structure, including a rapidly varying phase and magnitude peaks and notches that occur at approximately the same frequencies in both transfer functions. Both transfer functions also share a qualitatively similar dependence on stimulus intensity. At the lowest levels, transfer function magnitudes appear nearly independent of level, consistent with a region of approximate linearity near threshold. At higher intensities, the transfer-function magnitudes generally decrease, consistent with compressive nonlinear growth in emission amplitudes. Although the strong qualitative similarity between the CEOAE and SFOAE transfer functions suggests that clicks and tones evoke emissions via similar mechanisms, definitive conclusions require more careful comparisons.

A. Matching click and tone intensities

At a minimum, the comparisons need to take into account that both CEOAEs and SFOAEs depend on stimulus intensity. Even if the responses are comparable in principle, comparisons made at different effective intensities may amount to comparing apples and oranges. Complicating the situation is the fact that click and tone intensities are conventionally specified in different ways. Whereas pure-tone intensities are measured in SPL, click intensities are measured in peak-equivalent SPL (pSPL), defined as the SPL of a pure tone with the same peak pressure as the click waveform. At what click intensity (in dB pSPL) should one measure CEOAEs in order best to compare them with SFOAEs measured at a given probe level (in dB SPL)? Whether or not this question has a meaningful answer depends on the nature of the nonlinearities involved in OAE generation.

To address this issue we investigated the dependence of CEOAE transfer functions on stimulus intensity and bandwidth. The black symbols in Fig. 7 show the magnitude of $T_{CE}(f_0;A)$ in a narrow frequency range (near $f_0 \cong 1.2$ kHz) measured using clicks of various intensities (pSPL) and bandwidths (data from subject 2, also without SSOAEs). Although $T_{CE}(f_0;A)$ appears nearly independent of intensity and bandwidth at the lowest intensities ($A < 50$ dB pSPL), a systematic dependence on both emerges at higher levels ($A > 70$ dB pSPL). Overall, the trends in the CEOAE data appear well described by a function with the approximate form

$$T_{CE} = T_{CE}(f_0;A_{pk},\mathrm{BW}) \cong T_0 \Big/ \left(1 + \frac{A_{pk}/A_0}{1+\mathrm{BW}/\Delta F}\right)^{\alpha}, \qquad (17)$$

where A_{pk} and BW are, respectively, the peak-equivalent stimulus pressure[2] and equivalent rectangular bandwidth of the stimulus.[3] The corresponding reference values, A_0 and ΔF, have units of pressure and frequency, respectively; the dimensionless constant T_0 sets the overall scale, and the exponent α determines asymptotic growth rates.

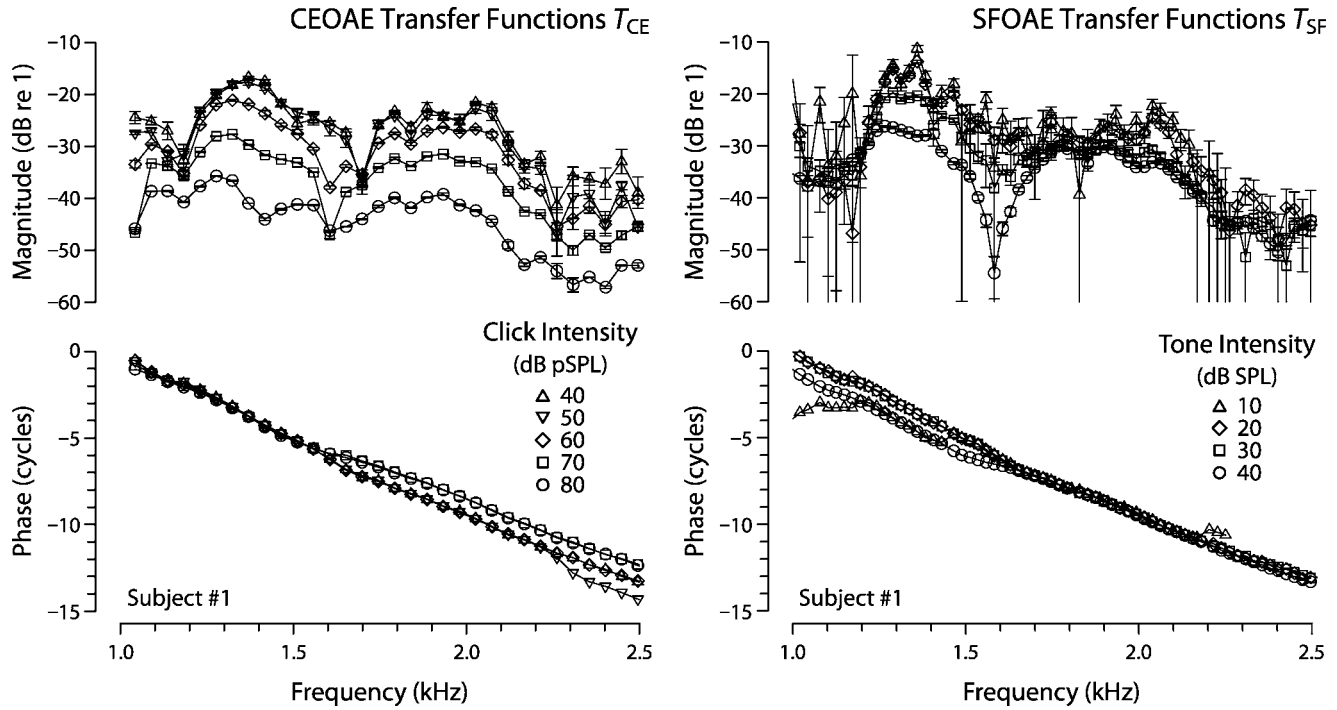


FIG. 6. CEOAE and SFOAE transfer functions.The two columns show the magnitude (top) and phase (bottom) of $T_{CE}(f;A)$ and $T_{SF}(f;A)$, respectively, for a subject without SSOAEs. The symbols identify the stimulus intensity, which ranged from 40 to 80 dB pSPL for clicks and from 10 to 40 dB SPL for tones. Error bars on the magnitude represent the standard error of the mean.

At large stimulus bandwidths and high intensities, the transfer function given by Eq. (17) increases with bandwidth (by about 6α dB/oct at fixed intensity) and decreases with intensity (at a rate approaching $-\alpha$ dB/dB at fixed bandwidth). In the opposite limits, Eq. (17) becomes independent of bandwidth when $\mathrm{BW} \ll \Delta F$ and independent of intensity when $A_{pk} \ll A_0$.

Equation (17) can be simplified for analysis and plotting by rewriting it as

$$T_{CE} = T_{CE}(f_0;A_{eff}) \cong \frac{T_0}{(1 + A_{eff}/A_0)^{\alpha}}, \tag{18}$$

a form obtained by combining the intensity and bandwidth dependence into an effective stimulus pressure, A_{eff}, defined by

$$A_{eff}(A_{pk}, r) \equiv A_{pk}/(1 + r), \tag{19}$$

where $r \equiv \mathrm{BW}/\Delta F$. To the extent that Eq. (17) approximates the data, Eq. (18) predicts that CEOAE transfer functions measured using different stimulus bandwidths can be made to fall along a single curve if plotted against the variable

$$L_{bwc} \equiv L_{pk} - 20 \log (1 + r), \tag{20}$$

where $L_{pk} = 20 \log (A_{pk}/20\ \mu\mathrm{Pa})$ is the click intensity in peak-equivalent SPL (pSPL) and L_{bwc} is what we call "bandwidth-compensated" sound-pressure level (cSPL). The black symbols in Fig. 8 show the CEOAE data replotted versus L_{bwc} with $\Delta F \cong 74$ Hz; as predicted, the data fall approximately along a single curve.

Equation (18) can be used to extrapolate the CEOAE growth function to stimulus bandwidths other than those measured in Fig. 7. As an extreme case, the equation can be used to predict the intensity dependence of SFOAE transfer functions by finding the limit $r \to 0$. This extrapolation regards the SFOAE stimulus as a "click" in which the bandwidth has been reduced so much that the stimulus comprises nothing but a single pure tone. If the extrapolation remains valid across this large reduction in stimulus bandwidth, then

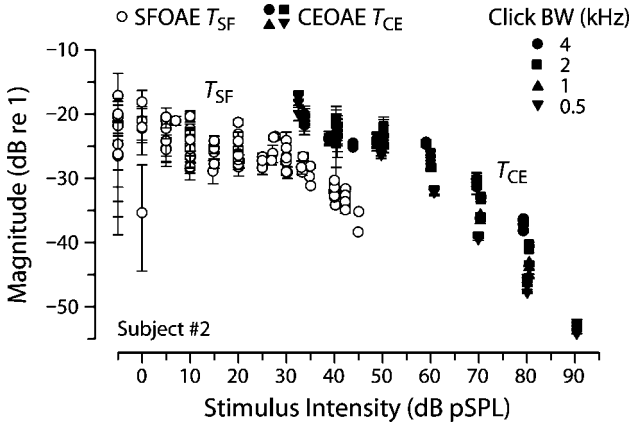


FIG. 7. Intensity and bandwidth dependence of CEOAE and SFOAE transfer functions. Black symbols show the intensity dependence of $T_{CE}(f_0;A)$ measured using clicks with nominal bandwidths of $\{4,2,1,0.5\}$ kHz, corresponding to spectral edges of $\{1-5,1-3,1-2,1-1.5\}$ kHz, respectively. Equivalent rectangular bandwidths are given in footnote 3. For comparison, filled symbols show the intensity dependence of the SFOAE transfer function, $T_{SF}(f_0;A)$. Both CEOAE and SFOAE data were measured at four frequencies in a narrow band near a peak in $T_{CE}(f;A)$ magnitude ($f_0 \in [1.15,1.2]$ kHz). Multiple points at the same intensity and bandwidth represent values at the four different f_0 frequencies in the range. All stimulus intensities are expressed in peak-equivalent SPL (pSPL) or its equivalent (SPL) for the pure tones used to measure SFOAEs.

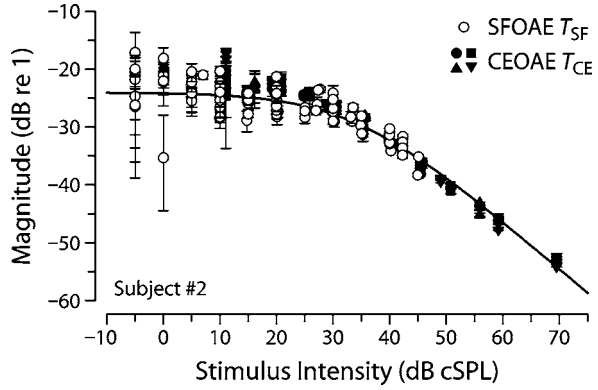


FIG. 8. Unification of CEOAE and SFOAE growth functions. CEOAE transfer functions $T_{CE}(f_0;A)$ from Fig. 7 are shown here with black symbols; SFOAE transfer functions $T_{SF}(f_0;A)$ are shown using open circles. Stimulus levels are expressed in "bandwidth-compensated" sound-pressure level (dB cSPL), defined by Eq. (20). The solid line shows Eq. (17) evaluated with the best-fit parameters given in the text.

$$T_{SF}(f_0;A) \cong \lim_{r\to 0} T_{CE}(f_0;A_{eff}) = \frac{T_0}{(1+A/A_0)^{\alpha}}, \tag{21}$$

where A (the rms pressure of the SFOAE stimulus) and A_{eff} are equivalent for pure tones. In the pure-tone limit, $L_{bwc}\to L_{pk}\to L$, where L is the stimulus sound-pressure level (in dB SPL). The extrapolation thus predicts that SFOAE growth function data lie along the same unified curve found for CEOAEs.

Figure 8 demonstrates that, remarkably, the extrapolation from click to pure-tone bandwidths appears valid: The single growth function shown to characterize the intensity dependence of $T_{CE}(f;A)$ applies also to SFOAEs at the same frequency. Although the measurements of $T_{SF}(f;A)$ are limited to stimulus intensities less than 45 dB SPL, the agreement between the two growth functions (plotted versus cSPL) is excellent throughout the measured range. Although variability in the measurements increases at the lowest intensities (especially for SFOAEs), the common growth function manifests a region of approximately linear growth below about 10–15 dB cSPL. The growth function then gradually transitions into a compressive regime whose slope is somewhat smaller than −1 dB/dB at higher intensities. Fitting Eq. (17) to the pooled CEOAE and SFOAE data yields the parameter estimates $T_0 \cong -24\pm 2$ dB *re* 1, $A_0 \cong 900\pm 400$ μPa (or 33 ± 4 dB SPL), $\Delta F \cong 74\pm 20$ Hz, and $\alpha \cong 0.8\pm 0.08$. (Essentially the same values are obtained when the CEOAE data are fit independently.) The uncertainties represent approximate 95% confidence intervals obtained by bootstrap resampling and are not independent of one another. The solid line in Fig. 8 shows Eq. (17) evaluated using the best-fit parameters.

B. Comparisons at matched stimulus intensities

Figure 8 suggests that CEOAE and SFOAE transfer functions might best be compared at stimulus intensities matched by expressing them in bandwidth-compensated SPL. If the unification achieved in Fig. 8 generalizes across frequency and subject, the transfer functions will then have very similar magnitudes. We test this suggestion in Fig. 9, which compares CEOAE and SFOAE transfer functions across frequency at two matched stimulus intensities in two subjects. The lower of the two intensities (20 dB cSPL) falls within or just above the low-level linear regime; the higher intensity (40 dB cSPL) evokes responses in the region of compressive OAE growth. The two columns shows transfer functions from different subjects, neither of whom had identifiable SSOAEs in the measured frequency range (1–2.4 kHz).

Figure 9 demonstrates that the magnitudes and phases of the transfer functions $T_{CE}(f;A)$ and $T_{SF}(f;A)$ are almost identical at matched intensities. The agreement extends even to the spectral notches, regions where one might expect the responses to be especially sensitive to small changes. Since details of the phase are obscured by the large delay and the phase unwrapping, Fig. 10 replots the transfer-function phases after subtracting out smooth trend lines that capture the secular variation of the phase. The resulting comparisons show that the agreement between $\angle T_{CE}(f;A)$ and $\angle T_{SF}(f;A)$ is generally excellent, even in microstructural detail.

Figure 11 extends the comparison to subjects with SSOAE. In these subjects, CEOAE transfer functions were obtained by averaging across measurement sessions, as described in Sec. III B. Although small differences between $T_{CE}(f;A)$ and $T_{SF}(f;A)$ are found, especially at the lower intensity [e.g., in subject 4, where sharp peaks in $|T_{SF}(f;A)|$ can be seen at SSOAE frequencies], the overall agreement is still excellent.

1. Discrepancies and their possible origin

Despite the compelling overall agreement between $T_{CE}(f;A)$ and $T_{SF}(f;A)$ at matched intensities, the two transfer functions are not always identical to within the measurement error. For example, in subject 2 at 40 dB cSPL (Fig. 9, right-middle panel), the spectral peaks of $T_{CE}(f;A)$ near approximately 1.2 and 1.4 kHz occur at slightly higher frequencies than do the corresponding peaks of $T_{SF}(f;A)$. Do these small differences result from normal session-to-session variability in the emission measurements? Or from methodological differences in the measurement of the two emission types? Or might the discrepancies be more interesting, perhaps reflecting some (presumably subtle) difference in the mechanisms of emission generation?

To characterize the session-to-session variability, we remeasured the two transfer functions multiple times in 1 day (removing, reinserting, and recalibrating the measurement probe each time) and on 9 different days over a 3-month period in a single subject (subject 2). To make the relatively large number of required measurements feasible, we limited the measurements of $T_{SF}(f;A)$ to five frequency points between 1 and 1.2 kHz. In this subject the variability of the measurements across different days was not significantly different from the variability of the measurements made within 1 day.[4] The solid curve with flanking gray region in Fig. 12 shows the mean of 62 measurements of $T_{CE}(f;A)$; the open circles show the mean and deviation of 30 measurements of $T_{SF}(f;A)$. The gray shaded area and error bars extend one standard deviation above and below the mean. The results show that the spectral offset between peaks of $T_{CE}(f;A)$ and

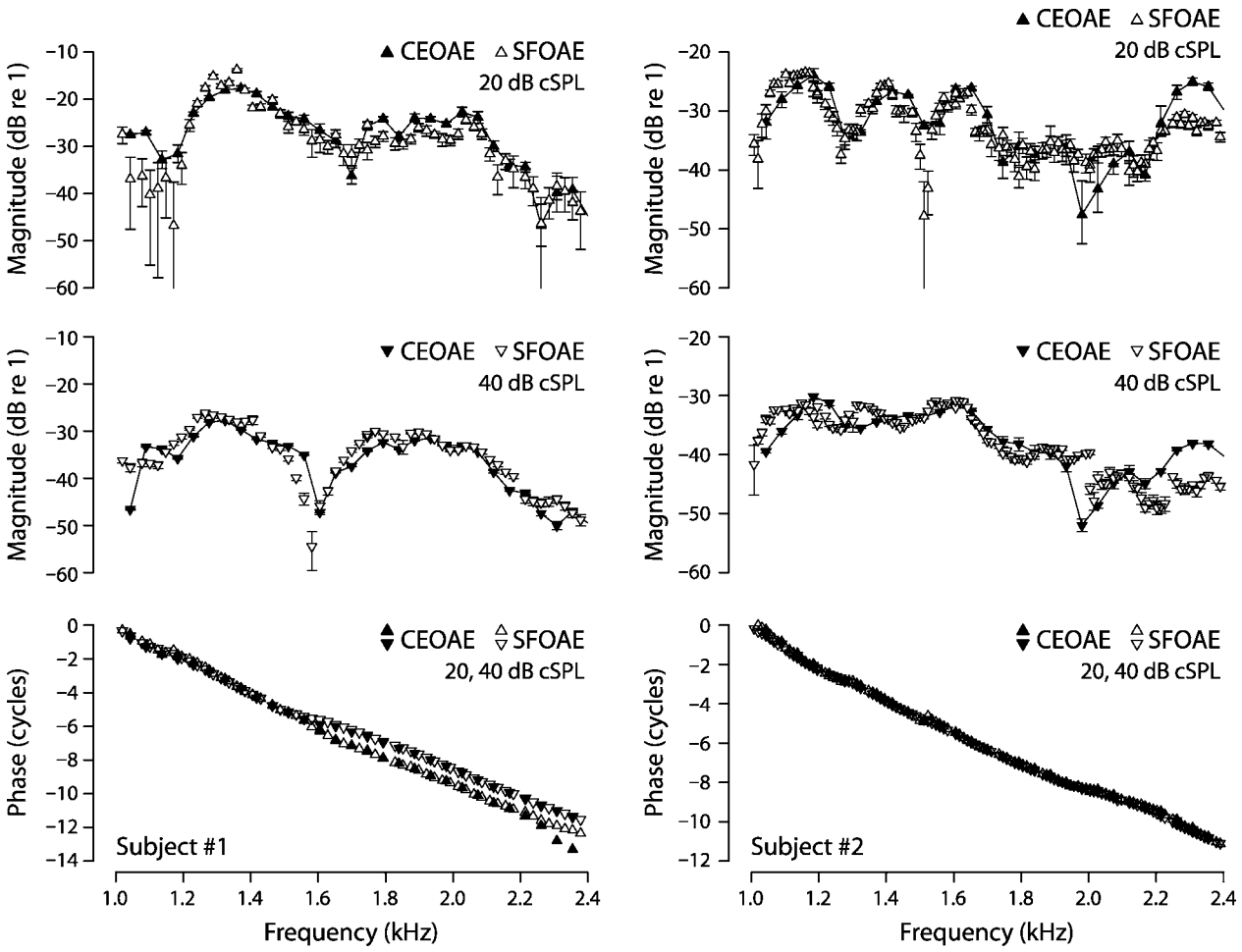


FIG. 9. CEOAE and SFOAE transfer functions at matched intensities in subjects without SSOAEs. Columns show measurements in two different subjects. The top row shows transfer-function magnitudes measured at 20 dB cSPL; the middle row shows magnitudes at 40 dB cSPL. The bottom row shows unwrapped phases at both intensities. In all panels, $T_{CE}(f;A)$ and $T_{SF}(f;A)$ are shown with filled and open symbols, respectively.

$T_{SF}(f;A)$ is larger than the measurement variability, indicating a statistically significant difference between the two measurements.

We doubt, however, that these discrepancies provide evidence for differences in the underlying mechanism of emission generation. Rather, we suspect that the differences are largely if not entirely methodological. As demonstrated in Fig. 2, overlap between the ringing portion of the stimulus and the CEOAE can affect details of the $T_{CE}(f;A)$ spectral shape. The solid curve with hatch marks in Fig. 12 illustrates how a small (ΔO_e=0.5 ms) increase in the offset of the OAE analysis window changes the value of $T_{CE}(f;A)$ extracted from the same measured waveforms. Although larger window offsets reduce interference artifacts due to ringing of the stimulus, they do so at the cost of eliminating short-latency components of the emission. Comparing the values of $T_{CE}(f;A)$ obtained from the same time waveforms using two different window offsets (O_e=5 and 5.5 ms) demonstrates that relatively small changes in the CEOAE measurement paradigm can produce variations in $T_{CE}(f;A)$ comparable to those observed between $T_{CE}(f;A)$ and $T_{SF}(f;A)$. Note, for example, how the peaks in $T_{CE}(f;A)$ near 1.2 and 1.4 kHz shift toward slightly lower frequencies when using the 5.5-ms offset, resulting in an improved match between the $T_{CE}(f;A)$ and $T_{SF}(f;A)$ at these frequencies.

V. DISCUSSION

At low and moderate stimulus intensities human CEOAE and SFOAE input/output transfer functions are nearly identical. When stimulus intensity is measured in bandwidth-compensated SPL (cSPL), we found that CEOAE and SFOAE transfer functions have equivalent growth functions at fixed frequency and equivalent spectral characteristics at fixed intensity. This strong similarity suggests that the OAEs evoked by broad- and narrow-band stimuli (clicks and tones) are generated by the same mechanism.

A. Possible limits of application

Although our conclusions may apply more widely, we summarize below the known limitations of our study. (1) Our comparisons between CEOAEs and SFOAEs were time consuming, and were therefore performed in a relatively small number of subjects (n=4). Nevertheless, since the subjects were selected only for having measurable emissions, and because we found similar results in all, it seems unlikely that the near equivalence we report is merely a statistical fluke. (2) All of our subjects had normal hearing. Although addi-

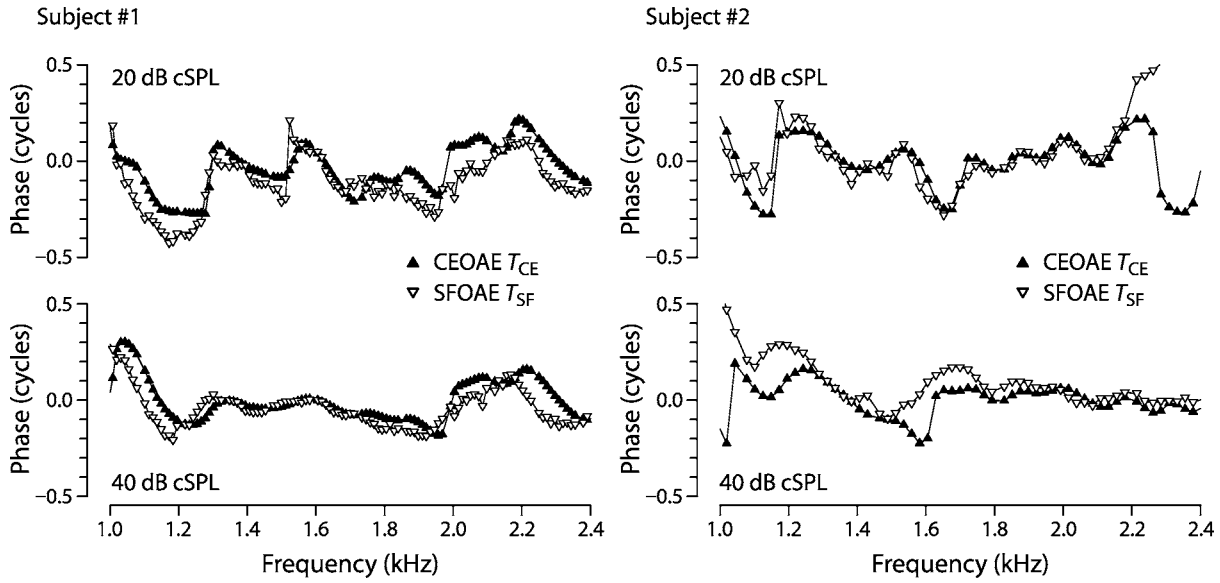


FIG. 10. Details of CEOAE and SFOAE transfer-function phase at matched intensities. The figure shows $\angle T_{CE}(f;A)$ and $\angle T_{SF}(f;A)$ reproduced from Fig. 9 after subtracting out smooth curves that capture the secular variation of the phase. Columns show measurements in two different subjects. The top row shows detrended phases measured at 20 dB cSPL; the bottom row shows detrended phase at 40 dB cSPL. At each level the same trend curves were subtracted from both $\angle T_{CE}(f;A)$ and $\angle T_{SF}(f;A)$.

tional studies are needed to determine whether our findings generalize, we have no reason to suspect that similar conclusions will not apply to impaired ears, so long as their emissions remain measurable. (3) We used low to moderate stimulus levels (35–80 dB pSPL for broadband clicks, 10–40 dB SPL for tones) and cannot rule out the possibility that SFOAE and CEOAE transfer functions differ more significantly at higher stimulus levels. (4) In order to reduce interference between the stimulus and the emission, we used time windows to eliminate CEOAE components arriving earlier than about 5 ms after the stimulus peak. In addition to removing high-frequency components of the response, this windowing may also have removed possible short-latency low-frequency components generated in the base of the cochlea. Although accurate estimates of the magnitudes of these components were compromised by system nonlinearities (see below), measurements in test cavities imply that any such short-latency components must be small relative to the long-latency components. (5) Our comparisons are limited to the frequency range of 1 to 3 kHz. In particular, we did not explore the behavior in more apical regions of the cochlea, where emission mechanisms may differ from those in the base (Shera and Guinan, 2003; Siegel *et al.*, 2005). (6) We did not systematically explore a wide range of stimulus presentation rates (e.g., for the click stimuli) in every subject. Since high-rate clicks are generally much more effective elicitors of efferent activity than the stimuli used to measure SFOAEs (Guinan *et al.*, 2003), we checked for differences related to efferent effects by varying the click-repetition period in two subjects. Although we found no obvious effects of click-repetition period in these subjects, the strength of otoacoustic efferent effects varies from individual to individual, and we may simply have "gotten lucky." It remains possible, even likely, that differences in the strength of efferent feedback elicited by the two stimuli can produce differences in $T_{CE}(f;A)$ and $T_{SF}(f;A)$ in some subjects, at least when $T_{CE}(f;A)$ is measured using high-rate clicks. (7) Finally, our measurements are in humans, a species whose OAE characteristics differ in some respects from those of many laboratory animals (e.g., humans have longer OAE latencies and smaller distortion-source emissions). The near-equivalence we find between $T_{CE}(f;A)$ and $T_{SF}(f;A)$ remains to be examined in other species.

B. Quasilinearity of the transfer functions

If the mechanisms of OAE generation and propagation were completely linear, the near-equivalence we find between CEOAE and SFOAE transfer functions would be entirely expected. The response of a linear system can be equivalently characterized either in the time domain using broadband stimuli (such as the click stimulus used to evoke CEOAEs) or in the frequency domain using narrow-band stimuli (such as the pure tone used to evoke SFOAEs). If the cochlea were a linear system, the principle of superposition would require that transfer functions measured in the time and frequency domains be identical, regardless of the details of emission generation.

Our data support the notion that cochlear responses are nearly linear at levels approaching the threshold of hearing. For example, we find that the transfer functions $T_{CE}(f;A)$ and $T_{SF}(f;A)$ are almost identical and independent of stimulus intensity and bandwidth at low levels. Furthermore, CEOAE transfer functions obtained using the nonlinear-residual method, a method that relies on nonlinear OAE growth to extract the emission, fall into the noise floor at low intensities. These results are consistent with previous OAE measurements (e.g., Zwicker and Schloth, 1984; Shera and Zweig, 1993), including those demonstrating approximate linear superposition among OAEs evoked by various low-level stimuli (Zwicker, 1983; Probst *et al.*, 1986; Xu *et al.*, 1994). Linearity of CEOAE and SFOAE responses at low

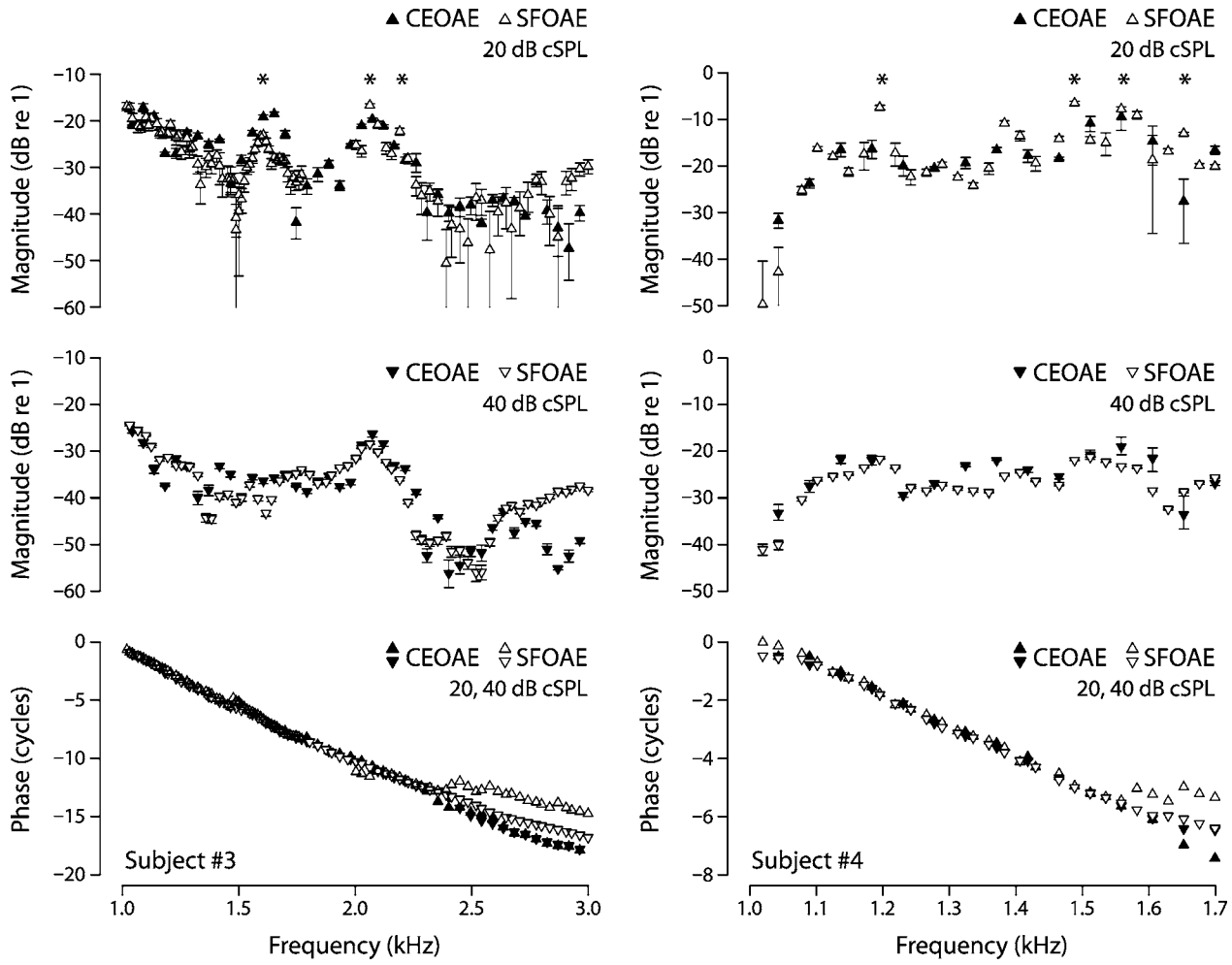


FIG. 11. CEOAE and SFOAE transfer functions at matched intensities in subjects with SSOAE. Columns show measurements in two different subjects. The top row shows transfer-function magnitudes measured at 20 dB cSPL; the middle row shows magnitudes at 40 dB cSPL. The bottom row shows unwrapped phases at both intensities. In all panels, $T_{CE}(f;A)$ and $T_{SF}(f;A)$ are shown with filled and open symbols, respectively. SSOAE frequencies are identified by asterisks (*) in the top panels.

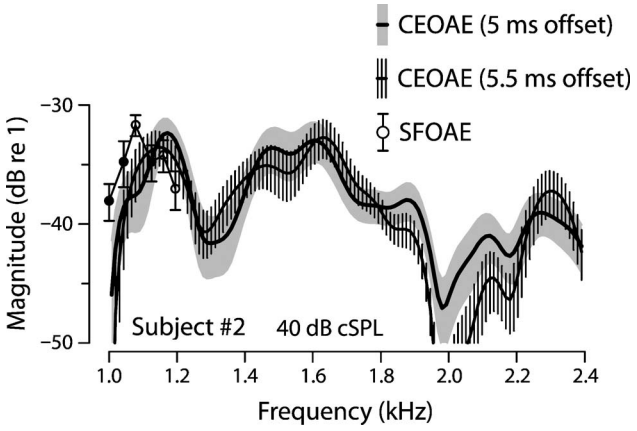


FIG. 12. Session-to-session and methodological variability. Solid curves show mean magnitudes of $T_{CE}(f;A)$ obtained from the same emission measurements (subject 2, 62 measurements distributed over 3 months) using different time offsets for the analysis window (O_e=5 and 5.5 ms). The flanking gray and hatched regions extend one standard deviation above and below the mean. The open circles with error bars show mean magnitudes of $T_{SF}(f;A)$ (30 measurements, same time frame). All stimulus intensities were 40 dB cSPL.

levels is also consistent with basilar-membrane mechanical responses (reviewed in Robles and Ruggero, 2001), which manifest approximate linearity at levels approaching threshold.

Since the operation of the cochlea is certainly nonlinear at intensities not far above threshold, our finding that CEOAE and SFOAE transfer functions continue to match even at moderate levels is more unexpected. The continuing match suggests that as stimulus intensities rise the cochlea emerges gracefully from the low-level linear regime. In particular, the observed match between the spectral structure of CEOAE and SFOAE transfer functions suggests that the reverse-traveling waves that combine to form CEOAEs arise by an approximately linear mechanism (e.g., "scattering") in which interactions among the various frequency components of the stimulus (e.g., intermodulation distortion) play only a secondary role. Although nonlinear interactions such as self- and mutual suppression affect the overall emission magnitudes (e.g., by influencing the gain of the cochlear amplifier), intermodulation distortion does not appear to be primarily responsible for generating the reverse-traveling waves themselves. Our results are consistent with studies of the OAEs evoked by broadband noise (Maat *et al.*, 2000), where

Wiener-kernel analysis indicates that although the overall emission amplitude varies with stimulus intensity, the cochlear response appears approximately linear at each level. Analogous results, including strong if imperfect matches between responses evoked by broad- and narrow-band stimuli, are found in measurements of basilar-membrane motion (e.g., Recio and Rhode, 2000; de Boer and Nuttall, 2002).

Our findings are also consistent with those of Prieve *et al.* (1996), who found that CEOAEs and tone-burst evoked OAEs (TBOAEs) have similar growth functions. They concluded that emissions evoked by the two stimuli share common mechanisms of generation and, in particular, that both are generated by mechanisms acting in independent frequency channels. This conclusion was questioned by Yates and Withnell (1999b), who pointed out that although the tone-burst bandwidths (which generally spanned an octave or more) were narrower than those of the clicks, they were still broad enough to excite the same complex cross-frequency interactions as the click. They therefore argued that the growth functions matched not because of OAE generation via independent frequency channels, but precisely the opposite: because both stimuli produce nonlinear interactions among the different spectral components of the stimulus. Our data do not support Yates and Withnell's suggestion: We measured SFOAEs using continuous narrow-band pure tones (rather than the relatively broadband TBOAEs used by Prieve *et al.*) and still found the reported match between wide- and narrow-band growth functions.

C. Interpreting the unification

The unification between CEOAE and SFOAE growth functions predicted by extrapolation from Eq. (17) and demonstrated experimentally in Fig. 8 was obtained using OAE data from a single individual measured in a narrow range of frequencies. Despite this limitation, the quantitative agreement between $T_{CE}(f;A)$ and $T_{SF}(f;A)$ apparent at matched (bandwidth-compensated) intensities demonstrated in Figs. 9 and 11 suggests that relations similar to Eq. (17) but with frequency- and perhaps subject-dependent parameters (e.g., T_0, A_0, ΔF, α) may apply more widely. [Equation (17) clearly breaks down at frequencies where monotonic emission growth is interrupted by "interference notches;" the frequencies f_0 represented by the data in Fig. 7 were chosen to avoid this behavior.]

Although we caution against overinterpretation of Eq. (17) until its empirical foundation and region of validity are more firmly understood, the frequency scale ΔF merits further comment. Its appearance in the CEOAE growth function presumably reflects an effective integration bandwidth for CEOAE generation, analogous to the equivalent rectangular band (ERB) of an auditory filter. In this regard we note that the best-fit value $\Delta F \cong 74 \pm 20$ Hz obtained from the OAE growth functions at $f_0 \cong 1.2$ kHz is comparable to the auditory-filter ERB [$\mathrm{ERB}(f_0) \cong 90 \pm 10$ Hz] obtained from independent otoacoustic and psychophysical measurements (Shera *et al.*, 2002; Oxenham and Shera, 2003). A more systematic study could determine whether Eq. (17) and the approximate equality of bandwidths observed here hold at other frequencies. Interpreted as an effective integration bandwidth, a nonzero ΔF implies that the "channels" associated with CEOAE generation are not truly independent (e.g., Prieve *et al.* 1996). Evidently, the CEOAE at any given frequency is affected by stimulus energy at nearby frequencies, presumably through suppression.

D. Consistency with the DP-place component of DPOAEs

The match we find between $T_{CE}(f;A)$ and $T_{SF}(f;A)$ is consistent with emission measurements that report strong similarities between CEOAEs and certain DPOAEs (Knight and Kemp, 1999), in particular upper-sideband DPOAEs and lower-sideband DPOAEs measured at f_2/f_1 ratios close to 1. DPOAEs are typically mixtures of emissions originating from at least two different regions of the cochlea, namely the region where the responses to the primaries overlap and the region tuned to the distortion-product frequency (e.g., Kim, 1980; Kemp and Brown, 1983; Gaskill and Brown, 1990; Brown *et al.*, 1996; Engdahl and Kemp, 1996; Brown and Beveridge, 1997; Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Kalluri and Shera, 2001; Knight and Kemp, 2001). Theory and experiment both indicate that the relative contribution of the components from these two locations varies systematically with stimulus parameters (e.g., Fahey and Allen, 1997; Knight and Kemp, 2001; Shera and Guinan, 2007). In particular, upper-sideband DPOAEs and lower-sideband DPOAEs measured with f_2/f_1 ratios close to 1 are generally dominated by emissions from the distortion-product place, oil whose characteristics are very similar to SFOAEs. Indeed, Kalluri and Shera (2001) showed by direct comparison that the DPOAE component originating from the DP place closely matched the SFOAE evoked at the same frequency. Previous results thus establish that (1) CEOAEs resemble the DP-place component of DPOAEs and (2) the DP-place component of DPOAEs matches SFOAEs. Taken together, these results are consistent with the equivalence reported here between CEOAE and SFOAE transfer functions.

E. Implications for emission mechanisms

Our results contradict two proposed models of CEOAE generation, both of which suggest that CEOAEs originate primarily by nonlinear mechanisms within the cochlea. Nobili and colleagues argue that CEOAEs arise from spatially complex, nonlinear "residual oscillations" of the basilar membrane that trace their origin to spectral irregularities in middle-ear transmission (Nobili, 2000; Nobili *et al.*, 2003a,b). Based on their model simulations, Nobili *et al.* conclude that transient evoked OAEs that occur in the absence of spontaneous emissions result from a "spatial imbalance" in cochlear nonlinearity and amplification caused by rapid frequency variations in forward middle-ear filtering. In this view, CEOAEs result from mechanisms that are both inherently nonlinear and fundamentally different from those responsible for generating SFOAEs. We note, for example, that Nobili *et al.*'s proposed middle-ear filtering mechanism for generating CEOAEs cannot produce SFOAEs at any

level of stimulation: Although CEOAEs are evoked by transient stimuli containing many frequency components, and are therefore potentially sensitive to frequency variations in middle-ear transmission as proposed, SFOAEs are evoked by pure (single-frequency) tones and, *ipso facto*, cannot originate via any mechanism that operates *across* frequency. We show here that the characteristics of CEOAEs and SFOAEs are nearly identical (in ears both with and without SSOAEs), in clear contradiction to Nobili *et al.*'s model predictions.

Our findings also contradict the notion that CEOAEs arise via nonlinear interactions among the frequency components of the stimulus. Based on measurements in guinea pig in which they evoked CEOAEs using high-pass filtered clicks and identified significant OAE energy outside the stimulus passband, Yates and Withnell (1999b) proposed that CEOAEs result primarily from intermodulation distortion within the cochlea. CEOAEs, they suggest, are "predominantly composed of intermodulation distortion energy; each component frequency of a click stimulus presumably interacts with every other component frequency to produce a range of intermodulation distortion products" (Withnell *et al.*, 2000). Our finding that CEOAE and SFOAE transfer functions are almost identical argues against this interpretation, at least in humans.

Although the contribution of nonlinear intermodulation distortion mechanisms to human CEOAEs appears small at low and moderate levels, our use of the windowing technique to measure CEOAEs may have eliminated short-latency distortion components present in the response (e.g., Knight and Kemp, 1999; Withnell and McKinley, 2005). Because of a stimulus artifact due to earphone nonlinearities we were unable to quantify accurately the size of any short-latency physiological component using the nonlinear residual method. Nevertheless we can report that any such short-latency component is small enough to be indistinguishable from the distortion measured in a test cavity of similar impedance. Any short-latency nonlinear component in human ears is therefore small relative to the long-latency linear response. Similar conclusions apply also to human SFOAEs (Shera and Zweig, 1993).

Although the observed equivalence between CEOAEs and SFOAEs contradicts these inherently nonlinear models of CEOAE generation, the equivalence is entirely consistent with predictions of the coherent-reflection model (e.g., Zweig and Shera, 1995; Talmadge *et al.*, 1998; Shera and Guinan, 2007). In this model, OAEs are generated by a process equivalent to wave scattering by preexisting (place-fixed) micromechanical perturbations in the organ of Corti. Not only does the coherent-reflection model predict the empirical equivalence between $T_{\mathrm{CE}}(f;A)$ and $T_{\mathrm{SF}}(f;A)$, the model also predicts the observed spectral characteristics of the transfer functions across frequency (e.g., their slowly varying amplitudes punctuated by sharp notches and their rapidly rotating phases).

Because different stimuli are used to evoke them, CEOAEs and SFOAEs are conventionally classified as different OAE types. Our results establish, however, that at low and moderate stimulus intensities these two OAE "types" are really the same emission evoked in different ways—CEOAEs and SFOAEs are evidently better understood as members of the same emission family. Our findings thus support the mechanism-based classification scheme proposed elsewhere (Shera and Guinan, 1999; Shera, 2004).

ACKNOWLEDGMENTS

We thank Christopher Bergevin, John Guinan, and the anonymous reviewers for helpful comments on the manuscript. This work was supported by Grant No. R01 DC03687 from the NIDCD, National Institutes of Health.

[1]In subjects without SSOAE, the transfer function's standard deviation is computed either from the noise floor or, when possible, by finding the deviation of multiple runs. In subjects with SSOAE, we always made multiple measurements, particularly at low stimulus levels where the SSOAE were likely to have the most influence. Because SOAEs typically have constant magnitude but random phase, complex averaging of multiple runs allows one to partially eliminate the SSOAE. The standard deviation represents the uncertainty of the mean value.

[2]The peak-equivalent pressure of the click is the rms pressure of the pure tone with the same peak pressure.

[3]The equivalent rectangular bandwidths of the click stimuli used to collect the data in Fig. 7 were $\{3.2, 1.8, 1, 0.67\}$ kHz.

[4]At frequencies near a SSOAE, however, the across-day variability can be considerably larger (see Fig. 5), perhaps because the strength or synchronizability of the SSOAE varies from day to day.

Brown, A. M., and Beveridge, H. A. (**1997**). "Two components of acoustic distortion: Differential effects of contralateral sound and aspirin," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. L. Hecht-Poinar (World Scientific, Singapore), pp. 219–225.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). "Two sources of acoustic distortion products from the human cochlea," J. Acoust. Soc. Am. **100**, 3260–3267.

Carvalho, S., Buki, B., Bonfils, P., and Avan, P. (**2003**). "Effect of click intensity on click-evoked otoacoustic emission waveforms: Implications for the origin of emissions," Hear. Res. **175**, 215–225.

de Boer, E., and Nuttall, A. L. (**2002**). "The mechanical waveform of the basilar membrane. IV. Tone and noise stimuli," J. Acoust. Soc. Am. **111**, 979–985.

Engdahl, B., and Kemp, D. T. (**1996**). "The effect of noise exposure on the details of distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **99**, 1573–1587.

Fahey, P. F., and Allen, J. B. (**1997**). "Measurement of distortion product phase in the ear canal of the cat," J. Acoust. Soc. Am. **102**, 2880–2891.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

Gelb, A., and Vander Velde, W. E. (**1968**). *Multiple-Input Describing Functions and Nonlinear System Design* (McGraw Hill, New York).

Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivo-cochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–540.

Heitmann, J., Waldman, 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.

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.

Kapadia, S., Lutman, M. E., and Palmer, A. R. (**2005**). "Transducer hysteresis contributes to "stimulus artifact" in the measurement of click-evoked otoacoustic emissions," J. Acoust. Soc. Am. **118**, 620–622.

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T., and Brown, A. M. (**1983**). "An integrated view of cochlear mechanical nonlinearities observable from the ear canal," in *Mechanics of*

Hearing, edited by E. de Boer and M. A. Viergever (Martinus Nijhoff, The Hague), pp. 75–82.
Kim, D. O. (**1980**). "Cochlear mechanics: Implications of electrophysiological and acoustical observations," Hear. Res. **2**, 297–317.
Knight, R. D., and Kemp, D. T. (**1999**). "Relationships between DPOAE and TEOAE amplitude and phase characteristics," J. Acoust. Soc. Am. **106**, 1420–1435.
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.
Krylov, N. M., and Bogolyubov, N. N. (**1947**). *Introduction to Nonlinear Mechanics* (Princeton U. P., Princeton).
Lapsley-Miller, J. A., Boege, P., Marshall, L., and Jeng, P. S. (**2004a**). "Transient-evoked otoacoustic emissions: Preliminary results for validity of TEOAEs implemented on Mimosa Acoustics T2K Measurement System v3.1.3," Technical Report No. 1232, Naval Submarine Medical Research Lab, Groton, CT.
Lapsley-Miller, J. A., Boege, P., Marshall, L., Shera, C. A., and Jeng, P. S. (**2004b**). "Stimulus-frequency otoacoustic emissions: Validity and reliability of SFOAEs implemented on Mimosa Acoustics SFOAE Measurement System v2.1.18," Technical Report No. 1231, Naval Submarine Medical Research Lab, Groton, CT.
Maat, B., Wit, H., and van Dijk, P. (**2000**). "Noise-evoked otoacoustic emissions in humans," J. Acoust. Soc. Am. **108**, 2272–2280.
Nobili, R. (**2000**). "Otoacoustic emissions simulated by a realistic cochlear model," in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 402–408.
Nobili, R., Vetešník, A., Turicchia, L., and Mammano, F. (**2003a**). "Otoacoustic emissions from residual oscillations of the cochlear basilar membrane in a human ear model," J. Assoc. Res. Otolaryngol. **4**, 478–494.
Nobili, R., Vetešník, A., Turicchia, L., and Mammano, F. (**2003b**). "Otoacoustic emissions simulated in the time domain by a hydrodynamic model of the human cochlea," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 524–530.
Oxenham, A. J., 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.
Prieve, B. A., Gorga, M. P., and Neely, S. T. (**1996**). "Click- and tone-burst-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears," J. Acoust. Soc. Am. **99**, 3077–3086.
Probst, R., Coats, A. C., Martin, G. K., and Lonsbury-Martin, B. (**1986**). "Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears," Hear. Res. **21**, 261–275.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **5**, 108.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
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. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, J. J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., and Guinan, J. J. (**2007**). "Mechanisms of mammalian otoacoustic emission," in *Active Processes and Otoacoustic Emissions*, edited by G. Manley, B. L. Lonsbury-Martin, A. N. Popper, and R. R. Fay, in press (Springer-Verlag, New York).
Shera, C. A., and Zweig, G. (**1993**). "Noninvasive measurement of the cochlear traveling-wave ratio," J. Acoust. Soc. Am. **93**, 3333–3352.
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.
Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Rugero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.
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.
Withnell, R. H., and McKinley, S. (**2005**). "Delay dependence for the origin of the nonlinear derived transient evoked otoacoustic emission," J. Acoust. Soc. Am. **117**, 281–291.
Withnell, R. H., and Yates, G. K. (**1998**). "Enhancement of the transient-evoked otoacoustic emission produced by the addition of a pure tone in the guinea pig," J. Acoust. Soc. Am. **104**, 344–349.
Withnell, R. H., Yates, G. K., and Kirk, D. L. (**2000**). "Changes to low-frequency components of the TEOAE following acoustic trauma to the base of the cochlea," Hear. Res. **139**, 1–12.
Xu, L., Probst, R., Harris, F. P., and Roede, J. (**1994**). "Peripheral analysis of frequency in human ears revealed by tone burst evoked otoacoustic emissions," Hear. Res. **74**, 173–180.
Yates, G. K., and Withnell, R. H. (**1999a**). "Reply to "Comment on 'Enhancement of the transient-evoked otoacoustic emission produced by the addition of a pure tone in the guinea pig'" [J. Acoust. Soc. Am. 105, 919–921 (1999)]," J. Acoust. Soc. Am. **105**, 922–924.
Yates, G. K., and Withnell, R. H. (**1999b**). "The role of intermodulation distortion in transient-evoked otoacoustic emissions," Hear. Res. **136**, 49–64.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.
Zwicker, E. (**1983**). "Delayed evoked oto-acoustic emissions and their suppression by Gaussian-shaped pressure impulses," Hear. Res. **11**, 359–371.
Zwicker, E., and Schloth, E. (**1984**). "Interrelation of different oto-acoustic emissions," J. Acoust. Soc. Am. **75**, 1148–1154.

Modeling comodulation masking release using an equalization-cancellation mechanism

Tobias Piechowiak, Stephan D. Ewert,[a] and Torsten Dau[b]
Centre for Applied Hearing Research, Acoustic Technology, Ørsted•DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

(Received 30 August 2006; revised 4 January 2007; accepted 5 January 2007)

This study presents an auditory processing model that accounts for the perceptual phenomenon of comodulation masking release (CMR). The model includes an equalization-cancellation (EC) stage for the processing of activity across the audio-frequency axis. The EC process across frequency takes place at the output of a modulation filterbank assumed for each audio-frequency channel. The model was evaluated in three experimental conditions: (i) CMR with four widely spaced flanking bands in order to study pure across-channel processing, (ii) CMR with one flanking band varying in frequency in order to study the transition between conditions dominated by within-channel processing and those dominated by across-channel processing, and (iii) CMR obtained in the "classical" band-widening paradigm in order to study the role of across-channel processing in a condition which always includes within-channel processing. The simulations support the hypothesis that within-channel contributions to CMR can be as large as 15 dB. The across-channel process is robust but small (about 2–4 dB) and only observable at small masker bandwidths. Overall, the proposed model might provide an interesting framework for the analysis of fluctuating sounds in the auditory system. [DOI: 10.1121/1.2534227]

PACS number(s): 43.66.Ba, 43.66.Mk [JHG] Pages: 2111–2126

I. INTRODUCTION

Many properties of auditory masking can be understood in terms of the responses of the basilar membrane within the inner ear. Each part of this membrane behaves like a filter that responds to a limited range of frequencies. When trying to detect a sinusoidal tone in background noise, it has been proposed that listeners use the output of a single auditory filter tuned to the frequency of the tone (Fletcher, 1940). That filter passes the tone at full intensity, but rejects most of the background noise. Although this theory can account for many aspects of masking, Hall *et al.* (1984a) and others showed that, when comodulated maskers were used, some of the results can be explained only if it is assumed that stimulus information is processed across the outputs of auditory filters. In fact, humans are often much better at detecting signals in comodulated maskers than in white noise, an effect called comodulation masking release (CMR; Hall *et al.*, 1984a). Various experiments on CMR have demonstrated that the human auditory system can exploit coherent envelope fluctuations very effectively and that substantial reductions in signal threshold can be the result. Since coherent across-frequency modulation is common in speech, music, animal vocalization and environmental noise, the ability to process such information is thought to be a powerful survival strategy in the natural world which aids in the detection of target sounds in the presence of competing sounds.

CMR was demonstrated initially by Hall *et al.* (1984a). In their "band-widening" experiment, the detection of a tone was measured as a function of the bandwidth of a noise masker, keeping the spectrum level constant. They used two types of maskers. One was a random noise with irregular fluctuations in amplitude that are independent in different frequency regions. The other was a random noise which was amplitude modulated using a low-pass filtered noise as a modulator. This modulation resulted in slow fluctuations in the amplitude of the noise that were the same in different frequency regions. For the random noise, the signal threshold increased as the masker bandwidth increased up to about the critical bandwidth at that frequency and then remained constant, as expected from the classical power spectrum model of masking (Fletcher, 1940; Patterson and Moore, 1986). The pattern for the modulated noise was quite different. Here, the threshold decreased as the bandwidth was increased beyond about 100 Hz (for a signal frequency of 2 kHz); thus, adding more noise to the masker made the signal easier to detect. This suggested that subjects may compare the outputs of different auditory filters to enhance signal detection. The fact that the decrease in threshold with increasing bandwidth only occurred with the modulated noise indicated that fluctuations in the masker are critical and that the fluctuations need to be correlated across frequency bands.

In a second class of experiments, CMR was demonstrated by using narrow bands of noise (of typically 20–50 Hz width), which inherently have relatively slow amplitude fluctuations. One band, the on-frequency band, was centered at the signal frequency. A second band, the flanker band, was placed remote from the signal frequency. When the flanking band was uncorrelated with the on-frequency band, there was typically no effect on signal threshold. How-

[a]Current address: Carl von Ossietzky Universität Oldenburg, Medizinische Physik D-26111 Oldenburg, Germany
[b]Author to whom correspondence should be addressed. Electronic mail: tda@oersted.dtu.dk

ever, when the flanking band was correlated with the on-frequency band, a flanking band produced a release from masking (Hall *et al.*, 1984a; Schooneveldt and Moore, 1987; Cohen and Schubert, 1987). CMR was also found even if the signal and on-frequency band were presented to one ear and the flanking band to the other ear (Schooneveldt and Moore, 1987; Cohen and Schubert, 1987).

Even though CMR has been investigated in a number of studies, the underlying mechanisms are still not clear. It has generally been assumed that CMR results from across-channel comparisons of temporal envelopes. Alternatively, it has been suggested that analysis of the output of a broad initial predetection filter, which encompasses frequencies generally thought to fall into separate auditory filters, can account for certain aspects of CMR (Berg, 1996). However, Buss *et al.* (1998) and Buss and Hall (1998) provided evidence against such a broad predetection filter; their results were, instead, consistent with an initial stage of auditory (bandpass) filtering. Other studies have proposed that within-channel cues, i.e., information from only the one auditory channel tuned to the signal frequency, can account for a considerable part of the effect in some conditions, which means that within-channel processing can lead to an overestimation of "true" across-channel CMR (e.g., Schooneveldt and Moore, 1987). This was supported by simulations of data from the band-widening experiment, using a modulation filterbank analysis of the stimuli at the output of the auditory filter tuned to the signal frequency (Verhey *et al.*, 1999). Additionally, for the CMR experiments using flanking bands, McFadden (1986) pointed out that it is imprecise to assume that one channel is receiving only the on-frequency band plus signal and another channel is receiving only the flanking band. Often, the two bands will be incompletely resolved. When this happens, the resulting waveform may contain envelope fluctuations resulting from beats between the carrier frequencies of the on-frequency and the flanker bands. These beats can facilitate signal detection without across-channel comparisons being involved. Thus, at least part of the masking release can be explained in terms of the use of within-channel rather than across-channel cues. Taken together, across-channel CMR appears to be a robust, but relatively small effect, which was found in monotic and dichotic conditions.

A recent study on effects of auditory grouping on CMR (Dau *et al.*, 2005) supported two forms of CMR. In their study, the effects of introducing a gating asynchrony between on-frequency and flanker bands or a stream of preceding (precursor) or following (postcursor) flanker bands were studied for conditions of CMR. Using widely (one octave) spaced flanking bands, CMR effects were eliminated by introducing a gating asynchrony and by introducing the pre-or postcursor flanking bands. Using narrowly spaced flanking bands (one-sixth octave), CMR was not affected by any of the stimulus manipulations. Their results supported the hypothesis that one form of CMR is based on within-channel mechanisms (Schooneveldt and Moore, 1987; Verhey *et al.* (1999)), determined by the envelope statistics. The fact that this effect was not susceptible to manipulations by auditory grouping constraints is in line with the assumption that the mechanism is peripheral in nature, based on the physical interaction of stimulus components within an auditory channel. The other form of CMR, mainly based on "true" across-channel comparisons, appeared to be dependent on auditory grouping constraints, consistent with the results from Grose and Hall (1993).

Several hypotheses have been suggested about the nature of the across-channel mechanism underlying CMR. One hypothesis is based on the assumption that the addition of the signal to the on-frequency masker band leads to a change in the modulation depth in the auditory filter centered at the signal frequency. By comparing this modulation depth to that of other auditory filters for which the modulation depth is unaltered, subjects would increase their sensitivity to the presence of the signal (Hall, 1986). A different explanation for CMR was proposed by Buus (1985), who suggested that the comodulated flanker band(s) provide valuable information about the moments at which the masker band has a relatively low energy. By attributing more weights to these valleys in the masker, the effective signal-to-noise ratio increases and detection improves. This mechanism was called "listening in the valleys." Also proposed by Buus (1985) was an equalization-cancellation (EC) mechanism, originally introduced by Durlach (1963), to account for various binaural masking release data. According to this mechanism, the envelope of the masker and flanking band are first equalized and then subtracted. The output of such a mechanism might have a considerable increase in the signal-to-noise ratio provided that the masker and the flanking bands are comodulated.

A fourth mechanism has been proposed by Richards (1987), where it was assumed that the cross *covariance* between the envelopes of the masker and the flanking bands is used for signal detection. The envelope cross covariance decreases when adding a signal to the masker and this cue might be used by the human auditory system. However, this model was later rejected because it was not compatible with experimental data by Edins and Wright (1984). They used two 100% sinusoidally amplitude modulated sinusoids of different frequencies, and the subjects had to detect the in-phase addition of a sinusoid to one of the SAM sinusoids. The cross covariance is not changed even though the modulation pattern is altered by the addition of the sinusoid. Thus, if changes in the cross covariance were essential for receiving CMR, this type of stimulus should not lead to a CMR. This, however, was in contrast to their data, which clearly showed CMR.

Later, van de Par and Kohlrausch (1998b) and van de Par (1998) found that CMR can better be described in terms of an envelope cross *correlation* mechanism than an envelope cross covariance mechanism. Their study was motivated by earlier findings by Bernstein and Trahiotis (1996) which showed that cross correlation was more successful than cross covariance when studying binaural detection phenomena. At high frequencies where these experiments were carried out, similar mechanisms may indeed underly monaural CMR and binaural masking level differences (BMLD, van de Par and Kohlrausch, 1998a). Moreover, the EC mechanism, which has been used to account for BMLD, was shown to be

equivalent to a decision mechanism based on cross correlation (Domnitz and Colburn, 1977; Green, 1992).

While potential mechanisms underlying CMR have been discussed in several studies, predictions that quantify the (relative) contributions of across- versus within-channel processing in different types of experiments have not been provided. The purpose of the present study was therefore to develop and evaluate a model that accounts for both effects in CMR. The modulation filterbank model by Dau *et al.* (1997a, b) was considered as the modeling framework. This model was used earlier to analyze within-channel cues in CMR obtained in the band-widening experiment (Verhey *et al.*, 1999), and applied to a variety of other detection and masking conditions, including tone-in-noise detection, modulation detection, and forward masking. In the Verhey *et al.* (1999) study, the model was exclusively tested in the band-widening experiment of CMR. The results from the simulations, performed only in the auditory channel tuned to the signal frequency, suggested that essentially no across-channel processing is involved in this type of CMR condition. Instead, temporal within-channel cues such as beating between components, evaluated by the modulation filterbank model, appear to account for the masking release in the model simulations. However, since the model does not contain any explicit across-channel processing, it will not be able to account for any true across-channel CMR. In the present study, an EC-based circuit was integrated into an extended version of the modulation filterbank model whereby the EC processing was assumed to take place at the level of the internal representation of the stimuli *after* modulation filtering.

First, the structure of the across-channel modulation filterbank model is described. The model is then evaluated in several experimental conditions: (i) CMR with four widely spaced flanker bands to study pure across-channel CMR (Experiment 1), (ii) CMR with one flanking band varying in frequency in order to study the transition between conditions dominated by within-channel processing and those dominated by across-channel processing (Experiment 2), and (iii) CMR obtained in the band-widening paradigm in order to study the contribution of across-channel processing in a condition which always includes within-channel processing (Experiment 3). For direct comparison, experimental data were obtained in the same conditions with exactly the same stimuli and using exactly the same threshold algorithm as in the simulations. The results and implications for further modeling work are discussed.

II. MODEL

The model presented here is based on the monaural detection model of Dau *et al.* (1997a). The original model was designed to account for signal detection data in various psychoacoustic conditions. It has proven successful in predicting data from spectral and spectro-temporal masking (Verhey *et al.*, 1999; Derleth and Dau, 2000; Verhey, 2002), nonsimultaneous masking (Dau *et al.*, 1996, 1997a; Derleth, *et al.*, 2001) and modulation detection and masking (Dau *et al.*1997a, b, 1999; Ewert and Dau, 2004). In the meantime, an additional model of amplitude modulation (AM) processing, the envelope power spectrum model (EPSM) has been developed (Ewert and Dau, 2000; Ewert *et al.*, 2002), based on Viemeister (1979) and Dau *et al.* (1999). The EPSM has a much simpler structure than the abovementioned processing model. It is similar to Viemeister's (1979) leaky-integrator model but assumes modulation bandpass filters instead of a single modulation lowpass filter. It consists of only three stages: Hilbert-envelope extraction, modulation bandpass filtering, and a decision stage based on the long-term, mean integrated envelope power. This model does not include any effects of peripheral filtering and adaptation, and timing information (as reflected in the envelope phase and modulation beatings) is neglected. While the EPSM demonstrated in a straightforward and intuitive way the need for modulation-frequency selective processing and can account for modulation masking data, it is conceptually less general than the perception model (Dau *et al.*, 1996, 1997a).

The model as described in Dau *et al.* (1997a), which forms the basis for the model developed here, consists of the following steps: Peripheral filtering, envelope extraction, nonlinear adaptation, modulation filtering, and an optimal detector as the decision device. To simulate the bandpass characteristic of the basilar membrane, the gammatone filterbank (Patterson *et al.*, 1987) is used. At the output of each peripheral filter, the model includes half-wave rectification and low-pass filtering at 1 kHz. While the fine structure is preserved for low frequencies, for high center frequencies this stage essentially preserves the envelope of the signal. Effects of adaptation are simulated by a nonlinear adaptation circuit (Püschel, 1988; Dau *et al.*, 1996). For a stationary input stimulus, this stage creates a compression close to logarithmic. With regard to the transformation of envelope fluctuations, the adaptation stage transforms the AM depth of input fluctuations with rates higher than about 2 Hz almost linearly. The stimuli at the output of the adaptation stage for each channel are then processed by a linear modulation filterbank. The lowest modulation filter is a second-order butterworth lowpass filter with a cutoff frequency of 2.5 Hz. For frequencies above 5 Hz there is an array of bandpass filters with a quality factor of $Q=2$. Modulation filters with a center frequency above 10 Hz only output the Hilbert envelope of the modulation filters, introducing a nonlinearity into the modulation processing through which the phase of the envelope is not preserved for the filters above 10 Hz. To model a limit of resolution, an internal noise with a constant variance is added to the output of each modulation filter. In the decision process, a stored, normalized temporal representation of the signal to be detected (the template) is compared with the actual activity pattern by calculating the cross correlation between the two temporal patterns (Dau *et al.*, 1996, 1997a). This is comparable to a "matched filtering" process (Green and Swets, 1966).

For the processing of arbitrary input stimuli, the function of the model can be considered as being separated in two (parallel) paths: (i) The stimulus representation after nonlinear adaptation is low-pass filtered at a cutoff frequency of 2.5 Hz, thereby essentially extracting the stimulus energy. With this processing alone, the model would be acting simi-

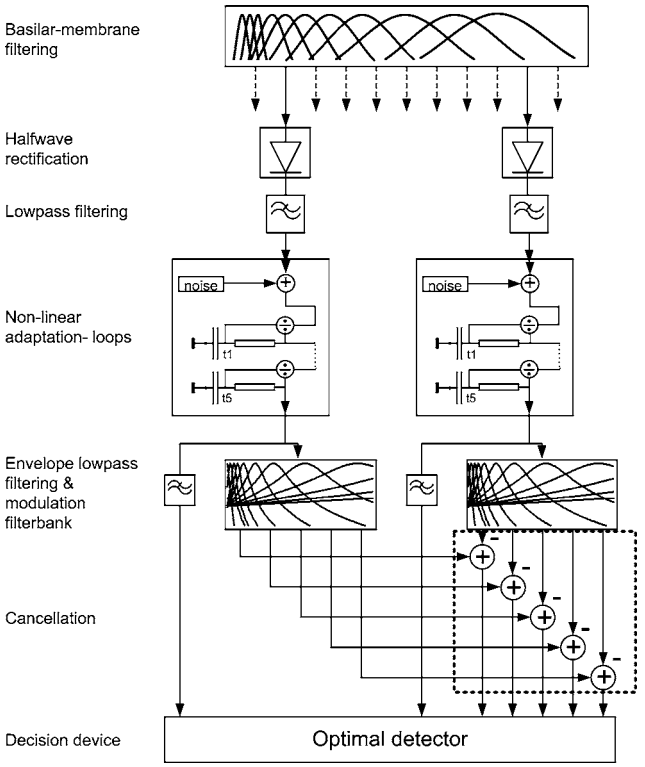


FIG. 1. Block diagram of the across-channel modulation filterbank model. The signals are filtered by the gammatone filterbank, half-wave rectified and low-pass filtered at 1 kHz, and subjected to adaptation. The adapted signal is then filtered by a modulation bandpass filterbank and a separate low-pass filter (at 2.5 Hz) at the output of each auditory filter. At the output of the individual modulation bandpass filters, the activity at the flanking bands is averaged across the flankers (E-process) and subtracted from the corresponding activity at the on-frequency band (C-process), illustrated here with only one flanking band and highlighted in the dashed box. The output activity is added to internal noise and finally subjected to an optimal detector as decision device.

larly to a power spectrum model (e.g., Patterson and Moore, 1996) and would account for certain aspects of spectral masking data (Derleth and Dau, 2000). In the second path, the bank of modulation bandpass filters captures the dynamic properties of the stimulus. It is expected that, in the model, a hypothetical process underlying across-channel CMR would use the output of the bandpass modulation filters. So far, however, the model in its original form does not contain any explicit across-channel processing and therefore fails to produce "true" across-channel CMR.

The present study introduces an explicit across-channel mechanism into the model. Figure 1 illustrates the model used in the present study. The modification of the model is comparable to the EC mechanism of Durlach's model (Durlach, 1960, 1963) for describing binaural masking level differences (BMLDs). However, while the EC mechanism in the original (binaural) model is applied essentially to the stimulus wave forms, and jitter is provided in the level and time domains in order to limit the resolution in the model, the (monaural) EC process in the current model is applied at a much later stage of auditory processing, and no additional limitations are introduced. In contrast to the original binaural EC model, it is assumed here that the limitations for performance are already included in the processing stages prior to the EC process.

The essential aspects of this approach are first illustrated for only two peripheral channels, i.e., using a channel centered at the on-frequency band including the signal, and a channel centered at one remote flanking band.

The across-channel processing within the model is assumed to occur at the output of all (bandpass) modulation channels tuned to frequencies at and above 5 Hz, which is the center frequency of the lowest modulation filter. The individual modulation filter outputs at the flanking band are subtracted from the corresponding outputs at the on-frequency channel. This process is denoted as cancellation in Fig. 1. The outputs of the low-pass filters in the different peripheral channels remain unaffected. The low-pass filtered outputs as well as the difference representations after modulation bandpass filtering are subjected to the decision stage, the optimal detector, which assumes independent observations for the different inputs. The specific case with only one flanking band does not require an equalization stage.

Typically, more than one flanking band will be presented. The generalized mechanism for the multi-channel case is indicated in Fig. 2. Here, the weighted sum of the activity of the flanking bands is computed and subtracted from the on-frequency channel. Calculating the weighted sum can be considered as equalization process, since it equalizes the summed activity in the different flanking bands with regard to the on-frequency band. The subtraction refers to a cancellation process as in the case with only one flanking band (Fig. 1).

In Fig. 2, a situation with N flanking bands and one on-frequency band is assumed. Here, the EC mechanism acts on the N peripheral channels, denoted as PC1, PC2, …, PCN. PCX indicates the channel centered at the on-frequency band. For simplicity, only the output of the jth modulation filter $s_{jn}(t)$ in the different peripheral channels $(n=1\ldots N)$ is indicated in the figure. The outputs of all other modulation filters are processed in the same way. The output $s_j(t)$ of the EC mechanism for N channels at the jth modulation filter can be expressed as

$$s_j(t) = s_{jx}(t) - c_j(t) = s_{jx}(t) - \frac{\sum_{i=1,\ i\neq x}^{N} w_i a_i s_{ji}}{\sum_{i=1,\ i\neq x}^{N} w_i a_i}, \qquad (1)$$

where the index x denotes the peripheral channel (PCX) tuned to the on-frequency band and $c_j(t)$ represents the cancellation term. The contributions $s_{j1}, s_{j2}, \ldots, s_{jN}$ are weighted by the factors $a_1, a_2, \ldots, a_N$. The weights a_i equal the root mean square (rms) of the low-pass filter output in the channels PCi $(i=1, \ldots, N)$. Since the rms value reflects the average energy of a signal, a_i equals the average energy in the ith peripheral channel. Thus, the weighting with a_i means that the channels that are excited by more input stimulus energy are emphasized relative to the filters which are excited by less. Specifically, filters without excitation by the stimulus do not contribute at all to the cancellation term $c_j(t)$. The cancellation term includes a normalization by the factor $\Sigma_{i\neq x,\ i=1}^{N} w_i a_i$ that is proportional to the overall energy of the stimuli in all peripheral channels except the on-frequency channel. In order to make sure that the EC stage operates *across* channels and does not subtract much

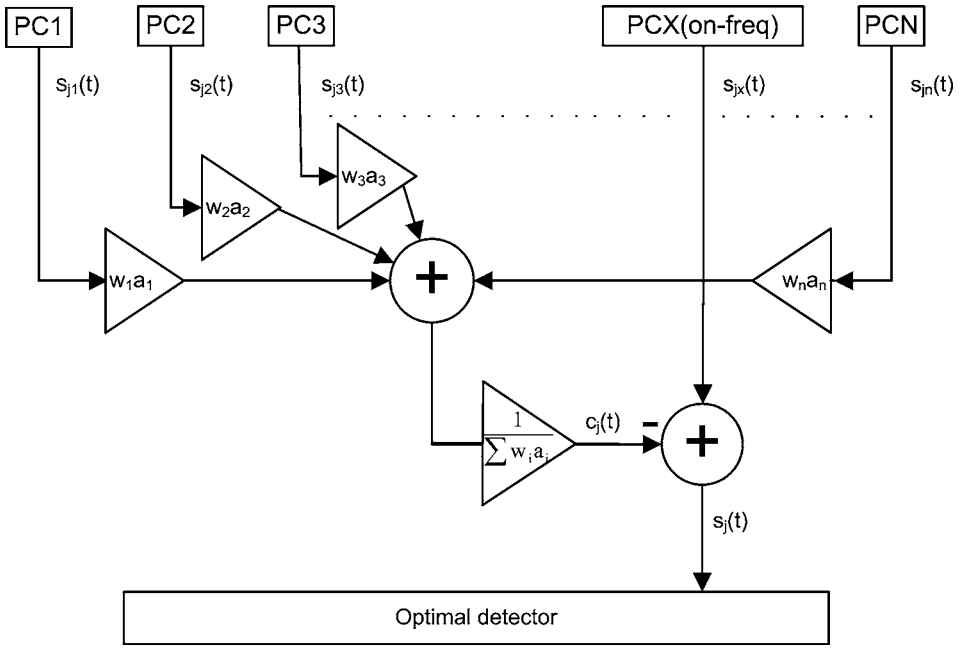


FIG. 2. Simplified block diagram of the across-channel EC process in the perceptual model for N peripheral channels PC1, …, PCN. Only one modulation filter at each peripheral channel is shown.

signal information from the signal channel, the off-frequency weight w_i was introduced. In the current implementation, w_i was set to zero if the overlap between the magnitude transfer function of the auditory channels at *PCi* and *PCX* is above a certain limit, and was set to one otherwise. The overlap of the filter transfer functions was calculated during the design phase of the model as the correlation value of broadband noise at the output of the two respective filters. The limit was chosen to be a correlation value of 5%. In this way, auditory filters tuned at $(i=x)$ and very close $(w_i=0)$ to the signal frequency were not considered in the EC process. The weights w_i ensure that, for example, in the case of a broadband noise as input, the stimuli in the channels contributing to the cancellation term are statistically independent from the excitatory on-frequency channel. Thus the EC mechanism in the model can be regarded as a true across-channel process.

In the most general version of the model, the EC process would be considered in all peripheral channels covering the whole audible frequency range, with each of the channels being regarded as a potential signal channel and with all respective surrounding channels being included in the cancellation term. In the simulations of the present study, however, the model was "told" in advance which was the signal frequency and thus which was the on-frequency channel. All remaining channels in the range from 500 to 6000 Hz were considered as the cancellation channels. This simplification is based on the assumption that the best signal-to-noise ratio is expected to be in the channel tuned to the signal and that detection is mainly based on this single channel (including the information from the other channels contained in the cancellation term of the EC process). An additional simplification was made in conditions when the stimulus was sparsely represented along the peripheral channels as, e.g., in the case of widely spaced narrowband flankers in first experiment. In this case, only channels tuned to the frequencies of the flanker bands were considered. The off-frequency weights w_i were then equal to one. If all flanker bands have equal energy (as in Experiment 1), all a_i have the same value a. The cancellation term $c_j(t)$ in Eq. (1) can then be simplified to

$$c_j(t) = \frac{\sum_{i=1,\ i\neq x}^{N} a s_{ji}}{\sum_{i=1,\ i\neq x}^{N} a} = \frac{\sum_{i=1,\ i\neq x}^{N} s_{ji}}{N-1} \tag{2}$$

and becomes the average over the number of flanking bands.

III. METHOD

A. Subjects

Four normal-hearing subjects participated in each experiment. Their ages ranged from 23 to 41 years. All subjects had experience in other psychoacoustic experiments. The authors T.P. and T.D. participated in the experiment. The other two subjects were paid for their participation on an hourly basis.

B. Apparatus and stimuli

The subjects were seated in a double-walled, sound attenuating booth and listened via Sennheiser HD580 headphones. Signal generation and presentation during the experiments were computer controlled using the AFC software package for MATLAB, developed at Universität Oldenburg and DTU. All stimuli were generated digitally on an IBM compatible PC and were then converted to analog signals by a high-quality 32 bit soundcard (RME DIGI-96PAD) at a sampling rate of 32 kHz. Three CMR experiments were performed where the subject's task was to detect a tone in the presence of one or more noise masker bands. The specific stimuli will be described in the respective experiments (Experiments 1–3).

C. Procedure

A three-interval, three-alternative forced-choice paradigm was used to measure detection thresholds. A two-down,

one-up procedure was used to estimate the 70.7% correct point of the psychometric function (Levitt, 1971) Subjects had to identify the one randomly chosen interval containing the signal. Subjects received visual feedback if the response was correct. The three observation intervals were separated by 500 ms of silence. The initial step size for the signal level was 4 dB and after every second reversal of the level adjustment the step size was halved until the step size of 1 dB was reached. The mean of the signal level at the last six reversals was calculated and regarded as the masked threshold value. For each stimulus configuration and subject, four masked threshold values were measured. The mean of these values was calculated and taken as the final threshold. For the model predictions, the identical procedure and the same alternative-forced-choice (AFC) framework as in the experiments were used.

IV. EXPERIMENT 1: CMR WITH FOUR FLANKING BANDS

A. Rationale

The first experiment was designed to investigate true across-channel CMR, where within-channel processing does not contribute. Four flanking bands with a spectral separation of one octave were used such that within-channel contributions to CMR can be assumed to be negligible at the (medium) sound pressure levels used in this experiment.

B. Stimuli

The signal was a 1000–Hz pure tone. The masker consisted of five bands of noise which were centered at 250, 500, 1000, 2000 and 4000 Hz, thus covering a frequency range of four octaves. Signal and masker had the same duration of 187.5 ms; 20–ms raised-cosine ramps were applied to the stimuli. Signal threshold was measured as a function of the bandwidth of the masker, which was 25, 50, 100 or 200 Hz. The masker bands were generated in the time domain, transformed to the frequency domain by Fourier transform where they were restricted to the desired bandwidth, and finally transformed back to the time domain by inverse Fourier transform. In the reference condition, the envelopes of the five bands were uncorrelated with each other. In the comodulated condition, the on-frequency noise masker was shifted to the center frequencies of the flanking bands in the Fourier domain, such that the envelopes of the different bands were fully correlated with each other. The presentation level of each of the maskers was 60 dB sound pressure level (SPL).

C. Results

Figure 3 shows the results of the experiment. Masked thresholds are plotted as a function of the masker bandwidth. The open symbols represent the experimental data, averaged across subjects. The circles and squares show the results for the uncorrelated and comodulated conditions, respectively. The right panel of Fig. 3 shows the amount of CMR, i.e., the difference between the uncorrelated and comodulated thresholds. There is a significant CMR effect of 4–5 dB for the

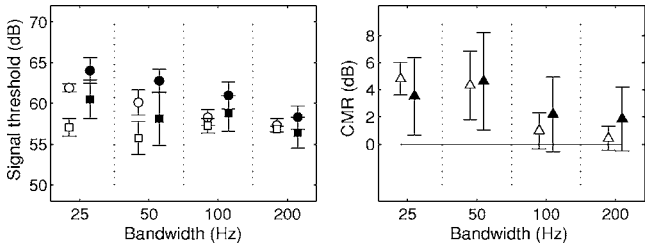


FIG. 3. Left panel: Detection thresholds for the 1–kHz tone in the presence of five noise bands as a function of the bandwidth of the noises. Open symbols indicate average experimental data and filled symbols show simulation results. Circles and squares represent results for the uncorrelated and comodulated conditions, respectively. Right panel: CMR effect for the conditions of the left panel.

small noise bandwidths of 25 and 50 Hz [one-way analysis of variance (ANOVA): $F(1,22)=38.59$, $p<0.001$ and $F(1,22)=32.18$, $p<0.001$], while no significant CMR was found for the larger bandwidths of 100 and 200 Hz [one-way ANOVA: $F(1,22)=1.67$, $p=0.21$ and $F(1,22)=0.02$, $p=0.89$] where statistical significance here and in the following is defined as having $p<0.01$. Thus, even though four flanking bands were used, the obtained CMR is relatively small compared to the results typically found with narrow spacing between the signal and flanking bands (see Experiment 2) or in the band-widening CMR paradigm (see Experiment 3). The results are consistent with results from previous studies (e.g., Moore and Emmerich, 1990), showing that CMR is restricted to narrowband noises with bandwidth smaller than 50 Hz. This indicates that across-channel CMR is a phenomenon that occurs only when the masker is dominated by relatively slow envelope fluctuations. The modulation spectrum of bandpass noise is directly related to the bandwidth of the noise (e.g., Lawson and Uhlenbeck, 1950; Dau *et al.*, 1997a). The rate of modulations will range up to the bandwidth of the noise, Δf.

The filled symbols in Fig. 3 show the simulations obtained with the processing model described in Sec. II. The simulations represent average thresholds of ten repetitions for each experimental condition. The model predicts slightly elevated overall thresholds (2–3 dB) and larger standard deviations in comparison to the empirical data. For the bandwidths 25 and 50 Hz, the model predicts a significant mean CMR effect of about 4 dB [one-way ANOVA: $F(1,18)=15.38$, $p<0.001$ and $F(1,18)=16.91$, $p<0.001$, respectively]. It does not produce a significant amount of CMR for the 100 and 200 Hz bandwidths [one-way ANOVA: $F(1,18)=6.48$, $p=0.02$ and $F(1,18)=6.29$, $p=0.02$].

D. Model analysis

The following describes how the EC mechanism affects the signal processing of the stimuli in the model. Since the EC process typically leads to a lower threshold in the comodulated condition compared to the uncorrelated condition, this should be reflected in the model's internal representations of the stimuli. As an example, the upper left panel of Fig. 4 shows the internal representation of a single 25-Hz-wide (comodulated) noise masker centered at 1 kHz. The outputs of the modulation filters are shown separately in

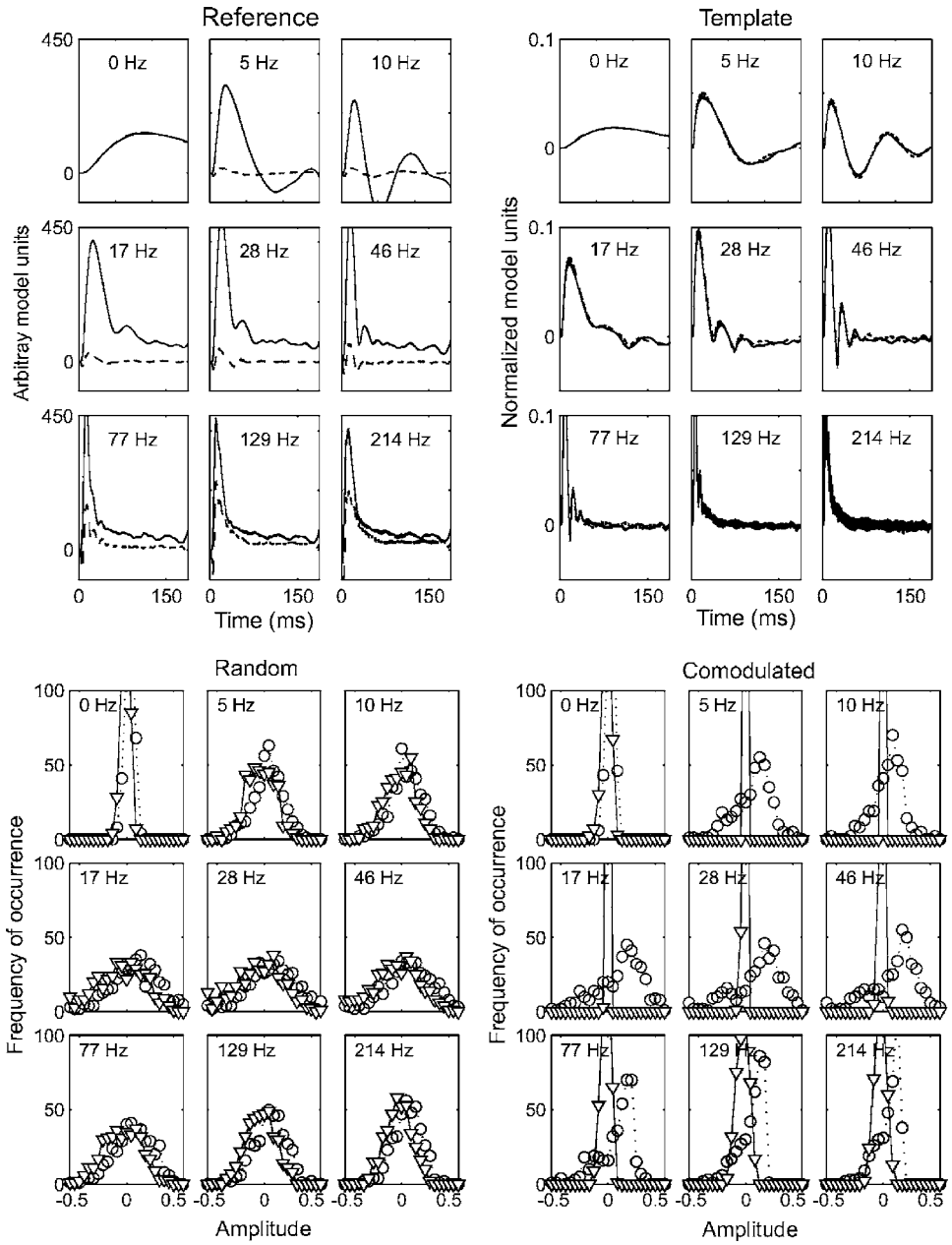


FIG. 4. Simulated internal representations at the output of the modulation filters (indicated by the center frequencies in the subpanels) in the on-frequency (peripheral) channel. Solid curves show outputs without EC process, dashed curves show results after the EC process. Left upper panel: Internal representation of (modulated) noise alone (i.e., no signal was added). Right upper panel: Internal representation of the template, i.e., the normalized difference between noise plus supra-threshold signal representation and noise alone representation. The lower panels show histograms of the cross-correlation coefficients between the noise-alone representation and template (triangles, solid line), and between the noise-plus-actual-signal representation and template (circles, dotted line), for the same individual modulation filters as considered in the top panels. This is shown for the random condition (left) and the comodulated condition (right), with EC mechanism applied.

the subpanels, including the modulation low-pass filter (indicated as 0 Hz), and the bandpass filters tuned to 5, 10, 17, 28, 46, 77, 129, and 214 Hz. The solid curves show the output obtained without EC process, i.e., when using the original model's (Dau *et al.*, 1997b) preprocessing. The dashed curves show the output when the EC process was included, i.e., after subtracting the average activity of the four flanking bands from the on-frequency band. As expected, the output representation (for the comodulated noise bands) after the EC process is reduced in amplitude compared to the result without the EC process. Note that modulation channels tuned to frequencies higher than the bandwidth of the noise (25 Hz) are activated as well, mainly reflecting the response to the onset of the adapted envelope of the stimulus.

As described in Sec. II and in previous publications (Dau *et al.*, 1996, 1997a), in the simulations, the internal representation of the noise is subtracted from the internal representation (either noise alone or signal plus noise) of each of the three intervals and then cross correlated with the template. The template represents the normalized difference between the internal representation of the noise plus *supra*-threshold signal and the noise-alone representation. The upper right panel of Fig. 4 shows the model's template using the same 25-Hz-wide noise (as used for the illustration of the reference) to which a supra-threshold 1–kHz tone was added. As for the reference representation, the individual modulation filter outputs are indicated in the subpanels. In the case of the template, there is essentially no difference between the situation with and without EC process since the internal representation of the template is dominated by the presence of the signal.

In order to evaluate the function of the EC mechanism, the two lower panels of Fig. 4 show a statistical analysis of the cross correlation between noise-alone representation and template (triangles), and between noise-plus-actual-signal representation and template (circles) including the EC

mechanism in the processing. The histograms of the cross-correlation coefficient are shown for the output of the same individual modulation filters as considered in the top panels. The "actual" signal level was chosen to be the simulated signal level at threshold (from Fig. 3, random condition). For the template, the same supra-threshold level (85 dB) was used as in the simulations. The lower left panel shows the results for the random noise condition at the output of the EC process. Since the signal level was chosen to be at detection threshold, the distributions are just separable (in terms of signal detection theory). The right panel shows the corresponding results for the comodulated condition. Here, the EC mechanism causes a strong sharpening of the distribution of correlations in the reference interval while the distributions in the signal interval remain essentially unaffected. This corresponds to an increased sensitivity and a decreased detection threshold in the simulations in the comodulated condition relative to the random condition, and represents the "noise reduction" caused by the EC mechanism. Without the EC mechanism, the histograms are similar in the random and comodulated condition.

The comparison of the histograms at the output of the different modulation filters suggests that all modulation filters contribute to signal detection (also those tuned to modulation frequencies higher than the noise bandwidth of 25 Hz). In other words, the decision in the model does not seem to be based on the activity at the output of only one or a few particular modulation filters. This is different from the situation in conditions of within-channel CMR, at least in the framework of the current model, where modulation cues like beatings between on-frequency and flanker bands components become effective and activate specific modulation filters in the signal interval (see the corresponding analysis in Experiment 2). In the EC model, a supra-threshold signal does not produce any specific modulation pattern that could be used as cue. The EC mechanism therefore does not lead to an enhancement of specific cues which would be reflected by different templates for the same condition with or without EC mechanism. The EC mechanism rather suppresses the noise fluctuations in the modulation filters, thereby enhancing signal detection.

Since the outputs of all bandpass modulation filters contribute to the function of the EC mechanism in the model, the question remains whether a modulation filterbank is necessary for the occurrence of CMR. To address this question, additional simulations were carried out with alternative modulation filtering stages: (i) A process referred to as "DC/AC" which separates the dc component of the Hilbert envelope spectrum from the remaining (ac) spectrum, (ii) a combination of a second-order Butterworth low-pass and a high-pass filter with cutoff frequencies of 2.5 Hz, referred to as "LH," (iii) a combination of the same low-pass filter at 2.5 Hz combined with a single bandpass filter centered at 5 Hz with a bandwidth of 5 Hz, referred to as process "LB5," and (iv) the same as (iii) but with a bandpass filter tuned to 50 Hz and a bandwidth of 25 Hz ($Q=2$; referred to as "LB50"). The EC process was applied to the ac-coupled output in dc/ac, the output of the high-pass filter in LH, and the output of the single bandpass filters in LB5 and LB50,

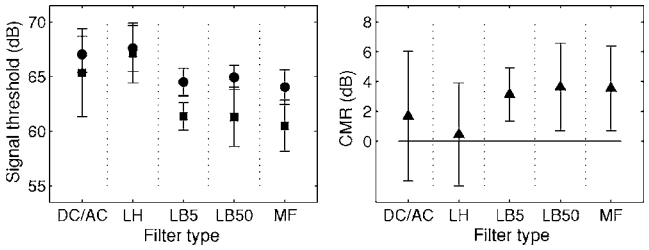


FIG. 5. Left: Signal thresholds obtained with the filter types dc/ac, LH, LB5, LB50, as defined in the main text and the complete modulation filterbank (MF). Circles and squares show results for random and comodulated noise, respectively. Right: Amount of CMR for the different filter types.

respectively. Figure 5 (left panel) shows the corresponding predictions obtained with the different processing schemes for the random and the comodulated noise conditions using the same symbols as in Fig. 3. The right panel shows the amount of CMR for the different schemes. The result obtained with the complete modulation filterbank, referred to as "MF," was replotted from Fig. 3 for direct comparison.

The DC/AC and LH processes do not produce any CMR [one-way ANOVA: $F(1,18)=1.51$, $p=0.24$ for DC/AC, $F(1,18)=0.16$, $p=0.68$ for LH]. In contrast, the processing of LB5 and LB50 produces a significant CMR effect of about 4 dB [one-way ANOVA: $F(1,18)=30.96$, $p<0.001$ and $F(1,18)=15.4$, $p<0.001$] which corresponds to the prediction obtained with the complete modulation filterbank MF [one-way ANOVA: $F(1,18)=38.59$, $p<0.001$]. Thus, within the model, across-channel CMR can only be produced if the stimulus after peripheral filtering, envelope extraction and adaptation is actually processed by frequency-selective (modulation) filters, whereby each individual filter would already be sufficient to produce significant CMR. The effect, however, disappears if only one broad (5–150 Hz) modulation bandpass filter is considered (not shown explicitly). The reason for the behavior in the model is that the input to the modulation filtering process, the adapted envelope, shows an onset response. This onset produces an excitation also at higher modulation frequencies. The EC process is only effective if the output of the modulation filtering process leads to a reasonable correlation between the flanking band and the signal band representations. This is only the case after (modulation) bandpass filtering, and cannot be obtained for the "broadband" schemes DC/AC and LH considered above. It is not clear, of course, to what extent the mechanisms in the real system are related to the ones proposed here on the basis of the model. The intention of the above analysis was to elucidate the functioning of the EC process of the proposed model.

In summary, the data from Experiment 1 confirm results from previous studies that across-channel processing in CMR is robust but small (even when several flanking bands are involved). Across-channel CMR is only observable at small bandwidths (below about 50 Hz), i.e., when the envelope fluctuations inherent in the stimuli are relatively slow. The simulations indicate that across-channel CMR can be accounted for quantitatively if an EC-like mechanism is introduced at the output of a modulation frequency-selective process.

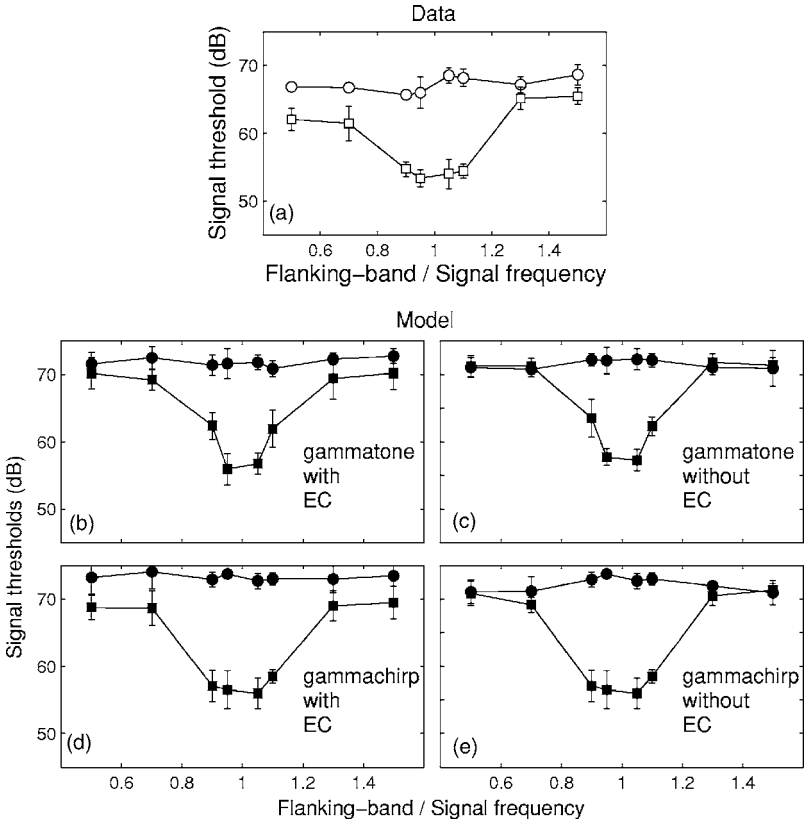


FIG. 6. (a) Measured data averaged across subjects. Signal threshold for a 2–kHz tone in 25–Hz wide noise as a function of the spectral separation between on-frequency band and flanking band. Circles and squares show results for random and comodulated noise, respectively. (b) Predictions with the EC model shown in Fig. 1, using gammatone filters as the peripheral filtering stage. (c) Predictions with the same model, but with EC process switched off. (d) Predictions as in (b) but with gammachirp filters. (e) Predictions with gammachirp filters, but with EC process switched off.

V. EXPERIMENT 2: CMR WITH ONE FLANKING BAND VARYING IN FREQUENCY

A. Rationale

This experiment investigates the transition between conditions where exclusively across-channel mechanisms determine CMR and those where primarily within-channel mechanisms generate CMR. Only one flanking band was used here, as in the study by Schooneveldt and Moore (1987). The amount of CMR was measured and simulated as a function of the spectral separation between the flanking and the on-frequency band. While for large separations of one octave or greater, CMR cannot be expected to exceed 2–4 dB, masking releases of about 14 dB and higher were observed in previous studies for separations of less than 1/10 octave where within-channel processing provides the most effective detection cues (Schooneveldt and Moore, 1987). A successful model of CMR needs to account for both within- and across-channel components.

B. Stimuli

The stimuli were similar to some of those used in Schooneveldt and Moore (1987). The signal was a 2000 Hz tone. The on-frequency masker was a 25-Hz-wide band of noise centered at the signal frequency. The flanking band had the same bandwidth as the on-frequency band and was centered at 1000, 1400, 1800, 1900, 2100, 2200, 2600 or 3000 Hz, corresponding to frequency ratios between flanking band and on-frequency band of 0.5, 0.7, 0.9, 0.95, 1.05, 1.1, 1.3, and 1.5. In contrast to the study by Schooneveldt and Moore (1987), the flanking band was not presented directly at the signal frequency or very close to it. The two noise bands were either uncorrelated or comodulated. As in Schooneveldt and Moore (1987) each band was produced by multiplying a sinusoid at the center frequency with a low-pass noise with a cutoff frequency of 12.5 Hz. In the comodulated condition, the noise bands were produced by multiplying the different sinusoids with an identical low-pass noise whereby a new noise was generated for each interval. Each band had an overall level of 67 dB SPL.

C. Results and model analysis

Panel (a) of Fig. 6 shows average data for the uncorrelated (open circles) and the comodulated (open squares) conditions. The signal threshold is plotted as a function of the ratio between flanking-band and signal frequency. The difference in threshold between uncorrelated and comodulated conditions, i.e., the amount of CMR, reaches 12–14 dB when flanker and signal frequency are close to each other (with ratios between 0.9 and 1.1). For large separations between on-frequency and flanking band, the data show a slight asymmetry: CMR of 3–4 dB in the presence of the high-frequency flankers and 5–6 dB for flanking bands presented at low frequencies. The data agree well with the results of Schooneveldt and Moore (1987).

Panel (b) of Fig. 6 shows the simulations obtained with the present model. As described in Sec. II, the EC mechanism was applied in all filters that overlap less than 5% with the on-frequency gammatone filter, i.e., in all channels except the two closest ones on both sides of the on-frequency channel. In this particular experiment, this means that the flanker bands were maximally contributing to the cancellation term of the EC process at frequency ratios of 0.5, 0.7,

1.3, and 1.5. The model accounts for the relatively flat threshold function obtained in the uncorrelated condition. For flanking-band frequencies close to the signal frequency (at the frequency ratios 0.95 and 1.05), the model predicts a large amount of CMR that corresponds to that found in the experimental data. This component depends on beating of the carrier frequencies of the on-frequency and flanking bands. In the model this can be accounted for by the processing within the (peripheral) channel tuned to the signal frequency. The model detects changes in the envelope statistic due to the addition of the signal to the on-frequency band (see Verhey *et al.*, 1999). This is effective for the comodulated condition while it does not provide additional detection cues in the uncorrelated condition. At very low and very high flanking band frequencies, the model predicts an average amount of CMR of about 3 dB which agrees well with the data at the high flanker frequencies but is slightly less than the measured effect at the low flanker frequencies. The simulated 3 dB effect is the result of the EC mechanism in the model as can be seen from direct comparison with the results obtained without EC circuit, shown in panel (c) of Fig. 6. As expected, without across-channel processing, no CMR is predicted at the large frequency separations between the on-frequency and the flanker band.

While certain aspects of the data can be described satisfactorily by the model, some other aspects cannot. First, the simulated threshold function for the comodulated condition increases too steeply with increasing spectral distance from the signal. Second, the predicted amount of CMR for the lowest flanker frequencies is smaller than in the data. The reason for these discrepancies might be related to the shape of the magnitude transfer function of the peripheral filters used in the simulations. The gammatone filters are symmetrical on a linear frequency scale. However, it has been demonstrated that below its center frequency, the skirt of the human auditory filter broadens substantially with increasing stimulus level, and above its center frequency the skirt sharpens slightly with increasing level (Lutfi and Patterson, 1984; Moore and Glasberg, 1987). In order to illustrate effects of frequency selectivity on CMR in the framework of the current model, additional simulations were carried out using gammachirp filters (Irino and Patterson, 1997). The gammachirp filter has an asymmetric magnitude transfer function, and the degree of asymmetry in this filter is associated with stimulus level. The gammachirp filter was shown to provide a very good fit to human notched-noise masking data. Its impulse response is well defined and includes only one parameter more than the gammatone filter (see Eq. 2, in Irino and Patterson, 1997). In the present study, the impulse responses of the gammachirp filters were calculated for a level of 67 dB SPL. Here, as a simplification, the simulations were run with selected gammachirp filters tuned to the on-frequency band and the flanking band, respectively. A complete gammachirp filterbank with well defined level-dependent overlap has not been developed yet. As in the previous simulations with gammatone filters, the EC process was applied when the overlap between the off-frequency channel and the signal channel was below 5% which was only the case for the four outer data points (frequency ratios 0.5, 0.7, 1.3, and 1.5). All other model parameters were kept the same as in the simulations with gammatone filters. The results are shown in panel (d) of Fig. 6.

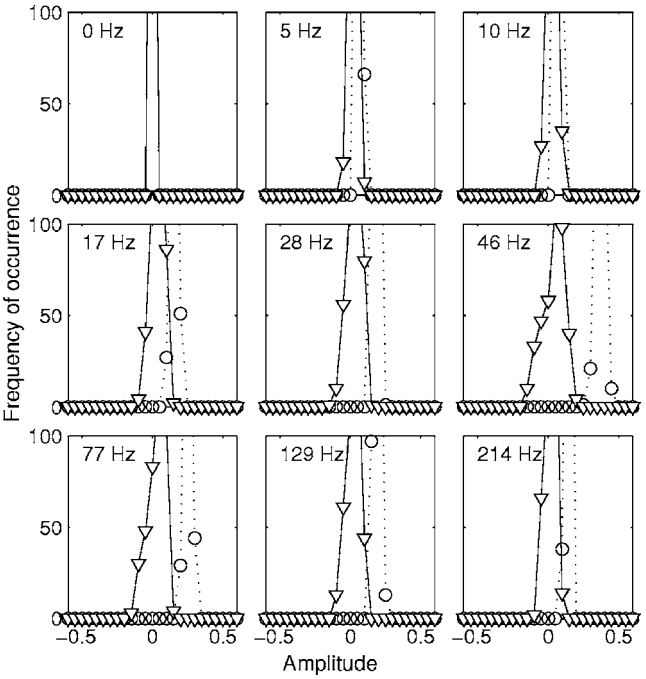


FIG. 7. Histograms of the cross-correlation coefficients at the output of nine modulation filters in an exemplary condition of Experiment 2 with 50 Hz separation between the on-frequency and flanking bands. Correlations for the reference (triangles, solid line) and reference plus signal (circles, dashed line) are shown for comodulated noise bands. The center frequency of the modulation filter is indicated within each panel. For the output of the modulation filter close to 50 Hz, the mean of the distributions is most different and the distributions are most separable in terms of signal detection.

The simulations with gammachirp filters account for many aspects of the experimental data. Due to the broader bandwidth of the gammachirp filter compared to the gammatone filter, within-channel cues become effective for a larger range of flanking-band frequencies. The plateau of low thresholds corresponds to that found in the data. At low and at high flanking-band frequencies, CMR amounts to 3–4 dB due to the EC processing in the model. However, the introduction of the gammachirp filter does not account for the slight asymmetry observed in the measured data, even though the transfer functions of the individual filters have an asymmetric shape. The simulated pattern for the comodulated condition actually produces the same thresholds at both ends. Still, the overall correspondence with the data is high. For direct comparison, panel (e) of the same figure shows the corresponding predictions without EC process. All data points except for the four outer ones are replotted from panel (d), since no EC process was applied for the inner data points in panel (d). As for the simulations with gammatone filters without the EC process, no CMR was obtained at the largest spectral separations between flanking and on-frequency band.

In order to illustrate the importance of within-channel cues available in the conditions where on-frequency band and flanking band are close to each other, Fig. 7 shows a statistical analysis similar to that presented in Experiment 1.

Histograms of the cross correlation between noise-alone representation and template (triangles) and noise-plus-actual-signal representation and template (circles) are shown for the outputs of the individual modulation filters. An exemplary frequency separation between on-frequency band and flanking band of 50 Hz was used for illustration. It can be seen in Fig. 7 that signal detection is mainly based on information at the output of the modulation filter tuned to about 46 Hz. Here, the mean of the signal distribution is clearly larger than that of the noise distribution. Thus, the addition of the signal to the masker causes changes in the internal representation of the stimuli such that it can effectively be evaluated in one (or only a few) modulation filters in this given task. This detection cue is qualitatively different from that discussed in connection with the across-channel process where signal detection was mainly based on the sharpening of the noise distribution at the output of the EC process in all modulation filters.

The results of Experiment 2 thus support the hypothesis that CMR has (at least) two components. One is restricted to flanking band frequencies around the signal frequency. This component reflects the use of within-channel cues (beating), rather than across-channel cues. The other component does not depend strongly on flanking-band frequency, but rather on across-channel cues. This across-channel component of CMR amounts to about 3 dB. While this has been proposed in earlier studies (e.g., Schooneveldt and More, 1987), the present study tried to provide quantitative modeling to test explicitly the (relative) contributions of within- and across-channel processing.

VI. EXPERIMENT 3: CMR AS A FUNCTION OF THE MASKER BANDWIDTH

A. Rationale

The third experiment considered the "classical" band-widening experiment where the masker was centered at the signal frequency and signal threshold was obtained as a function of the bandwidth of the masker. In contrast to the two previous experiments, the band-widening experiment does not allow for a separation between within- and across-channel processes; within-channel contributions will always contribute to CMR, even for large masker bandwidths when many auditory filters are excited by the noise. Verhey *et al.* (1999) showed that a single-channel analysis, which uses only the information in one peripheral channel tuned to the signal frequency, quantitatively accounts for the main CMR effect in the band-widening experiment. This suggested that across-channel processes are not involved or not effective in this class of CMR experiments, even though several auditory filters are excited by the noise. This was directly investigated here with the extended model that includes an explicit across-channel process while it keeps the ability to process within-channel cues, as shown in Experiment 2.

B. Stimuli

The signal was a 300-ms-long, 2000–Hz pure tone. The masker was a band-limited noise centered at the signal frequency. The masker bandwidth was 50, 100, 200, 400, 1000 or 2000 Hz. The duration of the masker was 600 ms with 10–ms raised-cosine onset and offset ramps. The signal was temporally centered in the masker. Two types of maskers were used, as in the original experiments in Hall *et al.* (1984a). One was a random noise with irregular and independent envelope fluctuations in different frequency regions. The comodulated noise was a random broadband noise which was modulated in amplitude at an irregular, low rate, and then restricted to the desired bandwidth. A low-pass noise with a cutoff at 50 Hz was used as a modulator. Other studies have shown that for modulator bandwidths larger than 50 Hz, CMR decreases with increasing modulator bandwidth whereas it remains roughly constant for modulator bandwidth, below this value (Schooneveldt and Moore, 1987; Carlyon and Stubbs, 1989). The modulation resulted in fluctuations in the amplitude of the noise which were the same in different frequency regions. The spectrum level of the bandpass noise was 30 dB, corresponding to overall levels of 47–63 dB SPL for the 50–2000 Hz bandwidth range.

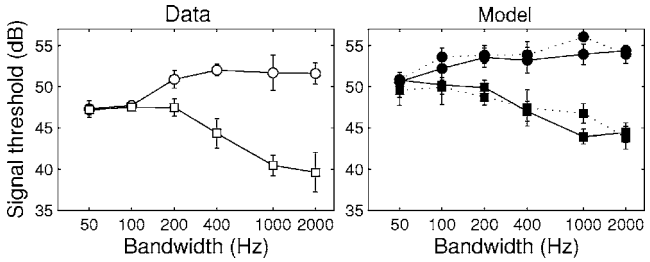


FIG. 8. Left panel: Average signal thresholds for four subjects are plotted as a function of the masker bandwidth in random noise (circles) and comodulated noise (squares). Right panel: Predicted signal threshold of the model when the EC mechanism is applied (dashed line) and when it is not applied (solid line). The modulator bandwidth was 50 Hz and the signal frequency was 2000 Hz.

C. Results

Figure 8 shows the results of the band-widening experiment. The left panel shows the experimental data, averaged across subjects. The signal threshold is plotted as a function of the masker bandwidth, for random noise (open circles) and comodulated noise (open squares). Consistent with the results from the earlier studies, for the random noise, the masked threshold first increases as the masker bandwidth is increased. Beyond a certain bandwidth (200 Hz in this case), the threshold no longer increases, but remains roughly constant. The increase of the threshold is caused by the fact that, up to the critical bandwidth, more noise passes through the auditory filter centered at the signal frequency, while beyond the critical bandwidth, the added noise falls outside the passband of the auditory filter. In contrast, for the comodulated noise, the threshold first stays constant and then decreases as the bandwidth is increased beyond about 200 Hz. The amount of CMR, defined as the difference in threshold between the random and comodulated conditions, is 12 dB for the largest bandwidth (2000 Hz).

The right panel of Fig. 8 shows the corresponding model predictions. For direct comparison, simulations are shown with EC process (dashed line) and without EC process (solid line). The two model versions essentially produce the same

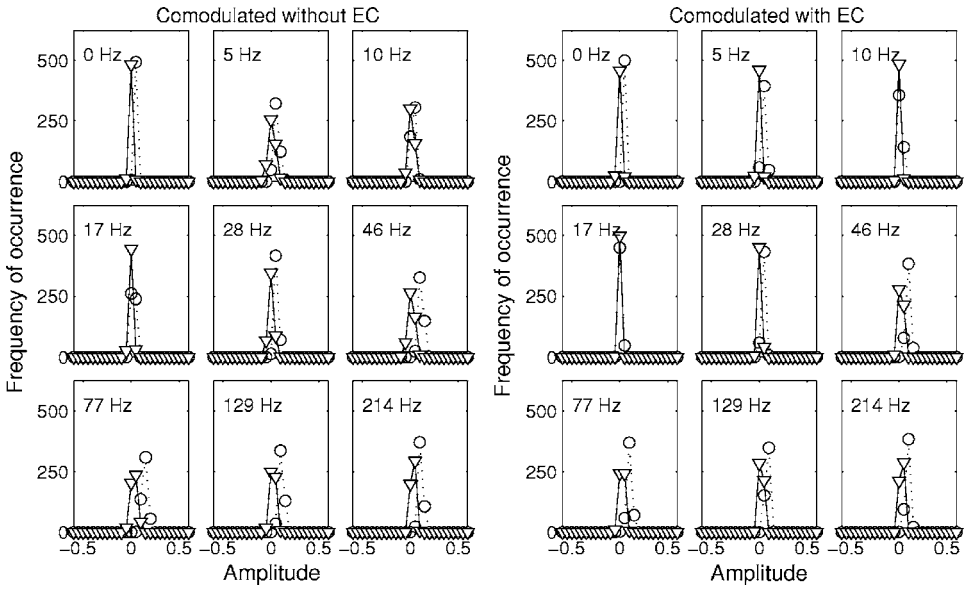


FIG. 9. Histograms of the cross-correlation coefficients at the output of nine modulation filters for the comodulated conditions of Experiment 3 with a noise bandwidth of 2000 Hz. Left panel: Reference alone (triangles, solid line) and reference plus signal (circles, dashed line) for comodulated noises without EC process. Right panel: Reference alone (triangles, solid line) and reference plus signal (circles, dashed line) for comodulated noises with EC process included.

results. Thus, the across-channel processing does not generate any change in the overall amount of CMR in the framework of this model, not even at the largest masker bandwidths where several auditory channels are excited. Figure 9 shows the statistical analysis of the decision variable in the simulations, as in the first two experiments.

The comodulated condition with the broadest noise bandwidth (2000 Hz) was considered for illustration with and without EC mechanism. At this bandwidth, the observed amount of CMR is maximal (12 dB). The analysis was carried out at a signal level of 55 dB which is about 10 dB above the simulated threshold in the comodulated condition. The left panel shows the distribution of the cross-correlation between noise-alone representation and template (triangles) and for the signal-plus-noise representation and template (circles) at the output of the single (peripheral) channel tuned to the signal frequency (single-channel analysis). It can be seen that there is a separation between the two distributions at several modulation filter outputs. Since the bandwidth of the noise (also after peripheral filtering) is larger in this experimental condition than in the previous experiments, the variability of the envelope amplitude fluctuations is smaller, leading to the relatively sharp distributions. The right panel shows the analysis including the across-channel process in the model, i.e., a multi-channel simulation was carried out in this case where the cancellation term in the EC process was derived from the off-frequency channels. The envelope correlation across the different peripheral channels is *not sufficient* to effectively increase the signal-to-noise ratio at the output of the EC process in the model. The EC process therefore does not contribute to signal detection in this type of experiment in the framework of the model.

These results therefore support the hypothesis that CMR obtained in the band-widening paradigm is strongly dominated by within-channel processing and is not a result of across-channel processing.

VII. OVERALL DISCUSSION

A. Within- versus across-channel processing

The modeling results of this study support the hypothesis that (at least) two mechanisms are contributing to what has been defined as CMR. The present model allows a distinction to be made between these two contributions. The simulations strongly support that one of the processes is based on within-channel mechanisms. Signal detection is based on the changes of the internal representation of the stimuli at the output of individual auditory filters—without the need for explicit across-frequency processing. The addition of the signal to the comodulated masker typically changes the (envelope) statistics of the stimuli significantly, while the changes are much smaller or absent in the case of random noise maskers (Schooneveldt and Moore, 1987; Verhey *et al.*, 1999). CMR resulting from within-channel contributions can be up to about 15 dB depending on the specific condition, and modulations (for example, resulting from beatings between signal and masker components) up to several hundred Hz can serve as a cue for signal detection. A pre-requisite for accounting for the full range of within-channel contributions to CMR is therefore a high sensitivity of the model to amplitude modulations (see also Verhey *et al.*, 1999), as is the case for the modulation filterbank used in the present framework. Specifically, the modeling results suggest that a few individual modulation filters (at the output of the single peripheral channel at the signal frequency) can process the changes in the internal representation of the stimuli effectively.

The other form of CMR is based on true across-channel processing. This effect is also robust but relatively small (2–4 dB) and becomes only effective when narrowband noises with bandwidths below about 50 Hz are presented, i.e., when the envelope fluctuations of the noises vary relatively slowly. The EC model described in the present study makes specific assumptions about how envelope information at the output of different auditory channels might be pro-

cessed. The EC process was assumed to take place at the output of each modulation bandpass filter. The effect of the EC process is that the variance of the external noise (originating from the masker) at the level of the internal representations after the EC process is reduced in the case of the comodulated noise condition. This leads to improved signal detection compared to the random noise condition. In the framework of the model, the detection cue is thus qualitatively very different from the situation where within-channel processing determines CMR.

It is clear that effects of nonlinear peripheral processing, such as the level-dependent auditory filter bandwidth, have an influence on the relative contributions of within- and across-channel processing to CMR. In fact, some of the effects that were considered as across-channel contributions in the past might become within-channel contribution with proper modeling of nonlinear filters. For example, at very high stimulus levels where the auditory filter bandwidth is markedly increased (compared to the gammatone filters used in the present study), it can be expected that even in conditions with very broad spacing between the on-frequency band and the flanking band(s), CMR might be dominated by within-channel contributions. Ernst and Verhey (2005) have shown that CMR over ranges of three octaves can be modeled as a suppression effect in a nonlinear single-channel model, using the dual resonance nonlinear filter (DRNL) model (Meddis *et al.*, 2001). In some of the conditions in their study, however, the level of the off-frequency flanker was much higher (up to 60 dB) than that of the on-frequency band. Although their results are not directly comparable to the experimental conditions used in the present study, it can be assumed that with proper modeling of the nonlinear auditory filters even more signal configurations that have been considered as across channel in the past might reveal a within-channel contribution. Our current definition of when across-channel processing is applied to a particular filter is based on the amount of overlap of its transfer function with that of the signal channel. This definition might be general enough to also apply to filters of different or varying shapes and to nonlinear filters; the approach was successful when analyzing the results of Experiment 2, where individual gammachirp filters were considered. This, however, needs to be further investigated using a complete filterbank of filters with different shape or a nonlinear filterbank such as, e.g, a bank of DRNL filters or a gammachirp filterbank.

The observation that two conceptually different mechanisms define CMR is compatible with the results from studies on effects of auditory grouping on CMR (Grose and Hall, 1993; Dau *et al.*, 2005). When widely spaced flanking bands were used (as in the first experiment of the present study), CMR effects could be eliminated completely by introducing a gating asynchrony between the on-frequency masker and the flanking bands, by introducing precursor flanking bands, and by introducing following flanking bands. Due to the large spacing (and the relatively low presentation levels), only across-channel processes contributed to CMR. In contrast, using narrowly spaced flanking bands with 1/6-octave spacing (similar to the conditions with close frequency spacings in Experiment 2), CMR was not affected by any of the stimulus manipulations. It was therefore suggested that (i) the within-channel mechanisms in CMR might be peripheral (brainstem level or below) in nature and therefore not susceptible to manipulation by auditory grouping constraints, and that (ii) the "slower" across-channel processing that is strongly dependent on auditory grouping constraints, might be of more central origin (Dau *et al.*, 2005). The model investigated in the present study is not able to identify or extract auditory objects based on comodulation. A more advanced version of the model might apply basic concepts of computational auditory scene analysis (Bregman *et al.*, 1990), where the EC-process would be switched on or off depending on the current spectro-temporal acoustical context.

B. Correlation with physiological CMR results

Even though a large number of studies have investigated CMR from a psychophysical perspective, little is known of its underlying physiological mechanisms (see, e.g., Verhey *et al.*, 2003, for a review). A few studies have addressed physiological mechanisms of across-frequency processing by estimating signal-detection thresholds from the recordings of single- and multi-unit recordings in CMR-like paradigms. Several stages of processing along the auditory pathway were considered. Some studies intended to investigate across-channel processing but actually studied mostly within-channel cues due to the specific choice of the stimuli (e.g., Mott *et al.*, 1990). Nelken *et al.* (1999) investigated the response of neurons in the primary auditory cortex to noise of varying bandwidth. They found a correlate for CMR in the band-widening paradigm in the disruption of the neurons' envelope following response. For most of the neurons in the population, the envelope locking was degraded by the addition of the pure tone signal. Using statistical criteria to estimate signal detection threshold, Nelken *et al.* (1999) demonstrated that the suppression of envelope locking lowers the detection thresholds for the single tones when comparing the responses of modulated versus unmodulated noise bands.

When considering true across-channel CMR, two possible correlates have been discussed recently. In the primary auditory cortex (of the cat), Rotman *et al.* (2001) in another study used a stimulus centered on the best frequency of the neuron and added two flanking bands equally spaced at either side of the best frequency. They showed that a single unit in the auditory cortex can demonstrate a response consistent with CMR in the flanking band paradigm. The correlate of CMR was again found as a disruption of the envelope following response. Thus, it appears that CMR is coded at a relatively late stage of auditory processing (in the primary auditory cortex) which appears conceptually compatible with the psychophysical findings on grouping constraints on CMR. Their finding of very similar correlates for CMR in the two stimulus paradigms seems to differ from the modeling analysis discussed in the present study, which suggests very different mechanisms for the two processes.

A second physiological correlate of across-channel CMR has been suggested to be wideband inhibition at brainstem level (e.g., Pressnitzer *et al.*, 2001; Meddis *et al.*,

2002). Here, it has been suggested, based on physiological experiments with the flanking-band paradigm with deterministic maskers, that cochlear nucleus onset units provide wideband inhibition at the level of the brainstem onto narrowband units in the ventral cochlear nucleus, and that this wideband inhibition could provide a possible physiological basis for a potential EC model of CMR (for details about hypothetical neural circuits underlying CMR in the cochlear nucleus, see Pressnitzer *et al.*, 2001; Verhey *et al.*, 2003). A problem with such a neural correlate at the level of the brainstem might be the perceptual findings in the context of auditory grouping which make it unlikely that across-channel CMR can be accounted for by processing in the auditory brainstem and below.

A very promising way to fully understand the physiological mechanisms underlying CMR might be to study the correlation between neural responses and performance in the same species. Such an investigation was undertaken by Langemann and Klump (2001) and Nieder and Klump (2001) using the starling. Nieder and Klump (2001) investigated across-channel CMR with the flanking band paradigm, but used 100-Hz-wide on-frequency and flanking bands amplitude modulated at 10 Hz. They showed that neural detection threshold was lowest when the probe tone was positioned in a dip of the masker envelope. They concluded that their multi-unit recordings in the auditory forebrain of the starling can be compared to the behavioral results in the same species. It would be interesting to specifically study the three basic paradigms of the present study in the same animal model both behaviorally and physiologically to learn more about the potential correlates of the different mechanisms underlying CMR.

C. Limitations of the current modeling approach

This study proposed an auditory signal processing model that accounts both for within-channel and across-channel processing in CMR. However, only three basic experiments were considered in order to evaluate the model and to discuss the main principles of auditory processing underlying CMR—in the framework of the model. A number of experimental conditions have been investigated in previous CMR studies, which have not been considered directly in the present study. These studies investigated in much more detail effects of signal frequency (e.g., Schooneveldt and Moore, 1987), masker spectral width (Haggard *et al.*, 1990; Hall and Grose, 1990) and masker spectral level (Moore and Shailer, 1991; Bacon *et al.*, 1997; Cohen, 1991; Hall, 1986; McFadden, 1986), the influence of the envelope statistic of the masker modulator (e.g., Eddins and Wright, 1994; Grose and Hall, 1989; Moore *et al.*, 1990; Hicks and Bacon, 1995), the effect of modulation frequency and modulation depth (Carlyon and Stubbs, 1989; Hall *et al.*, 1996; Lee and Bacon, 1997; Bacon *et al.*, 1997; Verhey *et al.*, 1999; Eddins, 2001), effects of flanking band number and flanking band level (e.g., Hatch *et al.*, 1995; Schooneveldt and Moore, 1987) and other effects. The current version of the model does not include a nonlinear peripheral filtering stage and therefore cannot account for level-dependent cochlear compression and effects associated with it such as level-dependent frequency tuning and suppression. While suppression does not seem to play a role in CMR with the level combinations in the present study (Hall *et al.*, 1984b; Hchooneveldt and Moore, 1987), effects of frequency selectivity certainly do, as was also shown in the present study. However, while corresponding modifications will change the details of the modeling outcomes, the main principles and implications discussed in the present study are expected to remain valid.

A further potential generalization of the model would be to include effects of dichotic presentation of flanker bands on CMR. The size of across-ear effects on CMR (2–3 dB) typically corresponds to that found in monaural across-channel CMR with one flanking band. The idea would be to apply the "central" EC mechanism to the stimuli after consideration of the inputs coming from the two ears. A binaural signal processing model based on the model by Breebaart *et al.* (2001a, b, c) but including a modulation filterbank stage is currently under development.

VIII. SUMMARY AND CONCLUSIONS

- A monaural auditory processing model was proposed that accounts for comodulation masking release (CMR) obtained in perceptual listening tests. The model distinguishes between contributions to CMR from within-channel processing and those resulting from explicit across-channel processing. For the across-channel process, an equalization-cancellation stage was assumed, conceptually motivated by models on binaural processing.
- The model accounts for the main findings in three critical experiments of CMR: (i) CMR with widely spaced flanking bands (where only across-channel processing contributes), (ii) CMR with one flanking band varying in frequency (where within-channel processing dominates at small separations while across-channel processing takes over at large separations), and (iii) CMR obtained in the classical band-widening experiment (where within-channel processing can never be eliminated).
- The simulation results support the earlier hypothesis that (at least) two different processes can contribute to CMR. The within-channel contributions can be as much as 15 dB and is caused by changes of the envelope statistics of the stimulus due to the addition of the signal to the (comodulated) masker—at the output of the auditory filter tuned to the signal frequency. The across-channel process is robust but small (about 2–4 dB) and only observable at small flanker bandwidths (below about 50 Hz).
- Specifically, in the classical band-widening experiment, which originally was used to define CMR as an across-channel process, the simulation results suggest that across-channel processing is not effective, not even at the largest noise bandwidth considered (2000 Hz) where several auditory filters are excited. CMR in this type of stimulus paradigm is dominated by within-channel processes.
- The current implementation of the model does not include a nonlinear, level-dependent cochlear filtering stage which limits its applicability in some of the experimental conditions tested in previous CMR studies. The effect of a level-

dependent frequency selectivity was investigated in one of the experiments of the present study using gammachirp instead of gammatone filters. A more complete implementation in the framework of the whole model is currently under investigation. Overall, the proposed model might provide an interesting framework for the analysis of fluctuating sounds in the auditory system.

ACKNOWLEDGMENTS

We thank our colleagues at the Centre for Applied Hearing Research for many interesting discussions, and John Grose and two anonymous reviewers for their constructive and very helpful criticism. This work was supported by the Danish Research Foundation and the Deutsche Forschungsgemeinschaft (DFG; SFB/TRR 31).

Bacon, S. P., Lee, J., Peterson, D. N., and Rainey, D. (**1997**). "Masking by modulated and unmodulated noise: Effects of bandwidth, modulation rate, signal frequency, and masker level," J. Acoust. Soc. Am. **101**, 1600–1610.

Berg, B. G. (**1996**). "On the relation between comodulation masking release and temporal modulation transfer functions," J. Acoust. Soc. Am. **100**, 1013–1023.

Bernstein, L. R., and Trahiotis, C. (**1996**). "On the use of the normalized correlation as an index of interaural envelope correlation," J. Acoust. Soc. Am. **100**, 1754–1763.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition. I.Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition. I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Bregman, A. S., Liao, C., and Levitan, R. (**1990**). "Auditory grouping based on fundamental frequency and formant peak frequency," Can. J. Psychol. **44**, 400–413.

Buss, E., and Hall, J. W. (**1998**). "The role of auditory filters in comodulation masking release (CMR)," J. Acoust. Soc. Am. **103**, 3561–3566.

Buss, E., Hall, J., and Grose, J. (**1998**). "Change in envelope beats as a possible cue in comodulation masking release (CMR)," J. Acoust. Soc. Am. **103**, 1592–1597.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Carlyon, R., and Stubbs, R. (**1989**). "Detecting single-cycle frequency modulation imposed on sinusoidal, harmonic, and inharmonic carriers," J. Acoust. Soc. Am. **85**, 2563–2574.

Cohen, M. F., and Schubert, E. D. (**1987**). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **81**, 721–723.

Cohen, M. (**1991**). "Comodulation masking release over a three octave range," J. Acoust. Soc. Am. **90**, 1381–1384.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the "effective-Bsignal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

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**, 2893–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrowband noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

Dau, T., Ewert, S. D., and Oxenham, A. J. (**2005**). "Effects of concurrent and sequential streaming in comodulation masking release," in *Auditory Signal Processing -Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. MacAdams, and L. Collet (Springer, New York), 335–343.

Derleth, R. P., and Dau, T. (**2000**). "On the role of envelope fluctuation processing in spectral masking," J. Acoust. Soc. Am. **108**, 285–296.

Derleth, R. P., Dau, T., and Kollmeier, B. (**2001**). "Modeling temporal and compressive properties of the normal and impaired auditory system," Hear. Res. **159**, 132–149.

Domnitz, R. H., and Colburn, H. S. (**1977**). "Lateral position and interaural discrimination," J. Acoust. Soc. Am. **61**, 1586–1598.

Durlach, N. (**1960**). "Note on the equalization and cancellation theory of binaural masking level differences," J. Acoust. Soc. Am. **32**, 1075–1076.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Eddins, D. A., and Wright, B. A. (**1994**). "Comodulation masking release for single and multiple rates of envelope fluctuation," J. Acoust. Soc. Am. **96**, 3432–3442.

Eddins, D. A. (**2001**). "Measurement of auditory temporal processing using modified masking period patterns," J. Acoust. Soc. Am. **109**, 1550–1558.

Ernst, S. M. A., and Verhey, J. L. (**2005**). "Comodulation masking release over a three octave range," J. Acoust. Soc. Am. **91**, 998–1006.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Ewert, S. D., and Dau, T. (**2004**). "Internal and external limitations in amplitude-modulation processing," J. Acoust. Soc. Am. **116**, 478–490.

Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). "Spectro-temporal processing in the envelope-frequency domain," J. Acoust. Soc. Am. **112**, 2921–2931.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Green, D. M. (**1992**). "On the similarity of two theories of comodulation masking release," J. Acoust. Soc. Am. **91**, 1769.

Grose, J. H., and Hall, J. W. (**1989**). "Comodulation masking release using SAM tonal complex maskers: Effects of modulation depth and signal position," J. Acoust. Soc. Am. **85**, 1276–1284.

Grose, J. H., and Hall, J. W. (**1993**). "Comodulation masking release: Is comodulation sufficient?," J. Acoust. Soc. Am. **93**, 2896–2902.

Haggard, M. P., Hall, J. W., and Grose, J. H. (**1990**). "Comodulation masking release as a function of bandwidth and test frequency," J. Acoust. Soc. Am. **88**, 113–118.

Hall, J. W., and Grose, J. H. (**1990**). "Comodulation masking release and auditory grouping," J. Acoust. Soc. Am. **88**, 119–125.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984a**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984b**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Hall, J. W., Grose, J. H., and Hatch, D. R. (**1996**). "Effects of masker gating for signal detection in unmodulated and modulated bandlimited noise," J. Acoust. Soc. Am. **100**, 2365–2372.

Hall, J. W. (**1986**). "The effect of across-frequency differences in masking level on spectro-temporal pattern analysis," J. Acoust. Soc. Am. **79**, 781–787.

Hatch, D., Arne, B., and Hall, J. (**1995**). "Comodulation masking release (CMR): Effects of gating as a function of number of flanking bands and masker bandwidth," J. Acoust. Soc. Am. **97**, 3768–3774.

Hicks, M. L., and Bacon, S. P. (**1995**). "Some factors influencing comodulation masking release and across-channel masking," J. Acoust. Soc. Am. **98**, 2504–2514.

Irino, T., and Patterson, R. D. (**1997**). "A time-domain, level-dependent auditory filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412–419.

Langemann, U., and Klump, G. M. (**2001**). "Signal detection in amplitude-modulated maskers. I. Behavioural auditory thresholds in a songbird," Eur. J. Neurosci. **13**, 1025–1032.

Lawson, J. L., and Uhlenbeck, G. E., editors (**1950**). *Threshold Signals, Vol. 24 of Radiation Laboratories Series* (McGraw–Hill, New York).

Lee, J., and Bacon, S. P. (**1997**). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration," J. Acoust. Soc. Am. **101**, 3688–3693.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A., and Patterson, R. D. (**1984**). "On the growth of masking asymmetry with stimulus intensity," J. Acoust. Soc. Am. **76**, 739–745.

McFadden, D. (**1986**). "Comodulation masking release: Effects of varying the level, duration, and time delay of the cue band," J. Acoust. Soc. Am. **80**, 1658–1667.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). "A computational algorithm for computing nonlinear auditory frequency selectivity," J. Acoust. Soc. Am. **109**, 2852–2861.

Meddis, R., Delahaye, R., O'Mard, L., Summer, C., Fantini, D. A., Winter, I., and Pressnitzer, D. (**2002**). "A model of signal processing in the cochlear nucleus: Comodulation masking release," Acta Acust. (Stuttgart) **88**, 387–398.

Moore, B. C. J., and Emmerich, D. S. (**1990**). "Monaural envelope correlation perception, revisited: Effects of bandwidth, frequency separation, duration, and relative level of the noise bands," J. Acoust. Soc. Am. **87**, 2628–2633.

Moore, B. C. J., and Glasberg, B. R. (**1987**). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns," Hear. Res. **28**, 209–225.

Moore, B. C. J., and Shailer, M. J. (**1991**). "Comodulation masking release as a function of level," J. Acoust. Soc. Am. **90**, 829–835.

Moore, B. C. J., Hall, J. W., Grose, J. H., and Schooneveldt, G. P. (**1990**). "Some factors affecting the magnitude of comodulation masking release," J. Acoust. Soc. Am. **88**, 1694–1702.

Mott, J. B., McDonald, L. P., and Sinex, D. G. (**1990**). "Neural correlates of psychophysical release from masking," J. Acoust. Soc. Am. **88**, 2682–2691.

Nelken, I., Rotman, Y., and Yosef, O. B. (**1999**). "Responses of auditory-cortex neurons to structural features of natural sounds," Nature (London) **397**, 154–157.

Nieder, A., and Klump, G. M. (**2001**). "Signal detection in amplitude-modulated maskers. II. Processing in the songbird's auditory forebrain," Eur. J. Neurosci. **13**, 1033–1044.

Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by Moore B. C. J. (Academic, New York), 123–178.

Patterson, R. D., Nimmo-Smith, J., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE.

Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (**2001**). "Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus," J. Neurosci. **21**, 6377–6386.

Püschel, D. (**1988**). "Prinzipien der zeitlichen Analyse beim Hören (Principles of Temporal Processing in Hearing)," doctoral thesis, Universität Göttingen.

Richards, V. M. (**1987**). "Monaural envelope correlation perception," J. Acoust. Soc. Am. **82**, 1621–1630.

Rotman, Y., Bar-Yosef, O., and Nelken, I. (**2001**). "Relating cluster and population responses to natural sounds and tonal stimuli in cat primary auditory cortex," Hear. Res. **152**, 110–127.

Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. **82**, 1944–1956.

van de Par, S., and Kohlrausch, A. (**1998a**). "Comparison of monaural (CMR) and binaural (BMLD) masking release," J. Acoust. Soc. Am. **103**, 1573–1579.

van de Par, S., and Kohlrausch, A. (**1998b**). "Diotic and dichotic detection using multiplied-noise maskers," J. Acoust. Soc. Am. **103**, 2100–2110.

van de Par, S. (**1998**). "A comparison of binaural detection at low and high frequencies," doctoral thesis, Eindhoven University of Technology.

Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.

Verhey, J. L., Pressnitzer, D., and Winter, I. M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.

Verhey, J. L. (**2002**). "Modeling the influence of inherent amplitude fluctuation simultaneous masking experiments," J. Acoust. Soc. Am. **111**, 1018–1025.

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

Interaural fluctuations and the detection of interaural incoherence. II. Brief duration noises

Matthew J. Goupell[a] and William M. Hartmann
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 12 December 2005; revised 20 December 2006; accepted 5 January 2007)

Listeners detected a small amount of interaural incoherence in reproducible noises with narrow bandwidths and a center frequency of 500 Hz. The durations of the noise stimuli were 100, 50, or 25 ms, and every one of the noises had the same value of interaural coherence, namely 0.992. When the nominal noise bandwidth was 14 Hz, the ability to detect incoherence was found to depend strongly on the size of the fluctuations in interaural phase and level for durations of 100 and 50 ms. For the duration of 25 ms, performance did not appear to depend entirely on fluctuations. Instead, listeners sometimes recognized incoherence on the basis of laterality. However, when the nominal bandwidth was doubled, leading to a greater number of fluctuations, detection performance at 25 ms resembled that at 50 ms for the smaller bandwidth. It is concluded that the detection of a small amount of interaural incoherence is mediated by fluctuations in phase and level for brief stimulus durations, so long as such fluctuations exist physically. This conclusion presents a promising alternative to models of binaural detection that are based on the short-term cross-correlation in the stimulus. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2436714]

PACS number(s): 43.66.Ba, 43.66.Pn, 43.66.Qp [AK] Pages: 2127–2136

I. INTRODUCTION

Interaural coherence is a measure of the similarity of signals in the left and right ears. Interaural incoherence occurs when the signals differ in some way apart from a simple delay or attenuation. Incoherence is introduced, for example, when an unrelated noise is added to the signal appearing at one of the ears. In a previous article (Goupell and Hartmann, 2006), to be called "article I," it was concluded that the ability of listeners to detect interaural incoherence in narrowband noises is mediated by fluctuations in the interaural phase and level differences. Those conclusions were based on experiments using hundreds of different reproducible noises, all with the same value of interaural coherence, namely 0.9922. The interaural coherence was defined, as usual, as the maximum value of the cross-correlation function, as computed for the stimulus signals themselves. In these experiments, with no reference interaural delay, the amount of admixed, unrelated noise was so small that the maximum value always occurred for zero lag. Although all the noises had the same coherence, different noises had interaural phase fluctuations and interaural level fluctuations of different size. The experiments showed that these fluctuations were critical in detection. Specifically, incoherence was far more easily detected for noises with large interaural fluctuations than for noises with smaller fluctuations.

The thrust of the conclusions in terms of auditory theory was to emphasize the central importance of fluctuations and to deemphasize the stimulus interaural coherence or cross-correlation function *per se*. It was suggested that interaural coherence is an average signal property that can indicate a range of expected interaural fluctuations, but that it has no further perceptual significance. Said another way, one can expect that when the interaural coherence is close to one, then decreasing the value of coherence normally makes the incoherence easier to detect, but only because this decrease broadens the distributions of the fluctuations in the interaural differences. When the coherence decreases, the standard deviations across different noise tokens differ more from the mean standard deviation. Article I also found that as the bandwidth of the noise increases, the distributions of fluctuations across different noise tokens become more narrow. Therefore, the value of the coherence becomes a better predictor of detection performance because it better predicts the variance over time of the interaural differences.

The experiments of article I used noise stimuli with a duration of 500 ms, and the common value of coherence (0.9922) was computed over the entire 500-ms duration. That relatively long duration opens the experiments and conclusions to a serious challenge because binaural processing can be rapid, with time constants much shorter than 500 ms (Hall *et al.*, 1998). It would seem entirely possible to argue that incoherence detection is fundamentally based on the cross-correlation function after all, but that the relevant cross-correlation is computed over brief time intervals. The objection would continue by suggesting that those particular noise tokens with incoherence that was easy to detect probably had important minima in the short-term coherence, but the experiments of article I would not have been sensitive to that fact because of the excessively long temporal average used in generating and selecting the stimuli.

The purpose of the present article is to address the objection by exploring the effects of duration on incoherence detection. Specifically, the reported experiments will try to determine whether a fixed value of coherence determined over small durations—100 ms and shorter—can predict incoherence detection. In this way, the experiments presented

[a]Electronic mail: goupell@kfs.oeaw.ac.at

here address the above-mentioned challenge, and the results will have implications for the validity of the conclusions reached in article I.

II. EXPERIMENT 1

Experiment 1 begins with the null hypothesis that the detection of incoherence is determined by the short-term cross-correlation as measured over a brief stimulus and not by the interaural fluctuations. The hypothesis predicts that if stimuli are as brief as the binaural computation time for short-term cross-correlation, then detection scores will be predictable from the stimulus cross-correlation peak (coherence) itself. In particular, if every noise in an ensemble has the same value of interaural coherence, as computed over the brief duration, then the incoherence ought to be equally detectable for all noises of the ensemble.

A. Stimuli

1. Stimulus generation

Three collections of 100 two-channel, narrowband noises (i.e., left-right noise-pairs, to be called "noises") were created for Experiment 1. The different collections were distinguished by the noise durations—100, 50, and 25 ms. The generation of the noises was a multistep process, which is described in the following paragraphs.

Initially, noises were constructed from equal-amplitude random-phase components that spanned a frequency range of 490–510 Hz with a frequency spacing of 2 Hz. Components between 495 and 505 Hz had equal amplitudes of unity. Components below 495 Hz and above 505 Hz were attenuated with a raised-cosine window, which zeroed the amplitudes at 490 and 510 Hz. Consequently, there were nine spectral components that had nonzero amplitudes, and the 3-dB bandwidth was 14 Hz. An orthogonalization procedure guaranteed that the interaural coherence of each noise was precisely 0.9922. Up to this point the stimulus generation procedure was identical to the procedure followed in article I.

Next, the stimuli were given temporal windows with total durations of 100, 50, or 25 ms including 10-ms Hanning edges for attack and decay. After the temporal shaping, the value of the coherence was recomputed. Noises were accepted only if the value of the coherence was 0.992 ± 0.001. In order to obtain 100 noises with a given duration and correct coherence, it was necessary to reject more than ten times that number.

After a collection of 100 noises was created, ten noises were selected to make a *phase set* and ten were selected to make a *level set*. As in article I, the phase set consisted of those five noises with the largest standard deviations in interaural phase, as computed over the duration, plus those five noises with the smallest standard deviation in interaural phase. Similarly, the level set was constructed by selecting noises with the five largest and five smallest standard deviations in interaural level. Occasionally a particular noise would be common to both phase and level sets. These two sets of stimuli were presented to listeners.

2. Stimulus spectra

The goal of the stimulus generation technique was to pack a number of components into a narrow band to create a noise with complicated fluctuations but a brief duration. However, the brief duration resulted in a bandwidth greater than the nominal value of 14 Hz. The spectra of the noises selected for the phase and level sets were measured. Average statistics were as follows: For the 100-ms noises, 90% of the energy was contained in a band 24 (±11) Hz wide, and 99% was contained in a band 76 (±22) Hz wide. For the 50-ms noises, the 90% and 99% bandwidths were 39 (±8) and 106 (±16) Hz. For the 25-ms noises, the corresponding bandwidths were 74 (±4) and 126 (±16) Hz. No correlation could be seen between the bandwidths of individual noises and the sizes of the interaural fluctuations in phase or level.

3. Stimulus interaural values

Figure 1 shows the fluctuations for all the stimuli used in this experiment. Fluctuations in interaural phase and interaural level for each noise are expressed, as in article I, as the standard deviations $s_t[\Delta\Phi]$ and $s_t[\Delta L]$ computed over the stimulus duration. For example, for a given noise, $s_t[\Delta\Phi]$ is computed by taking the square root of the sum of the squared deviations of the instantaneous interaural phase from the mean value. (Although an ensemble-average mean value would be zero, mean values for individual noises can be important for brief durations.) The average value of $s_t[\Delta\Phi]$, namely $\mu(s_t[\Delta\Phi])$, and the standard deviation of $s_t[\Delta\Phi]$, namely $\sigma(s_t[\Delta\Phi])$, are computed over all the noises of the ensemble, i.e., over the collection of 100 noises. This ensemble standard deviation appears in the legends in the figure panels on the left, together with the average value and the standard deviation for $s_t[\Delta L]$ as well as the correlation between phase and level fluctuations. It is interesting to observe that the average values decrease as the duration decreases for this fixed nominal bandwidth. By contrast, it was found in article I, that the average values were roughly constant for a long fixed duration when the bandwidth varied. It is worth noting that the 10-ms edges did not affect the computation of the fluctuations. In our approximations, the calculated interaural phase and interaural level depend on ratios in which the multiplicative Hanning edge is canceled.[1] The left-hand panels of Fig. 1 make it clear that noises with a large fluctuation in interaural phase also tend to have a large fluctuation in interaural level, and vice versa. The correlation between phase and level fluctuations ranges from 0.53 to 0.75.

The right-hand panels of Fig. 1 show the mean and standard deviations of the five stimuli with the greatest interaural fluctuations and of the five stimuli with the least. Circles are for the phase sets; squares are for the level sets. These stimuli were selected from the collections shown in the left-hand panels. These are the ten stimuli that were heard by the listeners. The difference between the larger fluctuations and the smaller fluctuations diminishes dramatically as the duration decreases to 25 ms.

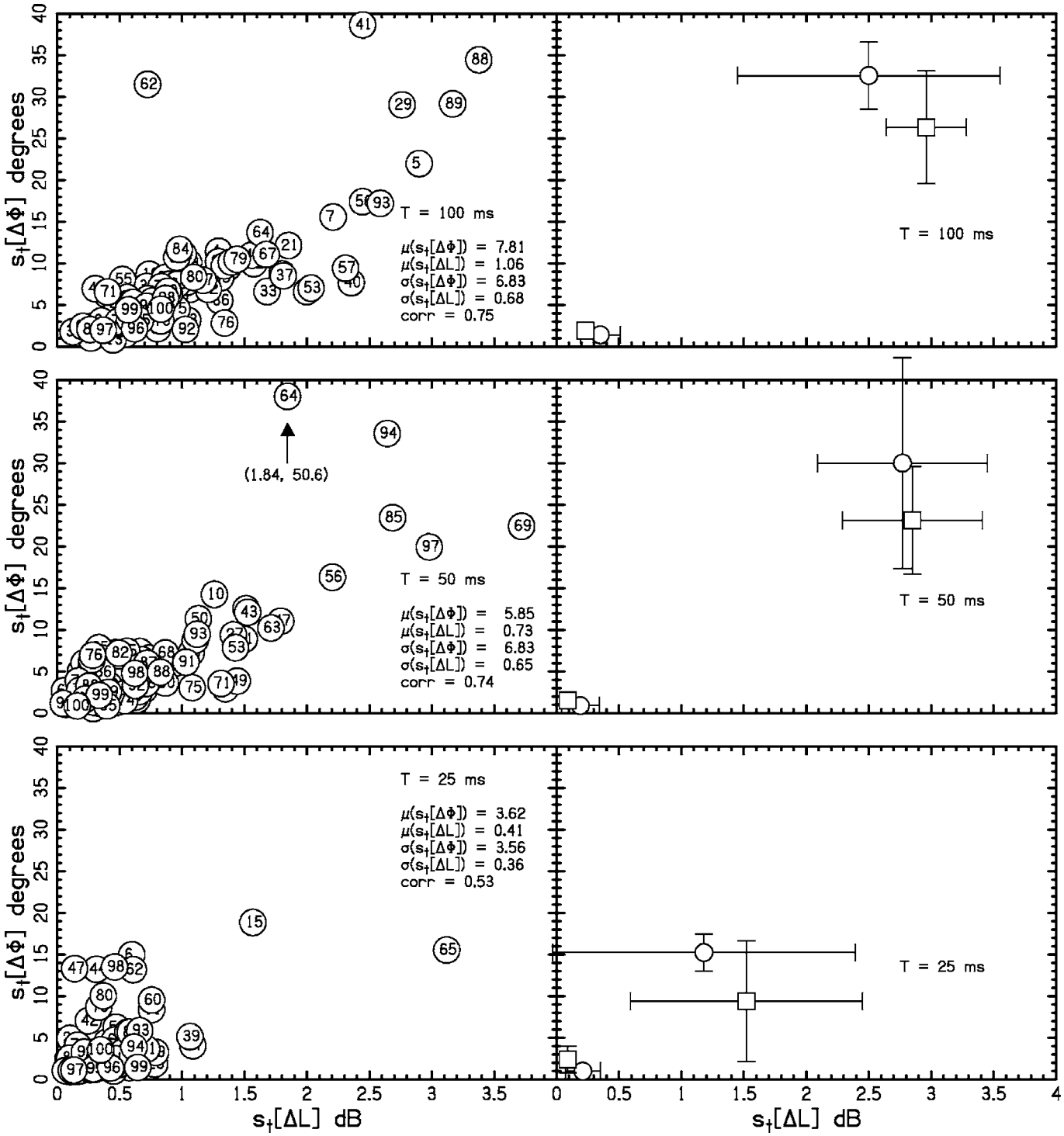


FIG. 1. Panels in the left column show the interaural fluctuations for the three collections of 100 noises. Fluctuations in interaural phase and interaural level for each noise are expressed as the standard deviations, $s_t[\Delta\Phi]$ and $s_t[\Delta L]$, computed over the stimulus duration. Panels in the right column show the mean and standard deviations of the five stimuli with the greatest interaural fluctuations and of the five stimuli with the least. (Circles are for phase sets; squares are for level sets.) These ten stimuli were presented to the listeners. Error bars are 2 s.d. in overall width.

4. Stimulus synthesis

Noises were computed by a Tucker-Davis AP2 array processor (System II) and converted to analog form by 16-bit DACs (DD1). The buffer size was 4000 samples per channel and the sample rate was 8 ksps. The noise was low-pass filtered with a corner frequency of 4000 Hz and a −115 dB/octave rolloff. The noises were presented at 70±3 dB with levels determined by programmable attenuators (PA4) operating in parallel on the two channels. The level was randomly chosen, in 1-dB increments, for each of the three intervals within a trial to discourage the listener from trying to use level cues to perform the task.

B. Procedure

Listeners were seated in a double-wall sound attenuating room and used Sennheiser HD414 headphones. Each experi-

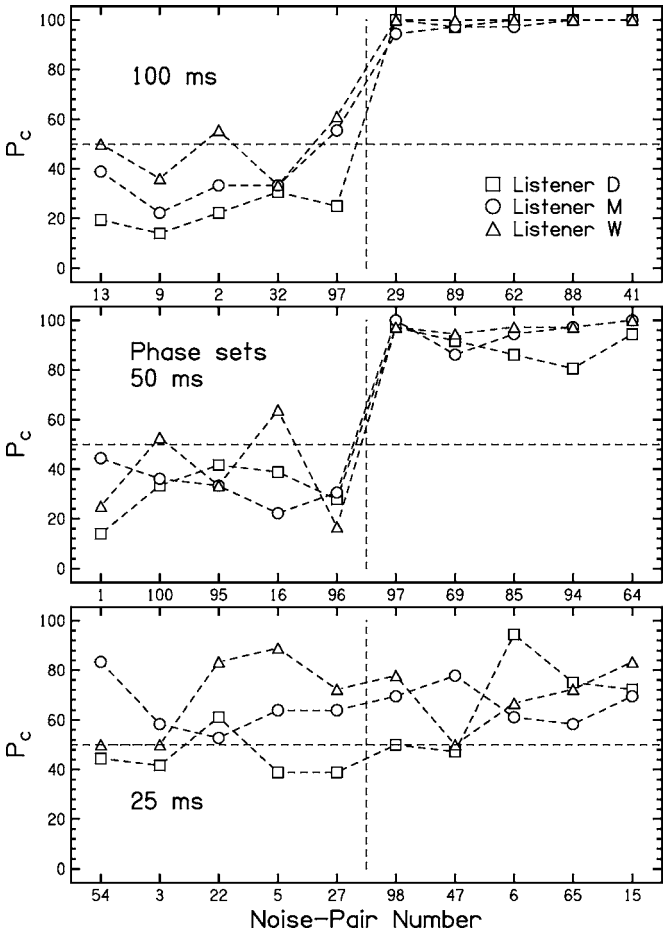


FIG. 2. The percent correct (P_c) for three listeners for the *phase set* from the 25-, 50-, and 100-ms collections. The noises were chosen to have the smallest and largest $s_t[\Delta\Phi]$ in the collection of 100 noises. The noises are rank ordered by increasing $s_t[\Delta\Phi]$ along the horizontal axis. The vertical dashed line represents 90 unused noises. The horizontal dashed line represents the level of guessing. Several of the 50- and 100-ms noises to the left of the dashed line are below the level of guessing.

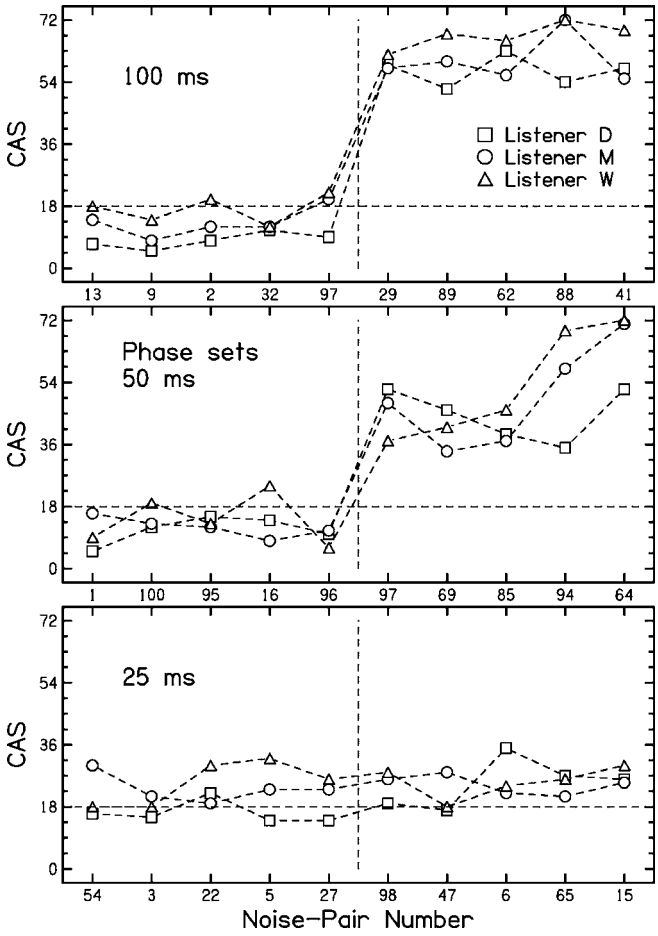


FIG. 3. Same as Fig. 2 except that the CAS values are plotted instead of the percentage of correct responses.

mental run was devoted to either a phase set or a level set. Within a set, the order of the reproducible noises was randomized—differently on each run. Six runs were devoted to the phase set and six to the level set. Listeners completed one set before moving on to the other.

The structure of runs, trials within a run, and the data collection procedure was the same as that in article I. It is briefly described as follows: A noise 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 noises in a set was presented incoherently a total of six times. Thus, a listener heard an individual noise 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 noises in the set except that they were required to be different from either channel of the incoherent "odd" interval and to be different from one another. The interinterval duration was 150 ms. The listener was required to decide which of the two latter intervals was the incoherent interval.

As described in article I, listeners were allowed to give a confidence rating. By making a correct choice and indicating confidence, a listener gained two points for the trial instead of just one point for a simple correct response. Making a wrong choice and indicating confidence was discouraged by allowing the listener only one such error in a run. With the second "confident" error, the run terminated and the listener was required to begin it again. This procedure led to a confidence-adjusted score (CAS) with a maximum possible score of 72 for a run of 36 trials. The data collection procedure kept track of both the percentage of correct responses (P_c) and the CAS.

C. Listeners

Experiments in this article employed three male listeners from article I—D, M, and W. Listeners D and M 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 those used in the experiment. Listeners M and W were the authors.

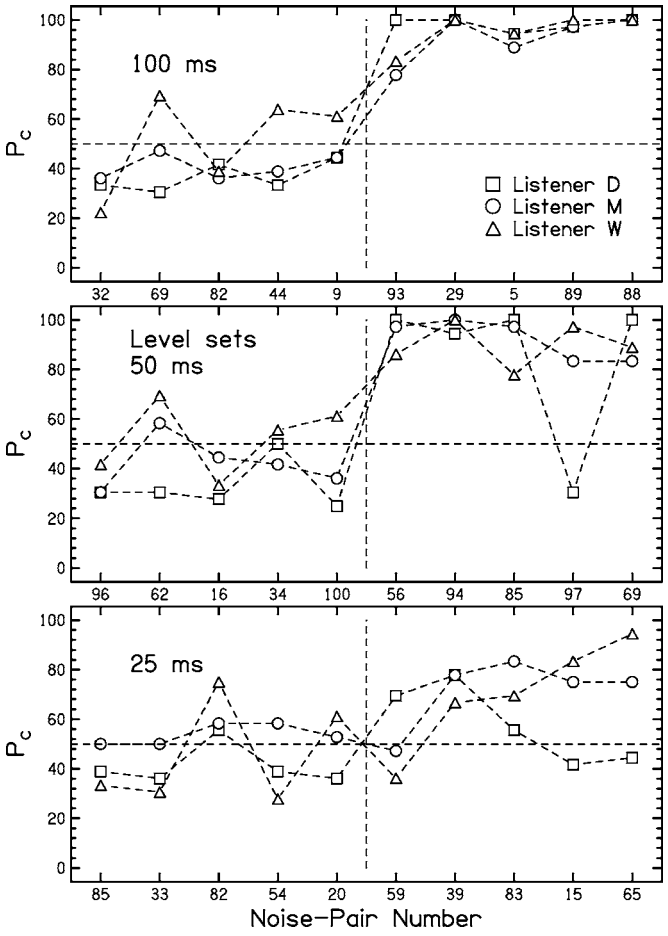


FIG. 4. The percent correct (P_c) for three listeners for the *level set* from the 25-, 50-, and 100-ms collections. The noises were chosen to have the smallest and largest $s_l[\Delta L]$ in the collection of 100 noises. The noises are rank ordered by increasing $s_l[\Delta L]$ along the horizontal axis. The vertical dashed line represents 90 unused noises. The horizontal dashed line represents the level of guessing. Several of the 50- and 100-ms noises to the left of the dashed line are below the level of guessing.

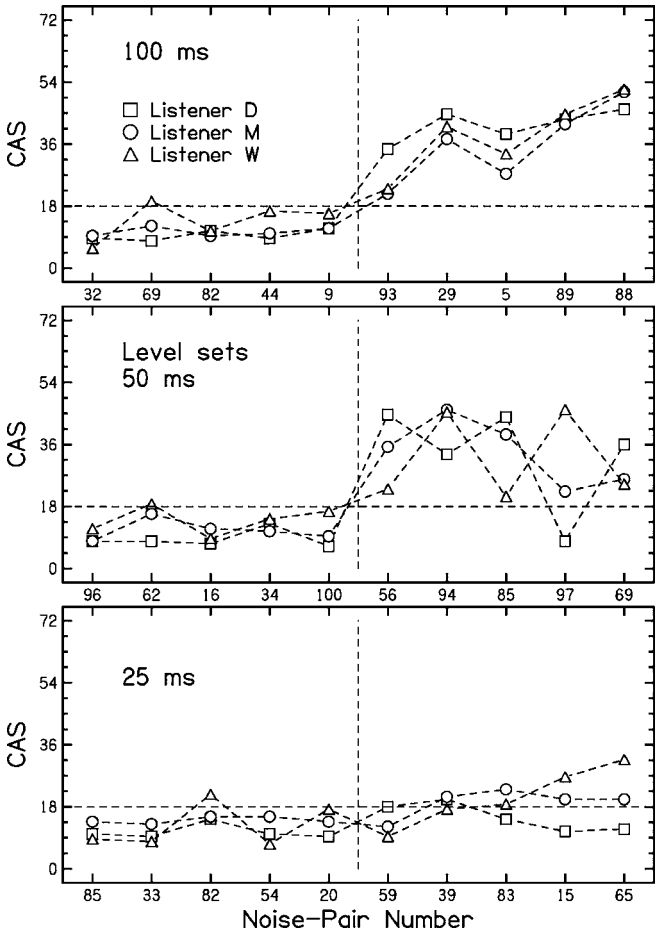


FIG. 5. Same as Fig. 4 except that the CAS values are plotted instead of the percentage of correct responses.

D. Results

The average value of P_c over phase and level sets over all three listeners was: 68% for a duration of 100 ms, 65% for 50 ms, and 61% for 25 ms. These values of P_c can be compared to the value from article I for a duration of 500 ms as obtained for the same three listeners. That value was P_c =92%, notably larger.

Figures 2–5 show the P_c and CAS values for the phase and level sets for the 100-, 50-, and 25-ms noises. The vertical dashed line shows the division between five noises with the smallest fluctuations and the five noises with the largest fluctuations. Thus, the vertical line represents 90 unused noises. The horizontal dashed line shows the value corresponding to guessing.

In Figs. 2–5, the 100- and 50-ms conditions show higher values of P_c and CAS for the five noises to the right of the vertical dashed line compared to the five noises to the left. These noises with the largest fluctuations have values of P_c and CAS that are near the ceiling (P_c=100% or CAS=72). This is different from the noises with a 25-ms duration, which had few noises with P_c values greater than 75% and no noises with CAS values greater than 36.

As in article I, t-tests were performed to determine whether detection of incoherence more frequently occurs in the five noises with the largest fluctuations in phase or level when compared to the five noises with the smallest fluctuations in phase or level. The p values from these t-tests are in Table I. For the 100- and 50-ms sets all 12 p values were significant at the 0.01 level for P_c, as were 11 of 12 for CAS. The significance of these tests is a strong indication that the hypothesis that began this experimental section does not apply for durations of 100 and 50 ms: For these durations, the coherence statistic is inadequate to predict detection performance. Instead, detection performance depends on interaural fluctuations, contrary to the hypothesis that detection is mediated by the interaural coherence of the short stimulus as a whole.

For the 25-ms sets, three of six p values were significant at the 0.05 level for P_c and also for CAS. Clearly, the number of significant p values and the levels of significance for the 25-ms sets did not match those for 100 and 50 ms.

E. Discussion: 50 and 100 ms

The results for noise durations of 50 and 100 ms differ so greatly from those at 25 ms that they are separately discussed.

TABLE I. The p values for the phase and level sets with a nominal bandwidth of 14 Hz and four durations: 25, 50, 100, and 500 ms. The 500-ms p values are taken from article I. Small p values indicate that incoherence is significantly more detectable for the five noises with large fluctuations than for the five noises with small fluctuations.

P_c	25 ms		50 ms		100 ms		500 ms	
Listener	Phase	Level	Phase	Level	Phase	Level	Phase	Level
D	0.029	0.039	<0.001	0.008	<0.001	<0.001	0.095	0.006
M	0.334	0.027	<0.001	<0.001	<0.001	<0.001	0.021	0.047
W	0.457	0.056	0.001	<0.001	<0.001	0.003	0.015	0.014
CAS	25 ms		50 ms		100 ms		500 ms	
Listener	Phase	Level	Phase	Level	Phase	Level	Phase	Level
D	0.027	0.039	<0.001	0.010	<0.001	<0.001	<0.001	<0.001
M	0.306	0.023	0.002	0.002	<0.001	0.004	0.002	0.002
W	0.457	0.065	0.002	0.015	<0.001	0.002	0.001	<0.001

1. Comparison with long-duration stimuli

The tests of significance for 100- and 50-ms durations can be compared with those from article I where noise durations were 500 ms. The CAS data obtained here for 100-ms noises show the same number of significant p values and levels of significance as the 500-ms CAS data from article I. However, compared to the 500-ms P_c data, the 100-ms P_c data show a larger number of significant p values. The reason for the difference in statistics is that when the duration was 500 ms, the P_c ran into a ceiling, which limited its statistical usefulness. Then the CAS measure became essential. However, as noted earlier, average P_c values were normally smaller for durations of 100 ms and shorter as studied here. Therefore, the P_c became statistically useful, and the CAS measure became less necessary. Nevertheless, the CAS does reveal more subtle features. For instance, for 100-ms duration, Figs. 2 and 4 show that incoherence could be detected with approximately equal success in both phase sets and level sets, but Figs. 3 and 5 show that listeners were more confident about their answers in the phase sets.

2. Responses worse than chance

Figures 2 and 4 for the percentage of correct responses show that many of the 100- and 50-ms noises with small fluctuations led to responses below the random guessing limit. For each duration there are 30 data points for small-fluctuation noises (3 listeners $\times$ 2 sets $\times$ 5 small-fluctuation noises). Both for the 50-ms noises and for the 100-ms noises, 23 of these led to P_c less than 50%. By contrast, in the comparable experiments of article I, with 500-ms noises and 14-Hz bandwidth, only one value of percent correct was less than guessing out of 40 points (4 listeners $\times$ 2 sets $\times$ 5 small-fluctuation noises).

The reason for this dramatic drop in P_c for the short durations is probably the single-channel envelope fluctuations, as were studied in Experiment 5 of article I. Physically, it is found that noises that have large interaural fluctuations also tend to have large envelope fluctuations. The envelope fluctuations, of course, are prominent even when the noise is heard diotically. In an experiment where it is very hard to hear the interaural fluctuations, as it was for the five selected small-fluctuation noises, a listener may choose the interval with large envelope fluctuation as the interval with the most "action." Apparently, listeners are more apt to confuse envelope fluctuations with interaural fluctuations for the durations of 100 and 50 ms than for a duration of 500 ms. Visual inspection of the envelopes for the small fluctuation noises reveals envelopes that are particularly flat.

Quantitatively, the performance worse than chance can be understood from the following logic: In the limit that noises with large interaural fluctuations always have large single-channel envelope fluctuations, and noises with large single-channel fluctuations are always chosen over noises with small interaural fluctuations in our task, values of P_c should be approximately 25% for the small-interaural-fluctuation noises. The reason is that, on average, dichotic small-interaural-fluctuation noises will be presented against diotic large-fluctuation noises half of the time. This, in turn, would reduce the P_c values of these noises from guessing level, which is 50%, to half of guessing level, which is 25%. As can be seen in Figs. 2 and 4, many small interaural fluctuation noises with a duration of 100 or 50 ms lead to P_c scores near 25%. Similarly, if listeners are misled, as hypothesized, by envelope fluctuations but are never confident about their decisions, this limit predicts a CAS of 9, half the guessing limit. Figures 3 and 5 show a number of scores as low as 9 for durations of 100 and 50 ms.

F. Discussion: 25 ms

The results for noises of 25-ms duration were unlike the results for any other duration because the fluctuations were few in number. Interaural fluctuations of a 14-Hz bandwidth noise are expected to vary with a time scale of $1/14 = 71$ ms. Therefore, it is expected that the interaural phase difference (IPD) and interaural level difference (ILD) would not vary much over a stimulus duration of 25 ms. The 25-ms noises used experimentally had bandwidths larger than this—about 74 Hz ($1/74 = 14$ ms). Nevertheless, fluctuations in the 25-ms noises were rare. Fluctuations can be roughly quantified by examining the number of changes in sign of the interaural phase and interaural level during the course of the stimulus. Out of 20 noises used in the phase and level sets, 3

were common to both sets. For the 17 different noises, 10 had no change in sign for either interaural level or phase, 4 had one change in sign, 2 had two, and 1 had three. This can be compared to the longer duration noises. For the 13 different 50-ms noises, 5 had no change in sign, 3 had one change in sign, 2 had two, and 3 had three or more. For the 15 different 100-ms noises, 3 had no change in sign, 2 had one change in sign, 4 had two, and 6 had three or more.

Noises that were 25 ms in duration led to only 11 values of P_c and CAS below chance level. It will be argued below that these stimuli, with this duration-bandwidth product, are likely to be too short to elicit perceptible fluctuations, monaural or binaural.

1. The laterality cue

All three listeners reported that when the duration was reduced to 25 ms, they sometimes used a laterality cue because there seemed to be little or no perceived width for most of the noises. This change in detection strategy is a likely explanation for the near-chance values of P_c and CAS and for the negative results of the t-tests for the 25-ms sets. To generate the stimulus sets, noises were selected by the s_t statistic, which is a fluctuation statistic associated with a width cue. As a standard deviation, this statistic is unaffected by a constant shift in mean value, which is associated with lateralization. When listeners used a laterality cue on sets that had been sorted by a statistic associated with a width cue, p values did not show as many significant differences between the noises.

2. Lateralization experiment

An auxiliary experiment was performed to test the idea that laterality is a salient cue for the 25-ms noises, as suggested earlier. Each of the ten noises from the 25-ms phase set and each of the ten from the 25-ms level set was alternated repeatedly with a diotic noise. The listener's task was to say whether the noise was to the left or right of the diotic standard. Results were compared with predicted values based on a sum of the compressed IPD and ILD lateralization according to a formula fitted to the data of Yost (1981). Out of 20 noises, the responses agreed with prediction on 17 for Listener M, on all 20 for Listener W, but only on 12 for Listener D. The agreement is an indication that laterality is salient, at least for some listeners. It also indicates that lateralization as computed from Yost's sine-tone data can be applied to brief noise bursts. This observation was previously made in connection with the modeling of incoherence detection (Goupell, 2005).

3. Modeling for 25 ms

Models for the detection of incoherence begin with the physical stimulus, and, to a greater or lesser extent, take into account known facts about the human auditory system. Such models are applied to the detection of interaural fluctuations in a future article in this series, where the predictions are compared with the results from experiments using long-duration noises. The present section is dedicated to simple, stimulus-based models for the shortest-duration noises, 25 ms. The models were tested only on the P_c data because confident responses were infrequent for this duration. The results of the tests are shown in Table II.

TABLE II. Linear correlation coefficients, r, for 25-ms sets with 14-Hz nominal bandwidth. The data were averaged over listeners. Several different laterality (mean, μ) and fluctuation (standard deviation, s_t) statistics were considered. The use of laterality compression allows IPD and ILD measures to be combined.

Statistic	No Lat. Comp.	Lat. Comp.
$\mu(\Delta\Phi)$	0.25	0.31
$\mu(\Delta L)$	0.35	0.32
$\max(\lvert\mu(\Delta\Phi)\rvert, \lvert\mu(\Delta L)\rvert)$	⋯	0.51
$\mu(\Delta\Phi)+\mu(\Delta L)$	⋯	0.48
$\lvert\mu(\Delta\Phi)\rvert+\lvert\mu(\Delta L)\rvert$	⋯	0.54
$s_t(\Delta\Phi)$	0.58	0.59
$s_t(\Delta L)$	0.54	0.61
$s_t(\Delta\Phi)+s_t(\Delta L)$	⋯	0.63
If $s_t(\Delta\Phi)+s_t(\Delta L)<1$ and $P_c>56\%$ then $\lvert\mu(\Delta\Phi)\rvert+\lvert\mu(\Delta L)\rvert$ else $s_t(\Delta\Phi)+s_t(\Delta L)$	⋯	0.72
Running short-term cross-correlation	0.38	⋯

The first two models assume that incoherence is detected on the basis of displacements from the midline, the average value of the IPD alone, $\mu(\Delta\Phi)$, or the average value of ILD alone, $\mu(\Delta L)$. Table II shows that the correlation with the experimental values of P_c is only about 0.3. It is possible to combine IPD and ILD displacements from the midline if both are put on the same scale of laterality using the compression function that fits Yost's 1981 data. Then choosing the maximum of the two displacements, or the sum of the displacements, or the sum of the absolute values of the displacements all lead to a correlation r of about 0.5.

A second set of models ignores average displacements from the midline and assumes that incoherence detection is based on fluctuations alone: phase fluctuations, $s_t(\Delta\Phi)$, level fluctuations, $s_t(\Delta L)$, or a combination. As shown in Table II, these models correlate with P_c data with r about 0.6.

Because both laterality models and fluctuation models were positively correlated with the data, there is an indication that a combination of the two models would make predictions that agree even better with the data. It was found that combining fluctuations and laterality conditional upon the P_c score could achieve a value of $r=0.72$. This conditional model says that if the fluctuations are small and yet performance is somewhat better than chance ($P_c>56\%$) then the model uses the sum of phase and level laterality magnitudes as the decision variable. Otherwise, it uses the sum of the phase and level fluctuations. We suspect that the reason that the correlation between model and data is not higher is that the treatment of the fluctuations omits important facts. These facts are included in the models of a future article.

A final calculation for the 25-ms data assumed that in-

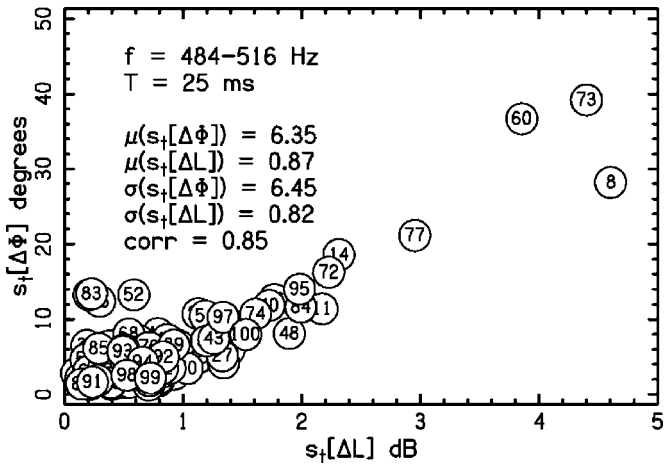


FIG. 6. Standard deviations in interaural phase and level showing the fluctuations for the 100 noises of Experiment 2, with nominal bandwidth of 28 Hz and duration of 25 ms. The scale is changed compared to Fig. 1. The ensemble-average fluctuations (μ) resemble those in Fig. 1 for 14-Hz nominal bandwidth and duration of 50 ms.

coherence is detected on the basis of the running short-term cross-correlation function. The cross-correlation was computed over a rectangular window, as long as 10 ms, for every instant in time. Then the average of this cross-correlation over the duration of the stimulus was used as a statistic to compare with P_c data. This model correlated with the data with $r=0.38$—better than the laterality models but not as successful as the fluctuations models. When the window was increased to 20 ms, the r value remained about the same.

III. EXPERIMENT 2

Experiment 1 found that the detection of incoherence in noises with a nominal bandwidth of 14 Hz depends on interaural fluctuations for durations of 50 ms and longer. Little fluctuation dependence was observed for 25 ms, but it seemed likely that the reason was that interaural fluctuations were too infrequent and too small for this duration, with mean values less than half of those at 100 ms, as shown in the legends of Fig. 1. An alternative possibility is that a duration of 25 ms is simply too short to perceive fluctuations. In order to decide between these two ideas, Experiment 2 again used 25-ms noises but doubled the nominal bandwidth to 28 Hz. The larger bandwidth was expected to

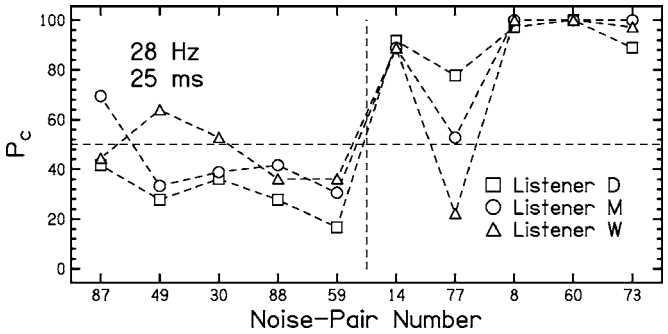


FIG. 7. Percent correct responses as in Figs. 2 and 4 except that the data are from Experiment 2, where the duration is 25 ms and the nominal bandwidth is 28 Hz. Phase sets and level sets are essentially the same for this experiment.

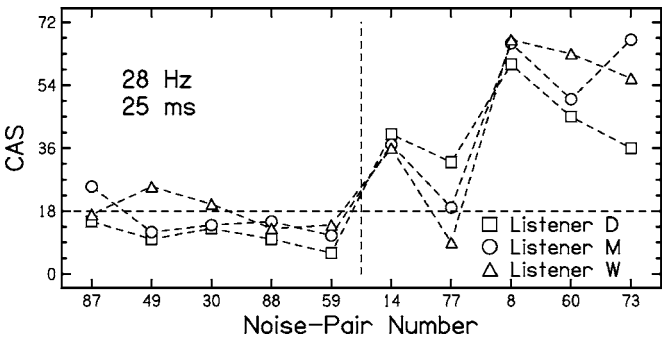


FIG. 8. CAS as in Figs. 3 and 5 except that the data are from Experiment 2, where the duration is 25 ms and the nominal bandwidth is 28 Hz. Phase sets and level sets are essentially the same for this experiment.

lead to faster interaural fluctuations, which would then lead to more fluctuations within the brief duration of the stimulus. If interaural incoherence detection scales perfectly, one would expect that 25-ms noises with a nominal bandwidth of 28 Hz would be comparable to 50-ms noises with a nominal bandwidth of 14 Hz, as tested in Experiment 1.

A. Method

The noises for Experiment 2 were generated in the same way as for Experiment 1 except for the bandwidth. Noises having a nominal 28-Hz bandwidth per the 3-dB down points were made by using 17 components with finite amplitude from 484 through 516 Hz. The amplitudes of the two components with the lowest frequencies and the two with the highest were shaped as for Experiment 1.

When it came to selecting noises for phase sets and level sets with the five largest and smallest values of interaural fluctuation, it was discovered that the five noises with the largest phase fluctuations also had the largest level fluctuations. Consequently, we thought it adequate to do the experiment with a single set of ten stimuli, presenting the five largest and smallest interaural phase fluctuations from the 100 noises shown in Fig. 6. For the 10 independent noises, 3 had no change in sign for either interaural level or phase, 4 had one change in sign, 1 had two, and 2 had three. Thus, the average number of sign changes per noise was 1.2, compared to 0.64 for the narrower band used in Experiment 1. The spectral distributions were computed for the ten noises of Experiment 2. On average, 90% of the energy was in a band 80 (±10) Hz wide, and 99% was in a band 130 (±15) Hz wide.

B. Results

The percentage of correct responses is shown in Fig. 7, which can be compared with Fig. 2. The CAS values are shown in Fig. 8, and these can be compared with Fig. 3. Both comparisons show that the performance for 28-Hz nominal bandwidth and 25-ms duration most closely resembles the performance for 14-Hz nominal bandwidth and 50-ms duration. One-tailed t-tests of the hypothesis that the percentage of correct responses is higher for the five noises on the right were significant at levels 0.001, 0.002, and 0.040, for Listeners D, M, and W, respectively. Similar tests for the CAS

values were significant at levels 0.001, 0.011, and 0.028, for Listeners D, M, and W, respectively. These values can be compared with those for either phase or level sets in Table I.

C. Discussion

As noted in the legends of Fig. 1, the fluctuations for 25-ms, 14-Hz noise are about half of those for 50-ms, 14-Hz noise. However, as shown in the legend of Fig. 6, when the nominal bandwidth is doubled in Experiment 2 the fluctuations for 25-ms, 28-Hz noise are about the same as those for 50-ms, 14-Hz noise. *A priori* one might expect the detection of incoherence for these two noises to be about the same as well. However, although the difference in detectability between large fluctuation noises and small fluctuation noises is clearly significant for 25-ms, 28-Hz noise, it is not quite as significant as for 50-ms, 14-Hz noise. Evidently, the time-bandwidth equality does not lead to identical performance. The reason may be that, by our energetic measures, the physical bandwidths of 25-ms noises with a 28-Hz nominal bandwidth are not exactly twice those for 14-Hz nominal bandwidth.

The most important conclusion from Experiment 2 is that listeners have no difficulty perceiving interaural fluctuations in noises that are 25 ms in duration if the fluctuations are physically present. Listeners can use these fluctuations to recognize interaural incoherence.

IV. CONCLUSION

This article began with a challenge to a previous work, article I (Goupell and Hartmann, 2006). That article had concluded that for 500-ms noises, with bandwidths equal to a critical band or narrower, the detection of interaural incoherence depends on interaural fluctuations in phase and level. Article I specifically denied that the detection could be predicted from the cross-correlation of the entire stimulus. That conclusion was based on experiments using noises, which all had the same interaural coherence. It was found that those noises with larger fluctuations were more readily identified as incoherent. The challenge noted that the stimulus duration of 500 ms is far longer than binaural integration times. The challenge continued by suggesting that coherence measured over 500 ms was possibly perceptually meaningless because the binaural system would likely base its decisions on coherence measured over much smaller time spans.

The present article addressed the challenge by experiments that were the same as those in article I except that the stimuli had durations of 100, 50, and 25 ms—intended to match probable binaural integration times. The coherence measured over those short durations for each of the stimuli was again 0.992. The bandwidth was again nominally 14 Hz, as in article I.

The experiments found that for durations of 100 and 50 ms, incoherence in those noises with large fluctuations was much more readily detected than incoherence in those noises with small fluctuations, even though coherence values were the same. Statistically, the results were even stronger than were found in article I. Consequently the conclusions of article I were upheld for these durations. Coherence, i.e. the peak of the cross-correlation function, is inadequate to predict incoherence detection.

For a duration of 25 ms, it was discovered that there were sometimes too few physical fluctuations to be perceptually useful when the bandwidth was 14 Hz. The small fluctuations for 25-ms, 14-Hz noises make this stimulus unique among our stimuli. The perceived difference between these noises and diotic signals is that these noises are sometimes lateralized. By contrast, all the other stimuli explored in this article lead to salient *fluctuations* in interaural differences. These subjective differences are our explanation for two facts about the 25-ms, 14-Hz noises: (1) Larger fluctuation noises are not better identified as incoherent than are smaller fluctuation noises. (2) Performance, as measured by percent correct or CAS values, is always near chance.

If stimuli are selected on the basis of fluctuations but the most salient cue is laterality, then detection performance differences are unlikely to reflect the selection criterion. The smaller physical range of fluctuations for 25 ms is likely to translate into a smaller perceptual difference between large and small fluctuations. In connection with the second fact, it appears that the lateralization cues that are available for brief stimuli are less salient than the fluctuation cues available for longer stimuli. Consequently, performance suffers.

We suspect that this distinction between laterality and fluctuation is also seen in forward (and backward) fringe experiments in NoSπ detection (McFadden, 1966). A homophasic (No) fringe is especially helpful when the signal is brief (Robinson and Trahiotis, 1972). As noted by Yost (1985), who used a 20-ms signal, a homophasic fringe (No) provides a reference such that an $S\pi$ signal leads to a variation in lateralization. The variation leads to improved detection. A similar observation was made by Gilkey *et al.* (1990), who referred to the sudden change in interaural parameters as an "onset-effect." By contrast, when the signal duration is long, the interaural fluctuations introduced by the out-of-phase signal produce temporal variations of their own, and a fringe becomes less valuable.

To try to understand the incoherence detection data for 25-ms, 14-Hz noises we tested some simple stimulus-based models for detection. Some of the models were based on laterality, defined as an average lateralization, others were based on fluctuation. The most successful was a model that depended on both, depending on circumstances, consistent with the idea that for this (most awkward) condition some noises have more prominent laterality while others have more prominent fluctuations. However, none of the models were highly successful.

Our conclusion that the anomalous results of the experiment with 25-ms, 14-Hz bandwidth noise were due to a lack of fluctuations was validated by a second experiment. This experiment too used 25-ms noises, but with a doubled bandwidth, effectively doubling the number of fluctuations presented to the listener. When the bandwidth was doubled, significant differences in fluctuations appeared amongst different stimuli, and significant differences in detectability of incoherence reappeared in the P_c data.

Assuming, as we do, that the short-term analysis capabilities of the binaural system can be correctly explored by using stimuli with short durations, it follows that there is no sense in which the short-term cross-correlation of the stimulus is, by itself, adequate to predict detection. Instead, the cross-correlation statistic provides a guide to the distribution of interaural fluctuations that can be expected in an ensemble of noises. It is inadequate to predict the fluctuations for any particular noise, and it is these fluctuations that are at the basis of incoherence detection.

The significance of these results for binaural modeling is to clarify the role of the binaural cross-correlator. A model that ascribes incoherence detection to a reduction of stimulus cross-correlation as measured at zero lag (midline) is inconsistent with the results of our research. A model that examines the output of the cross-correlator off the midline, where momentary fluctuations occur, is consistent. In addition, the fluctuations in interaural phase measured by the cross-correlator must be supplemented with the fluctuations in interaural level in order to make a successful detector of incoherence.

Although the short-duration experiments presented here seem to have dealt with the challenge to article I in a satisfactory way, this article together with article I have only established the importance of *stimulus* interaural fluctuations in comparison with the cross-correlation function. This article has not established what properties of these fluctuations, or what transformations, or what combinations of IPD and ILD fluctuations are used by the auditory system in detection. This article has also not compared the fluctuation model with the performances of a model *neural* cross-correlator. A future work will develop a signal-based auditory model that attempts to treat fluctuation detection in a realistic way and will make the necessary comparisons.

ACKNOWLEDGMENTS

This work was partially supported by the NIDCD under Grant No. DC 00181, and partially supported by a dissertation completion fellowship to M.J.G. from the College of Natural Science of Michigan State University. Dr. A. Kohlrausch, Dr. C. Trahiotis, and Dr. H. S. Colburn made useful comments on an earlier version of this article.

[1]To determine whether the values of fluctuations in the legends of Fig. 1 were representative of a large population, 30 000 noises were computed for each duration, and interaural fluctuations were calculated. The comparison between the legend and the large-population values showed that means, standard deviations, and correlations for 100- and 500-ms durations differed by less than 10%. For 50- and 25- ms durations, the discrepancy was less than 15% except that the standard deviation of phase fluctuations for 50 ms was 5.63° in the large ensemble instead of 6.83, and the standard deviation of level fluctuations for 25 ms was 0.50 dB in the large ensemble instead of 0.36. Also for 25 ms, the large-ensemble correlation was 0.69 instead of 0.53.

Gilkey, R. H., Simpson, B. D., and Weisenberger, J. M. (**1990**). "Masker fringe and binaural detection," J. Acoust. Soc. Am. **88**, 1323–1332.

Goupell, M. J. (2005). "The use of interaural parameters during incoherence detection in reproducible noise," Ph.D. dissertation, Michigan State University, East Lansing, MI.

Goupell, M. J., and Hartmann, W. M. (**2006**). "Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects," J. Acoust. Soc. Am. **119**, 3971–3986.

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.

McFadden, D. (**1966**). "Masking level differences with continuous and with burst masking noise," J. Acoust. Soc. Am. **40**, 1414–1419.

Robinson, D., and Trahiotis, C. (**1972**). "Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions," Percept. Psychophys. **12**, 335–338.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.

Yost, W. A. (**1985**). "Prior stimulation and the masking level difference," J. Acoust. Soc. Am. **78**, 901–907.

Loudness changes induced by a proximal sound: Loudness enhancement, loudness recalibration, or both?

Daniel Oberfeld[a)]
Department of Psychology, Johannes Gutenberg—Universität Mainz, 55099 Mainz, Germany

(Received 21 September 2006; revised 30 January 2007; accepted 30 January 2007)

The effect of a forward masker on the loudness of a target tone in close temporal proximity was investigated. Loudness matches between a target and a comparison tone at the same frequency were obtained for a wide range of target and masker levels. Contrary to the hypothesis by Scharf, Buus, and Nieder [J. Acoust. Soc. Am. **112**, 807–810 (2002)], these matches could not be explained by an effect of the masker on the comparison loudness, which was measured by loudness matches between the comparison and a fourth tone separated in frequency from the comparison and the masker. The data thus demonstrate that a forward masker has an effect on the loudness of a proximal target. The results are compatible with the suggestion by Arieh and Marks [J. Acoust. Soc. Am. **114**, 1550–1556 (2003)] that the masker triggers two processes. The data indicate that the effect of the slower-decaying process resulting in a reduction in the loudness of a following tone saturates at masker-target level differences of 10–20 dB. The faster-decaying process causing loudness enhancement or loudness decrement has the strongest effect at a masker-target level difference of approximately 30 dB. A model explaining this mid-difference hump is proposed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2710433]

PACS number(s): 43.66.Cb, 43.66.Mk, 43.66.Ba [AJO] Pages: 2137–2148

I. INTRODUCTION

Does a sound presented in close temporal proximity to a target sound have an effect on the loudness of the target, and how can this effect be understood? Effects of a *masker*[1] temporally separated from a *target* by less than 1 s have received considerable interest in the field of auditory intensity processing, although not all phenomena are currently well understood (for recent reviews see Plack and Carlyon, 1995; Scharf, 2001; Oberfeld, 2005). Concerning effects on loudness, several studies conducted in the 1970s reported that a forward masker higher in level than a target following it by less than about 500 ms caused an increase in target loudness (*loudness enhancement*; e.g., Galambos *et al.*, 1972; Zwislocki and Sokolich, 1974), while the loudness of the target was reduced if the masker level was lower than the target level (*loudness decrement*; e.g., Zwislocki and Sokolich, 1974; Elmasian *et al.*, 1980). In the experiments, listeners adjusted the level of a *comparison* presented approximately 1000 ms after the target, until the comparison loudness matched the loudness of the target. Brief tone or noise bursts were used. The comparison was presented at the same frequency as the masker and the target. The upper row labeled "Three-tone task" in Fig. 1 displays a typical temporal structure. Forward maskers higher in level than the target resulted in the comparison level being adjusted to higher levels than in quiet, which was taken as evidence for loudness enhancement of the target. To summarize the experimental results, loudness enhancement increases with the level difference between masker and target, at least for level differences up to 30 dB, amounting to as much as 20 dB (Elmasian and Galambos, 1975; Elmasian *et al.*, 1980). On the other hand, for the masker level fixed to 90 dB sound pressure (SPL), the maximum amount of loudness enhancement was observed at *intermediate* target levels (40–60 dB SPL) in experiments by Zeng (1994) and Plack (1996). At low target levels, the effect of the masker on the loudness level of the target was small, however, representing an analog to the midlevel hump in intensity discrimination (e.g., Zeng *et al.*, 1991; Carlyon and Beveridge, 1993). Oberfeld (2003) showed that forward masking has a significant effect on the loudness of a low-level target, but that this effect is maximal at intermediate masker-target level differences in the range between 20 and 40 dB, resulting in a mid-*difference* hump. Zwislocki and Sokolich (1974) reported that loudness enhancement vanished if the masker-target inter-stimulus interval (ISI) was longer than 400 ms. Loudness enhancement was also observed if the masker followed the target (backward masking; Elmasian *et al.*, 1980; Plack, 1996).

Recently, Scharf *et al.* (2002) raised the question of whether the loudness matches obtained in the three-tone matching task really reflect a change in the loudness of the target. They used a comparison tone much higher in frequency than the masker and the target and found no evidence for loudness enhancement. Scharf *et al.* (2002) interpreted this finding as to showing that in previous studies the masker did not enhance the loudness of the target presented proximally to the masker, but rather caused loudness recalibration (Marks, 1994), or "induced loudness reduction (ILR)," as Scharf *et al.* termed it, in the comparison. The loudness recalibration experiments showed that a moderately strong tone (e.g., 80 dB SPL) reduces the loudness of a weaker tone following with an ISI of more than about 200 ms (for a review see Wagner and Scharf, 2006). The effect was observed only for targets similar in frequency to the masker (Marks,

a)Electronic mail: oberfeld@uni-mainz.de

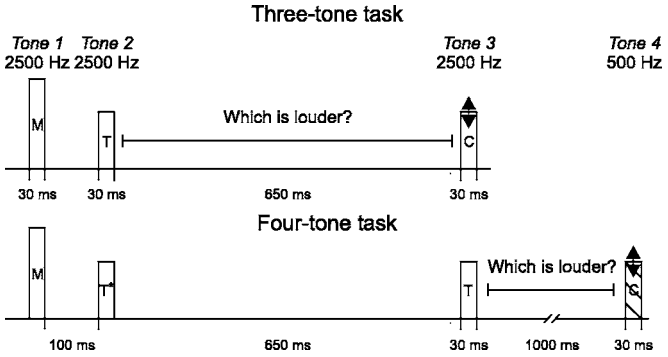


FIG. 1. Trial configurations used in the experiments. In both tasks, the listener decided whether the target tone T or the comparison tone C had been louder. The level of the target was fixed. The level of the comparison was adjusted by an adaptive procedure. The level of the masker M was varied between blocks. In the three-tone task (upper row), the frequency of all tones was 2500 Hz. Listeners produced a loudness match between Tone 2 and Tone 3. In the four-tone task (lower row), listeners produced a loudness match between Tone 3 (2500 Hz) and Tone 4 (500 Hz). The target (Tone 3) was preceded by exactly the same stimuli as the comparison in the three-tone task. Tone 2 (T^*) was identical in level to the target. For the baseline matches, the masker or the masker and the tone T^* were omitted. All stimuli were 30 ms tone bursts.

1994). Scharf *et al.* pointed out that a reduction in comparison loudness would result in the comparison being adjusted to higher levels than the target, that is, in "loudness enhancement" as defined in the three-tone matching task. To explain why the target is not also subject to loudness reduction, however, it was necessary to assume that "*[…] a weak tone presented in close temporal proximity to a stronger tone somehow is protected from ILR*." (Scharf *et al.*, 2002, p. 809). While the reason for such a pattern remained unclear, data compatible with the hypothesis were reported by Arieh and Marks (2003a), who presented the masker (80 dB SPL) and the target (60 dB SPL) at 2500 Hz, and the comparison at 500 Hz. The masker had virtually no effect at masker-target ISIs smaller than 200 ms, but caused a reduction in the loudness level of more than 11 dB at ISIs longer than 500 ms. The average loudness match produced for the masker and the comparison presented at the same frequency corresponded to the average difference between the loudness reduction induced in a target presented 575 or 75 ms after the masker.

These recent data make it necessary to reinterpret the results obtained in experiments using the same frequency for all tones. It remains to be shown, however, whether the entire set of loudness enhancement data can be explained by ILR. For instance, Nieder *et al.* (2003) reported that presenting the masker at either 80 dB SPL or 95 dB SPL produced a difference of only about 2 dB in the loudness reduction induced in 70 dB SPL targets. In contrast, Elmasian and Galambos (1975) found the average loudness matches for a 70 dB SPL target obtained with an 80 dB SPL and a 100 dB SPL masker to differ by about 11 dB. Another important issue is the observation of loudness *decrement* (e.g., Elmasian and Galambos, 1975). The loudness of the target being matched by a comparison *lower* in level than the target could be explained by loudness recalibration only if a masker less intense than the target produced an *increase* in comparison loudness, contrary to results by Mapes-Riordan and Yost (1999).

In the present study, the hypothesis that the masker-induced change in loudness level measured with the masker and the comparison sharing the same frequency is due to a reduction in comparison loudness was tested directly in two experiments. Loudness matches in the traditional three-tone loudness matching procedure, depicted in the upper row of Fig. 1, were obtained for masker-target level combinations found to produce the most pronounced changes in loudness level. For the same listeners, the effect of the masker on the loudness of the comparison (Tone 3 in the lower row of Fig. 1) was measured by obtaining loudness matches between the latter tone and a fourth tone presented at a much lower frequency. This made it possible to compare the change in loudness level in the three-tone task to the change in comparison loudness, for each masker-target level combination. An important detail was that in the four-tone task depicted in the in the lower row of Fig. 1, the same two stimuli as in the three-tone task preceded Tone 3, namely the masker and what had been the target. In the experiments by Arieh and Marks (2003a), only one tone (the masker) preceded the test tone. To estimate the reduction in comparison loudness effective in the three-tone task, however, it is important to use exactly the same temporal configuration.

II. EXPERIMENT 1: CAN "LOUDNESS ENHANCEMENT" BE EXPLAINED BY LOUDNESS RECALIBRATION?

Experiment 1 tested the hypothesis that the loudness matches obtained with the target and the comparison presented at the same frequency can be explained by a reduction in comparison loudness (Scharf *et al.*, 2002). Target level (L_T) was 60 dB SPL, a level at which both considerable "loudness enhancement" (Zeng, 1994; Plack, 1996; Oberfeld, 2003) and ILR (Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003a; Nieder *et al.*, 2003) have been reported. The maximum masker level was 90 dB SPL, corresponding to the condition producing the largest amount of loudness enhancement (Zeng, 1994; Plack, 1996; Oberfeld, 2003). A 70 dB SPL masker was presented because ILR can be expected to be most pronounced for a 10–20 dB level difference between masker and target (Mapes-Riordan and Yost, 1999). To answer the question of whether loudness *decrement* is due to a change in comparison loudness rather than in target loudness, two maskers lower in level than the target (L_M=40 and 50 dB SPL) were included.

A. Method

Nine students at the Johannes Gutenberg—Universität Mainz participated in the experiment voluntarily (eight female, one male, age 20–27 years). They either received partial course credit or were paid for their participation. All reported normal hearing. For the ear tested, detection thresholds were better than 13 dB hearing level (HL) at all octave frequencies between 0.5 and 4 kHz.

All stimuli were pure tones with a steady-state duration of 20 ms, gated on and off with 5 ms $\cos^2$ ramps. On each trial, listeners decided whether the target (the penultimate

tone in the trial) or the comparison (the final tone) had been louder. The level of the target was fixed at 60 dB SPL. The level of the comparison was adjusted by an adaptive procedure. The level of the masker was varied between blocks.

The stimuli were generated digitally, played back via one channel of an RME ADI/S digital-to-analog converter (sampling rate 44.1 kHz, 24 bit resolution), attenuated (TDT PA5), buffered (TDT HB7), and presented to the right ear via Sennheiser HDA 200 headphones calibrated according to IEC 318 (1970). The attenuator setting remained constant within a trial. The experiment was conducted in a single-walled soundproof chamber.

The experiment comprised two tasks. In the *three-tone task* depicted in the upper row of Fig. 1, the frequency of all tones was 2500 Hz. Listeners produced a loudness match between Tone 2 (target) and Tone 3 (comparison). The silent interval between masker offset and target onset was 100 ms. The interval between target offset and comparison onset was 650 ms. In the *four-tone task* (Fig. 1, lower row), listeners produced a loudness match between Tone 3 (target: 2500 Hz) and Tone 4 (comparison: 500 Hz). Tone 3 was preceded by exactly the same two tones as the comparison in the three-tone task. Tone 2 (T^*) was identical in level to the target (Tone 3). The silent interval between Tone 3 and Tone 4 was 1000 ms. For the baseline matches, the masker M or the masker and tone T^* were omitted. Listeners were instructed to ignore the masker and tone T^*. No feedback was provided. The inter-onset interval between the target in a given trial and the target in the following trial was fixed at 5.7 s in both tasks.

A two-interval, two-alternative forced-choice (2I, 2AFC) interleaved-staircase procedure (Jesteadt, 1980) was used. Each experimental block comprised two randomly interleaved tracks. The upper track converged on the comparison level corresponding to the 70.7% *comparison louder* point on the psychometric function. If the listener indicated on two consecutive trials that he or she had perceived the comparison as being louder than the target, the level of the comparison was reduced. After each response indicating that the target had been perceived as being louder than the comparison, the level of the comparison was increased. In the lower track, a 1-down, 2-up rule was used to track the 29.3% *comparison louder* point on the psychometric function. In the three-tone task, the upper and the lower track started with a comparison level 15 dB above and below target level, respectively. In the four-tone task, the two tracks started with the level of Tone 4 15 dB above and below, respectively, the baseline match obtained at the beginning of a given session. The step size was 5 dB until the fourth reversal. The track continued with a step size of 2 dB until ten reversals had occurred or 60 trials had been presented. If in one of the tracks ten reversals had already occurred before the other track had also reached ten reversals, trials from the former track were still presented with an a priori probability of 0.25.

Listeners received only one task in each session. Sessions with the two tasks alternated. In each block, only one masker-target level combination was presented. A session started with a baseline match. In the following blocks, the masker (three-tone task) or the masker and tone T^* (four-tone task) were included. Masker level increased from block to block, starting with $L_M = 40$ dB. Therefore, the largest masker level was always presented at the end of a session. The experiment started with a practice session.

For each block, the arithmetic mean of the level differences between comparison and target ($L_C - L_T$) at all but the first four reversals was computed separately for the upper and for the lower track, with the restriction that for each track an even number of reversals entered the computation (e.g., if 11 reversals had occurred in one of the tracks, reversal 11 was excluded). The arithmetic mean of these two values was taken as the loudness match, corresponding to the comparison level at the point of subjective equality (PSE) minus target level. A run was discarded if the standard deviation of $L_C - L_T$ at the counting reversals was greater than 6 dB in either the upper or the lower track. Three runs were obtained in different sessions for each task and masker level, resulting in a total of six experimental sessions. Time permitting, additional runs were presented in a seventh session if the standard deviation of the loudness matches exceeded 5 dB.

Because the upper and the lower track converge on the 70.7% and the 29.3% *comparison louder* point on the psychometric function, respectively, half the difference between $L_C - L_T$ in the upper track and $L_C - L_T$ in the lower track was taken as a measure of loudness variability, denoted as *jnd* $= (x_{0.707} - x_{0.293})/2$ (Schlauch and Wier, 1987; Zeng, 1994; Plack, 1996).

B. Results and discussion

For the three-tone task, the individual results are displayed in Fig. 2 as the level difference between comparison (Tone 3) and target (Tone 2) at the point of subjective equality (PSE), relative to the baseline match. Positive values indicate that the masker either enhanced the loudness of Tone 2, or reduced the loudness of Tone 3, or both. For the four-tone task, the results are displayed as the level difference between target (Tone 3) and comparison (Tone 4) at the PSE, relative to the baseline match (without masker and tone T^*). Positive values indicate that the masker reduced the loudness of Tone 3. For each block in which a masker was presented, the arithmetic mean of the baseline matches was subtracted from the loudness match; the latter was defined as the arithmetic mean of average $L_C - L_T$ in the upper and in the lower track. For the four-tone task, the resulting value was multiplied by −1.

What can be concluded about the three-tone task (squares in Fig. 2)? For masker levels lower than the target level, most loudness matches (relative to the baseline) were negative. If the masker level was higher than the target level, virtually all loudness matches were positive. Therefore, the traditional interpretation of the results would be that the masker produced loudness decrement or loudness enhancement, depending on its level relative to the target level. The size of the effects was compatible with previous data, with a maximum amount of "loudness decrement" of about 7 dB (Elmasian and Galambos, 1975), a maximum amount of

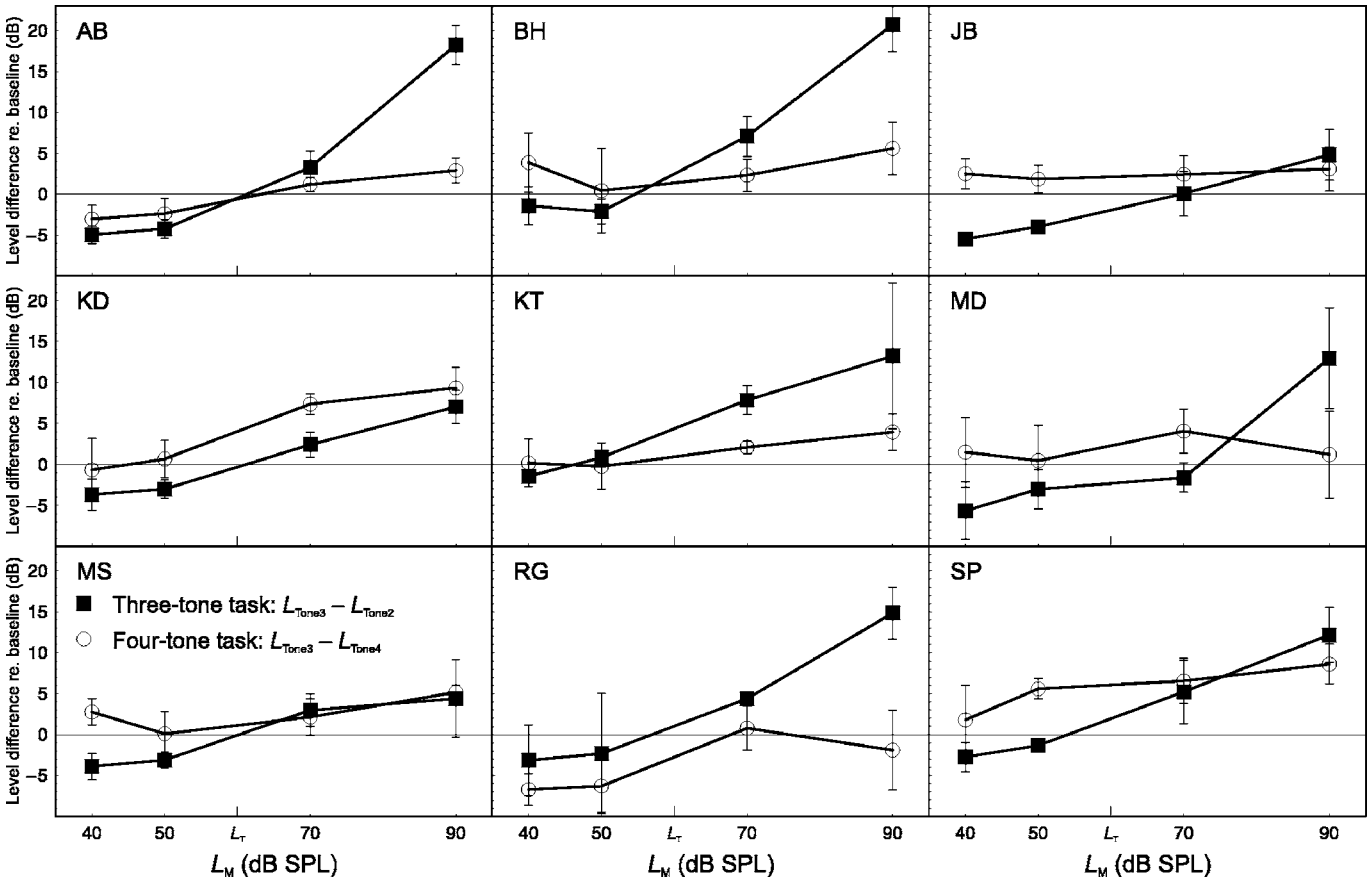


FIG. 2. Experiment 1. Individual masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Squares: level of the comparison (Tone 3) matching the loudness of the target (Tone 2) in the three-tone task, relative to the baseline match. Circles: level of Tone 3 (2500 Hz) matching the loudness of Tone 4 (500 Hz) in the four-tone task, relative to the baseline match. Target level was 60 dB SPL. Panels represent listeners. Error bars show plus and minus one standard deviation of the three or more measurements obtained for each data point.

loudness enhancement of about 20 dB (Elmasian and Galambos, 1975; Oberfeld, 2003), and a considerable amount of inter-individual variability (Plack, 1996).

Can the effects observed in the three-tone task be explained by a change in the loudness of the comparison? If so, the two lines in Fig. 2 representing matches from the two tasks should lie on top of each other. Yet, only a few of the data points are compatible with this prediction. First, for masker levels lower than the target level, the match in the four-tone task (circles in Fig. 2) was close to zero for most listeners (e.g., listener KD), compatible with data by Mapes-Riordan and Yost (1999). Therefore, the *negative* level difference between comparison and target required to produce the loudness match in the three-tone task cannot be explained by a change in comparison loudness in most cases.

For masker levels higher than the target level, the data do in fact indicate a reduction in the loudness of Tone 3 (the comparison in the three-tone task), except for listener RG at

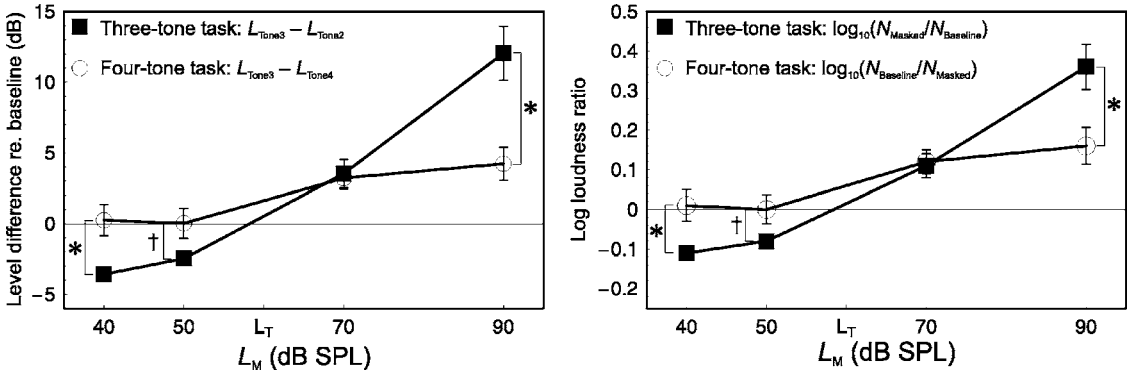


FIG. 3. Experiment 1. Left panel: Mean masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Same symbols as in Fig. 2. Right panel: Mean logarithm of the ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in quiet, as a function of masker level and task. Brackets indicate significant differences (paired-sample t tests; *: $p < 0.05$, †: $p < 0.1$; two tailed). Error bars show plus and minus one standard error of the mean (SEM) of the nine individual estimates.

the highest masker level. The maximum effect was approximately 10 dB. Mean results are displayed in the left panel of Fig. 3. Mean loudness reduction was 4.2 dB (SD=3.5 dB) at L_M=90 dB SPL. This value is smaller than the 11 dB of loudness reduction Arieh and Marks (2003a) found for an 80 dB SPL masker combined with a 60 dB SPL target (50 ms, 2500-Hz tone bursts), but comparable to the amount of ILR reported by Mapes-Riordan and Yost (1999) for the same level combination and 500 ms, 2500 Hz tones. It therefore remains unclear whether the tone T^* interpolated between masker and target in the four-tone task reduced the amount of ILR. With the 90 dB SPL masker, the reduction in Tone 3 loudness was considerably smaller than the loudness enhancement of Tone 2 observed in the three-tone task for six of the nine listeners. With the 70 dB SPL masker, the difference between the two measures was generally smaller.

A repeated-measures analysis of variance (ANOVA) with Huynh-Feldt correction for the degrees of freedom was conducted. The factors were masker level and task. There was a significant effect of masker level [$F(3,24)=72.8$, $p=0.001$, $\tilde{\varepsilon}=0.7$]. As a posthoc analysis, two separate ANOVAs with the factor masker level were conducted. The effect of masker level was significant for both the three-tone and the four-tone task [$F(3,24)=59.83$, $p=0.001$, $\tilde{\varepsilon}=0.46$ and $F(3,24)=11.10$, $p=0.001$, $\tilde{\varepsilon}=0.95$, respectively]. One-sample t tests were conducted for each data point. For the three-tone task, all matches differed significantly from 0 dB ($p<0.05$, two tailed). For the four-tone task, the matches were significantly different from 0 dB only if the masker level was higher than the target level, confirming the observation by Mapes-Riordan and Yost (1999) that a masker lower in level than the target produces no ILR. Loudness decrement thus reflects a reduction in the loudness of the target presented proximally to the masker.

In the two-factorial ANOVA, the effect of task was not significant [$F(1,8)=0.124$]. There was a significant Masker Level×Task interaction, however, demonstrating that the matches of Tone 3 versus Tone 2 (three-tone task) and Tone 3 versus Tone 4 (four-tone task) were not identical [$F(3,24)=20.0$, $p=0.001$, $\tilde{\varepsilon}=0.71$]. Posthoc pairwise comparisons indicated that the two matches obtained at each masker level differed significantly at a masker level of 40 dB SPL [$t(8)=-3.2$, $p=0.012$], and marginally significantly at a masker level of 50 dB SPL [$t(8)=-2.3$, $p=0.054$]. These results confirm that loudness decrement cannot be explained by a change in comparison loudness. At a masker level of 70 dB SPL, the difference between the two different matches was not significant [$t(8)=0.22$]. This observation is compatible with data by Arieh and Marks (2003a) obtained for an 80−dB SPL masker combined with a 60 dB SPL target. For the 90 dB SPL target, however, there was a significant difference between the loudness match obtained in the three-tone and in the four-tone task [$t(8)=3.2$, $p=0.013$]. Thus, the masker had an effect on target loudness at this masker level, contrary to the hypothesis by Scharf *et al.* (2002).

The upper panel of Fig. 4 shows the loudness matches in the three-tone task plotted against the loudness matches in the four-tone task. Scharf *et al.* (2002) predicted that the reduction in Tone 3 loudness measured in the four-tone task

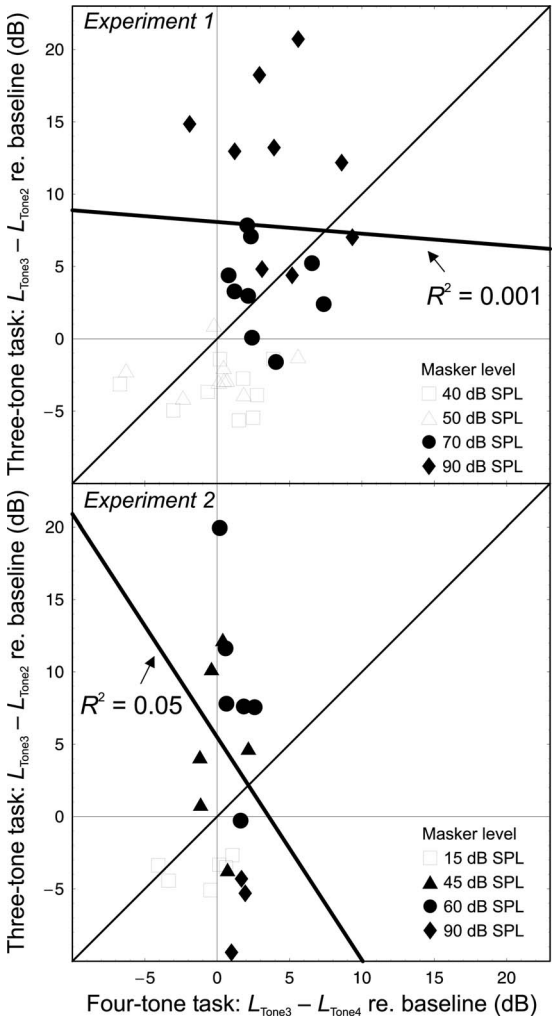


FIG. 4. Scatterplot of individual masker-induced changes in the loudness level of Tone 3 in the four-tone task (x axis), and of Tone 2 in the three-tone task (y axis). Upper panel: Experiment 1. Lower panel: Experiment 2. Each data point represents one listener and one masker level. The different masker levels are indicated by the different symbols. The thick line shows the best fitting linear regression line, computed only for the data points where masker level was higher than target level (filled symbols). All of these data points should lie on the diagonal if the loudness match in the three-tone task was determined completely by the loudness reduction of the comparison.

be identical to the change in loudness level measured in the three-tone task, that is, that all data points where the masker was higher in level than the target (filled symbols in Fig. 4) should lie on the diagonal. The thick line shows the best-fitting linear regression line. The Pearson product-moment correlation was not significant ($r=-0.04$, $N=18$), confirming the pronounced differences between the loudness matches in the two tasks. In an additional analysis, the data points where $L_M<L_T$ were also included. The correlation between the two measures was significant ($r=0.40$, $N=36$, $p=0.017$), but the proportion of the variance of the three-tone matches explained by the four-tone matches was still only $R^2=0.16$.

Taken together, the data show that the effects of a forward masker on the loudness matches in experiments presenting the masker and the comparison at the same frequency can only partly be attributed to an effect on comparison loudness. Thus, a forward masker induces loudness changes in a

proximal target, but the loudness matches are contaminated by ILR of the comparison if the masker and the comparison share the same frequency.

There are three issues which might be a reason to qualify the above interpretation of the loudness matches. The first issue arises because the loudness matches in the three-tone and in the four-tone task were obtained for different comparison frequencies. The second issue is a potential effect of the sequence of comparison levels presented during the course of a session or during an experimental block. The third issue is related to the variation in the level of Tone 3 introduced by the adaptive procedure in the three-tone task.

Concerning the first issue, from the difference between the comparison level at the PSE in forward masking and the comparison level at the PSE in quiet observed in the three-tone task on the one hand and the four-tone task on the other hand, the conclusion was drawn that the effect of the masker on the loudness level of the target was larger than can be explained by a reduction in comparison loudness. However, the direct comparison of the level difference in dB at 2500 and 500 Hz is valid only if the slope of the loudness functions (Lf) for 500 and 2500 Hz tones is identical. A recent estimate of the slope of the Lf at various frequencies has been provided by Suzuki and Takeshima (2004), who estimated the exponent of the power function relating sound intensity and loudness at levels above 30 dB SPL to be 0.31 at 500 Hz and 0.29 at 2500 Hz (see Fig. 7 in Suzuki and Takeshima, 2004). To test for a potential effect of this difference in the exponents on the conclusion presented above, the ratio between the comparison loudness at the PSE in forward masking and the comparison loudness at the PSE in quiet (baseline condition) was estimated for each individual block. The loudness function proposed by Zwislocki (1965) was used, because it provides a better account of loudness near threshold than the loudness function used by Suzuki and Takeshima (cf. Buus *et al.*, 1998). According to Zwislocki (1965)

$$N = a[(p^2 + 2.5p_t^2)^{\alpha} - (2.5p_t^2)^{\alpha}], \tag{1}$$

where N is loudness, p is sound pressure, p_t is sound pressure at the detection threshold, α is the frequency dependent slope of the Lf, and a is a scale constant. Individual detection thresholds for 500 and 2500 Hz tones (duration 30 ms, including 5 ms $\cos^2$ ramps) in quiet were used. As no detection thresholds were available for listener RG at 2500 Hz and for listener KT at both frequencies, the average threshold of the remaining listeners at the respective signal frequency was used in these three cases. The ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in the baseline condition was estimated using Eq. (1)

$$\frac{N_{\text{Masked}}}{N_{\text{Baseline}}} = \frac{(p_{\text{Masked}}^2 + 2.5p_t^2)^{\alpha} - (2.5p_t^2)^{\alpha}}{(p_{\text{Baseline}}^2 + 2.5p_t^2)^{\alpha} - (2.5p_t^2)^{\alpha}}, \tag{2}$$

where p_{Masked} and p_{Baseline} denote the sound pressure of the comparison matching the loudness of the target in the presence of the forward masker and in quiet, respectively. Because the ratio $N_{\text{Masked}}/N_{\text{Baseline}}$ ranges from unity to infinity for N_{Masked} greater than N_{Baseline}, but only from zero to unity for N_{Masked} smaller than N_{Baseline}, the logarithm of the ratio was used in the analyses. For the data obtained in the four-tone task, the ratio $N_{\text{Masked}}/N_{\text{Baseline}}$ was inverted so that positive values of the log loudness ratio indicate a masker-induced reduction in the loudness of Tone 3.

As the right panel of Fig. 3 shows, the mean log loudness ratios exhibited the same pattern as the masker-induced changes in loudness level displayed in the left panel. A repeated-measures ANOVA with the factors masker level and task was conducted. There was a significant effect of masker level [$F(3,24)=66.9$, $p=0.001$, $\tilde{\varepsilon}=0.75$]. The Masker Level × Task interaction was also significant [$F(3,24)=13.9$, $p=0.001$, $\tilde{\varepsilon}=0.74$], demonstrating that the masker-induced loudness changes were not identical in the two tasks. The effect of task was not significant [$F(1,8)=0.01$]. Post-hoc pairwise comparisons indicated that the log loudness ratios obtained in the two tasks differed significantly at a masker level of 40 dB SPL [$t(8)=-3.1$, $p=0.015$], and at a masker level of 90 dB SPL [$t(8)=2.5$, $p=0.036$]. At a masker level of 50 dB SPL, the difference was marginally significant [$t(8)=-2.2$, $p=0.061$]. At a masker level of 70 dB SPL, the difference was not significant [$t(8)=-0.26$]. This is the same pattern of statistical results that had been observed for the changes in loudness level.

Concerning the second issue, in a paper published after Experiments 1 and 2 had been completed, Epstein and Gifford (2006) reported that obtaining the baseline match at the beginning of a session may result in an underestimation of ILR. Because the comparison level will be higher during the baseline match than in a subsequent block in which a condition producing ILR is presented, the comparisons in the second block might be subject to ILR caused by the comparisons presented in the first block, due to the slow recovery from ILR (Arieh *et al.*, 2005; Epstein and Gifford, 2006). Epstein and Gifford found the estimate of ILR to be about 3 dB smaller if the experimental condition (80 dB SPL masker, 70 dB SPL target) was run immediately after the baseline match rather than after a delay of 15 or 120 min. For the present experiment, there might thus have been a carry over effect from the first block (baseline) to the second block (40 dB SPL masker). Moreover, as the loudness match with the 40 dB SPL and the 50 dB SPL masker was rather similar to the baseline match for most listeners, a carry over effect from Block 2 to Block 3, and from Block 3 to Block 4 cannot be precluded either. On the other hand, as the minimum duration of Block 4 presenting the 70 dB SPL masker was 5 min, which is larger than the ILR recovery time of 130 s (Arieh *et al.*, 2005), it can be assumed that the problem was not relevant for the block presenting the 90 dB SPL masker. Even if it had been, the 1–4 dB underestimation of ILR reported by Epstein and Gifford (2006) cannot account for the 8 dB discrepancy between the loudness reduction of Tone 3 (measured in the four-tone task) and the change in loudness level observed in the three-tone task. The results by Epstein and Gifford also point to a potential problem associated with the interleaved-staircase procedure used in the current study as well as by Mapes-Riordan and Yost (1999). The comparisons in the upper track are on average higher in level

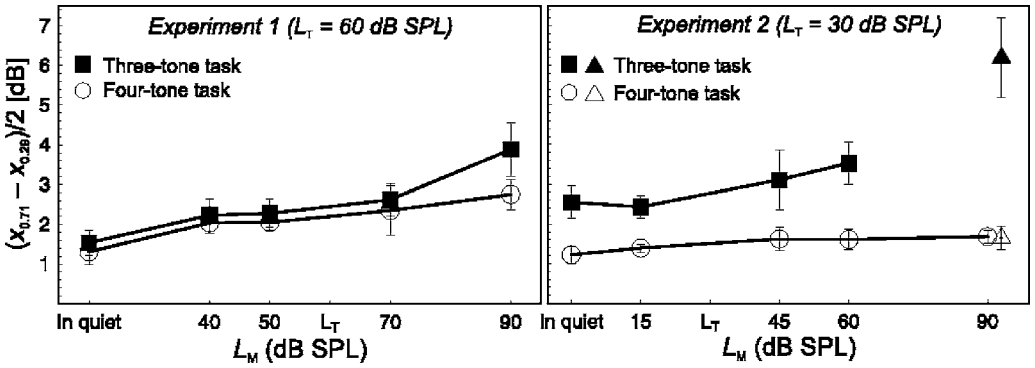


FIG. 5. Experiment 1. Mean loudness variability measured as half the difference between average reversal level in the upper and in the lower track $[(x_{0.71} - x_{0.29})/2]$, as a function of masker level and task. Squares: three-tone task. Circles: four-tone task. Left panel: Experiment 1. Right panel: Experiment 2. Only three listeners were tested with the 90 dB SPL masker in the three-tone task. The filled and open triangles show their average data in the three-tone task and the four-tone-task, respectively. Error bars show ±1 SEM.

than the comparisons in the lower track, so that the latter might be subject to ILR. The same effect would also apply to the baseline matches, however. Because ILR is defined as the loudness match in a block presenting a masker minus the baseline match, a potential loudness reduction of the tones in the lower track would cancel, provided that the level difference between upper and lower track does not differ between the forward masked conditions and the baseline condition. The average difference (with the *SD* in parentheses) between mean comparison level in the upper track and mean comparison level in the lower track was 8.4 dB (2.8 dB), 9.2 dB (2.8 dB), and 7.9 dB (2.7 dB) for blocks presenting the 70 dB SPL masker, the 90 dB SPL masker, and the baseline condition, respectively. Thus, the level difference in blocks presenting the 90 dB SPL masker was on average 1.4 dB larger than in the baseline condition, but it is unlikely that this small difference significantly affected the estimate of loudness reduction induced in Tone 3. Note also that because the loudness match in a given block is defined as the arithmetic mean between the average reversal levels in the two tracks, any influence of a change in loudness level of the lower track on the estimate will be attenuated by a factor of 2.

Turning to the third issue, the induced reduction in Tone 3 loudness was measured in the four-tone task with the level of Tone 3 fixed at target level. In the three-tone task, however, the level of Tone 3 (the comparison) was varied by the adaptive procedure, so that it could have levels above and below target level. If now the 90 dB SPL masker caused a disproportionately greater amount of ILR at lower levels of Tone 3, then the reduction in the loudness of Tone 3 could have been greater in the three-tone task than in the four-tone task, due to Tone 3 being presented at levels below the target level in the lower track of a three-tone block. This would result in the impression of additional loudness enhancement above that predicted by ILR on the four-tone task. To estimate the importance of this issue, mean comparison level in the lower track was computed for each block obtained in the three-tone task, for all trials following the fourth reversal (i.e., the part of the track that was used in the calculation of the loudness match $L_C - L_T$). At the 90 dB SPL masker level, mean comparison level in the counting part of the lower track was lower than target level in 11 of the total of 32 blocks only, with a minimum value of 7.1 dB below target level. In 18 blocks, on the other hand, mean comparison level in the lower track was at least 5 dB higher than target level. Thus, on average ILR of Tone 3 should have been *smaller* in the three-tone than in the four-tone task. With the 11 blocks in which mean comparison level in the counting part of the lower track was lower than target level excluded, the mean loudness match (relative to the baseline condition) obtained with the 90 dB SPL masker was 14.7 dB ($SD = 4.9$ dB) and 4.4 dB ($SD = 3.7$ dB) for the three-tone and the four-tone task, respectively. The difference between these two matches was significant [$t(7) = 4.05$, $p = 0.005$; note that only eight of the nine listeners contributed to this analysis], just as in the original analysis presented above. The variation in the level of Tone 3 in the three-tone task thus has no implications for the interpretation of the results.

The mean loudness variability is displayed in the left panel of Fig. 5. Carlyon and Beveridge (1993) suggested that in intensity discrimination experiments, the *jnd* elevation caused by a forward masker might be related to the masker-induced loudness change. In fact, loudness enhancement and intensity-difference limens or loudness variability were found to be (weakly) correlated (Zeng, 1994; Plack, 1996; Oberfeld, 2005). For the present data, the average increase in the loudness variability observed with the 90 dB SPL masker was only 2.3 dB in the three-tone task, while elevations of 4–15 dB in the intensity DL have been observed for 80–90 dB SPL maskers combined with a 60 dB SPL standard (e.g., Plack *et al.*, 1995; Oberfeld, 2005). The data speak against a one-on-one relation between the masker-induced loudness change and loudness variability: Paired-sample *t* tests indicated that for both tasks, the loudness variability was significantly larger in forward masking than in quiet at all masker levels except the lowest ($p < 0.05$, two tailed). Thus, there was an effect on the loudness variability in conditions producing loudness enhancement, ILR, and virtually no effect on the loudness (cf. Fig. 3, left panel). For the three-tone task, a repeated-measures ANOVA conducted for the forward-masked conditions showed that the increase in the loudness variability with masker level was significant

$[F(3,24)=4.4,\ p=0.038,\ \tilde{\varepsilon}=0.56]$. For the four-tone task, an ANOVA conducted for the data obtained under forward masking showed no significant effect of masker level $[F(3,24)=1.1]$.

III. EXPERIMENT 2: ENHANCEMENT VERSUS RECALIBRATION AT A LOW TARGET LEVEL

Marks (1996), Arieh and Marks (2003a), and Wagner and Scharf (2006) suggested that ILR occurs only if the masker is presented at a relatively high sound pressure level (i.e., above 60 dB SPL). It therefore seemed unlikely that the change in the loudness level of a 25 dB SPL target caused by 40 and 55 dB SPL maskers reported by Oberfeld (2003) was due to a reduction in comparison loudness. In Experiment 2, exactly the same design as in Experiment 1 was used, but a 30 dB SPL target was presented. Masker-target level differences ranged from −15 to +60 dB. The effect of the masker on the loudness level in the three-tone task was expected to be most pronounced at intermediate masker-target level differences (Oberfeld, 2003).

A. Method

The same stimuli, apparatus and procedure as in Experiment 1 were used, except for the lower target level and the different masker levels.

Seven students took part in the experiment voluntarily; only one of them (KD) had participated in Experiment 1. The listeners either received partial course credit or were paid for their participation. All reported normal hearing. For the ear tested, detection thresholds were better than 10 dB HL at all octave frequencies between 0.5 and 4 kHz. One of the listeners showed a systematic shift of the loudness matches in the three-tone task during the course of the experiment. With the 60 dB SPL masker, for instance, she adjusted the comparison to a level 33.7 dB above target level in the first session presenting the three-tone task. In the following sessions, the level difference between comparison and target required for the loudness match gradually shifted toward negative values. In the last session presenting the three-tone task, it was −13 dB, resulting in a range of more than 45 dB. A comparable pattern was observed with the 15 dB SPL and the 45 dB SPL masker in the three-tone task. The data of this listener were excluded from the analysis. The remaining listeners (five female, one male) ranged in age between 19 and 26 years.

Detection thresholds were obtained for 500 Hz and 2500 Hz tones with a duration of 30 ms including 5 ms $\cos^2$ ramps. A 2I, 2AFC, adaptive procedure was used (3–down, 1–up rule; Levitt, 1971). Three runs were obtained per condition. For the 500 Hz tones presented in quiet, the individual thresholds ranged from 11.9 to 21.6 dB SPL ($M=17.2$ dB SPL, $SD=4.0$ dB). For the 2500 Hz tones presented in quiet the individual thresholds ranged from 4.5 to 9.6 dB SPL ($M=7.3$ dB SPL, $SD=2.1$ dB). The 2500 Hz tones were also presented with 30 ms, 2500 Hz foward maskers and a masker-signal ISI of 100 ms. Mean thresholds (with *SD*s in parentheses) were 16.9 dB SPL (7.9 dB) and 19.1 dB SPL (8.3 dB) at a masker level of 60 dB SPL and 90 dB SPL, respectively. For listener KS, the average threshold measured with the 90 dB SPL masker was 31.6 dB SPL and thus above the level of the target in the loudness matching task. For listener TG, even the 60 dB SPL masker caused a large threshold elevation to 29.6 dB SPL, although she reported to clearly hear the target in three-tone loudness matching blocks presenting this masker level. No threshold was obtained for this listener in the condition presenting the 90 dB SPL masker. Listener MM reported that she was not able to distinguish the target from the 90 dB SPL masker in the three-tone task, but only perceived some sort of echo, even though the 30 dB SPL target was well above her forward masked detection threshold (19.4 dB SPL). These three listeners were not tested with the 90 dB SPL masker in the three-tone task.

B. Results

Individual data are shown in Fig. 6. In the three-tone task and at a masker level of 15 dB SPL, all listeners adjusted the level of the comparison to a lower level than in the baseline condition, indicating loudness decrement (squares in Fig. 6). For the 45 dB SPL and the 60 dB SPL masker, the level of the comparison matching the loudness of the target was higher than in the baseline condition for all listeners but SD, indicating loudness enhancement. Note that only three listeners were tested with the 90 dB SPL masker. The matches indicated a reduction in loudness rather than loudness enhancement in this condition. These results are in accordance with the expected mid-difference hump.

In the four-tone task (circles in Fig. 6), the change in loudness level effected by the masker was smaller than in the three-tone task. It was also smaller than for the 60 dB SPL target presented in Experiment 1. This observation is compatible with the assumption that ILR is most pronounced for targets presented at an intermediate level (Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003a; Wagner and Scharf, 2006), although no study systematically measured ILR for a low-level target combined with different masker levels. The data are not compatible with the hypothesis by Scharf *et al.* (2002), according to which in Fig. 6, the two lines representing matches from the two tasks should lie on top of each other.

Mean data are shown in the left panel of Fig. 7. An ANOVA with the within-subjects factors masker level and task showed a significant Masker Level × Task interaction $[F(2,10)=16.6,\ p=0.002,\ \tilde{\varepsilon}=0.79]$. The data obtained with the 90−dB SPL masker were excluded from this analysis because only three of the six listeners had been tested in this condition. Pairwise comparisons indicated that the matches obtained in the two tasks differed significantly at a masker level of 15 dB SPL $[t(5)=-3.2,\ p=0.024]$, 60 dB SPL $[t(5)=-2.65,\ p=0.046]$, and 90 dB SPL $[t(2)=-6.0,\ p=0.027]$. For the 45 dB SPL masker, the difference was not significant $[t(5)=1.84]$. The main effect of task was not significant $[F(1,5)=2.7]$. There was a significant main effect of masker level $[F(2,10)=21.1,\ p=0.001,\ \tilde{\varepsilon}=0.81]$. As a posthoc analysis, two separate repeated measures ANOVAs with the factor masker level were run. The effect of masker level was significant for both the three-tone and the four-tone task

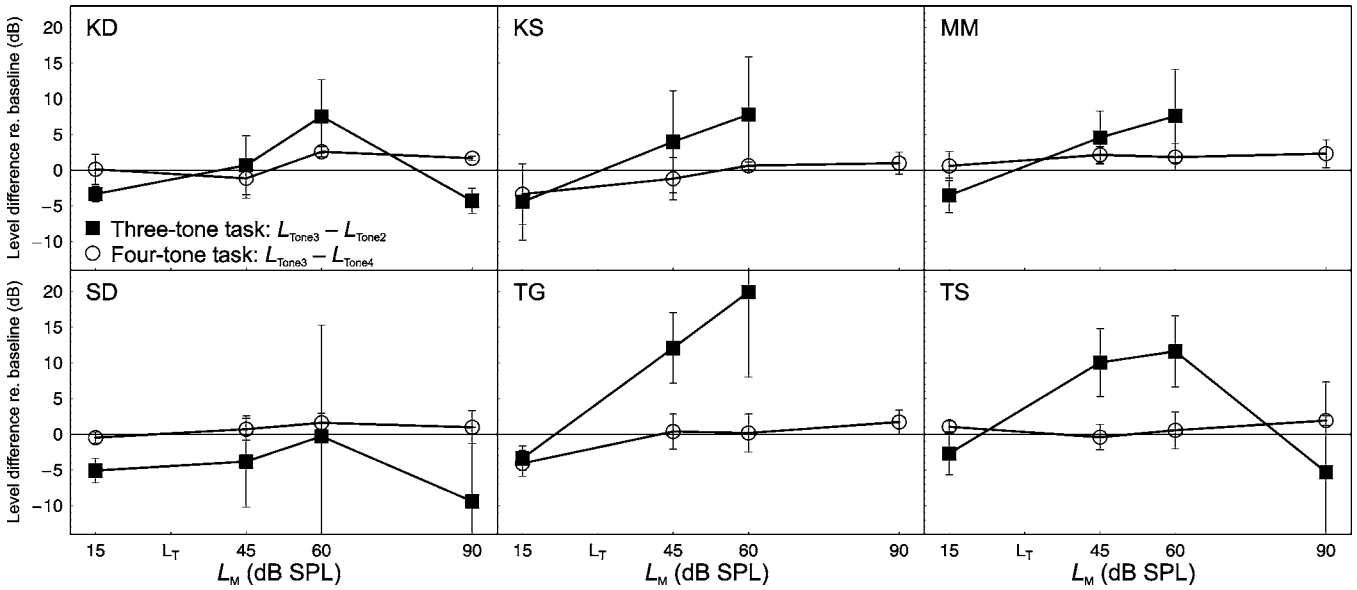


FIG. 6. Experiment 2. Individual masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Target level was 30 dB SPL. Same format as Fig. 2.

$[F(2,10)=21.6,\ p=0.001,\ \tilde{\varepsilon}=0.72$ and $F(3,15)=6.2,\ p=0.006,\ \tilde{\varepsilon}=1.0$, respectively]. One-sample t tests conducted for each data point showed that for the three-tone task, only the matches obtained with the 15 dB SPL and the 60 dB SPL masker differed significantly from 0 dB. For the four-tone task, the matches indicated significant loudness reduction only for the 60 dB SPL and the 90 dB SPL masker.

The lower panel of Fig. 4 shows the loudness matches in the three-tone task plotted against the loudness matches in the four-tone task. Each data point represents one masker level and one listener. For the data points where the masker was higher in level than the target (filled symbols), the correlation between the two measures was not significant ($r=-0.23$, $N=15$). The thick line shows the corresponding regression line. If the data points where $L_M < L_T$ were also included, the correlation between the two measures was again not significant ($r=0.14$, $N=21$).

The data were additionally analyzed in terms of the ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in quiet, using Eq. (2) and individual detection thresholds. As it can be seen in the right panel of Fig. 7, the mean log loudness ratios exhibited approximately the same pattern as the loudness matches displayed in the left panel of Fig. 7. In a repeated-measures ANOVA with the factors masker level and task, the Masker Level × Task interaction was significant $[F(2,10)=12.2,\ p=0.005,\ \tilde{\varepsilon}=0.77]$, indicating that the masker-induced loudness changes were not identical in the two tasks. Note that the data obtained with the 90 dB SPL masker were again excluded from the analysis. Posthoc pairwise comparisons indicated that the log loudness ratios observed in the two tasks differed significantly at a masker level of 15 dB SPL $[t(5)=-2.91,\ p=0.033]$, and 90 dB SPL $[t(2)=-5.96,\ p=0.027]$. For the 45 dB SPL masker, the difference was not significant $[t(5)=1.62]$. This pattern was also observed in the analysis of the masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task). Unlike in the latter analysis, however, the difference was not

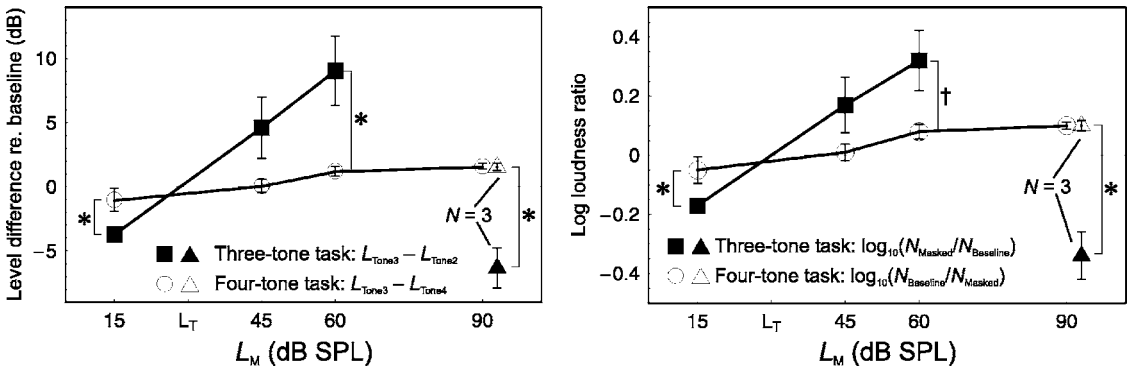


FIG. 7. Experiment 2. Left panel: Mean masker-induced changes in the loudness level of Tone 2 (three-tone task) and of Tone 3 (four-tone task), as a function of masker level and task. Right panel: Mean logarithm of the ratio between comparison loudness at the PSE in forward masking and comparison loudness at the PSE in quiet (baseline condition), as a function of masker level and task. Squares: three-tone task. Circles: four-tone task. Only three listeners were tested with the 90 dB SPL masker in the three-tone task. The filled and open triangles show their average data in the three-tone task and the four-tone-task, respectively. Brackets indicate significant differences (*: $p<0.05$, †: $p<0.1$; two tailed). Error bars show ±1 SEM of the individual estimates.

significant at the 60 dB SPL masker level [$t(5)=-2.01$, $p=0.10$]. For this masker level, the significant difference between the change in the loudness level of the target in the three-tone task and the change in the loudness level of Tone 3 in the four-tone task must thus be viewed with some caution.

As discussed above, the fact that the level of Tone 3 was fixed in the four-tone task but varied in the three-tone task presents a potential problem for the interpretation of the data. At the 60 dB SPL masker level, at which the masker-induced change in loudness level was significantly larger in the three-tone than in the four-tone task, the mean comparison level in the counting part of the lower track was lower than the target level in ten of the total of 24 blocks obtained in the three-tone task. With the former 10 blocks excluded, the mean loudness match obtained with the 60 dB SPL masker was 14.9 dB ($SD=5.9$ dB) and 1.2 dB ($SD=0.9$ dB) for the three-tone and the four-tone task, respectively. The difference between these two matches was significant [$t(5)=13.7$, $p=0.004$]. Thus, the difference between the masker-induced changes in loudness level observed in the two tasks cannot be attributed to the variation in comparison level in the three-tone task.

Taken together, the results from Experiment 2 (30 dB SPL target) confirm the conclusion for the 60 dB SPL target (Experiment 1) that the masker-induced changes in loudness level measured in a three-tone matching task with comparison frequency equal to masker frequency cannot be explained exclusively by a reduction in the loudness of the comparison. Instead, the data demonstrate that the masker had an effect on the loudness of the proximal target.

The mean loudness variability is displayed in the right panel of Fig. 5. An ANOVA with the within-subjects factors masker level and task was conducted. The data obtained with the 90 dB SPL masker were excluded from the analysis. The effect of masker level was only marginally significant [$F(3,15)=2.9$, $p=0.071$, $\tilde{\varepsilon}=1.0$]. The *jnd* obtained with the 90 dB SPL masker in the three-tone task for three listeners did not differ significantly from the *jnd* in quiet, either [$t(2)=2.85$]. The effect of the forward masker on the loudness variability was thus generally smaller than in Experiment 1, which is compatible with reports that an intense masker (e.g., 90 dB SPL) has a larger effect on the loudness variability for midlevel than for low-level targets (Zeng, 1994; Plack, 1996; Oberfeld, 2005). The loudness variability was significantly larger in the three-tone than in the four-tone task [$F(1,5)=11.3$, $p=0.020$]. The Masker Level × Task interaction was not significant [$F(3,15)=0.6$].

IV. GENERAL DISCUSSION

The present study tested the hypothesis by Scharf *et al.* (2002) that the change in loudness level measured in experiments presenting the masker and the comparison at the same frequency does not reflect an effect of the masker on the loudness of the proximal target, but rather a reduction in the loudness of the comparison. The results show that a forward masker does induce changes in target loudness, although Scharf *et al.* (2002) conjectured correctly that there are also effects on comparison loudness. Maskers lower in level than the target had no effect on comparison loudness, so that the observed loudness decrement must represent a change in target loudness. For a masker-target level difference of 30 dB, the increase in the loudness level of the target observed in the three-tone task was significantly larger than the loudness reduction of the comparison, demonstrating loudness enhancement of the target. The correlation between the change in the loudness level measured in the three-tone task and the reduction of the comparison loudness level was not significant.

A. Two-process model for the loudness changes caused by a proximal sound

The data collected in the current study are evidence supporting the hypothesis by Arieh and Marks (2003a) and Arieh *et al.* (2004) that the masker triggers two processes, loudness enhancement and loudness recalibration. They can also be used to refine the two-process model. Experiment 1 showed that for a 70 dB SPL masker combined with a 60 dB SPL target, the loudness match in the three-tone task can be explained by a loudness reduction of the comparison. This result is compatible with the observation by Arieh and Marks (2003a) that at a masker-target ISI of 75 ms, an 80 dB SPL masker had no effect on the loudness match between a 60 dB SPL target and a comparison presented at a much lower frequency. Arieh and Marks noted that this finding can be interpreted in two different ways: either both loudness enhancement and loudness recalibration of the target were absent, or both effects were present but equally strong, so that they cancelled. Put differently, Arieh and Marks (2003a) suggested that the masker might trigger two processes, both with a fast onset: loudness enhancement with a decay time of some 100 ms, and loudness recalibration with a decay time of several seconds. This would account for their finding of only very small loudness changes at masker-target ISIs smaller than 200 ms, without the need to assume a delayed onset of loudness recalibration, which would be at odds with the fast onset of inhibitory processes observed at various stages of the auditory system (see Arieh and Marks, 2003a, for a discussion). Moreover, Arieh *et al.* (2004) reported that an 80 dB SPL masker caused "residual loudness recalibration." In the first block of their experiment, the loudness of a 60 dB SPL target was unaffected when it followed the masker by 100 ms. However, a directly following experimental block in which the masker was omitted showed a reduction in target loudness relative to the baseline match. To explain the latter result under the assumption that the masker had caused only recalibration, but no enhancement, it would be necessary to assume that in the first block, the maskers triggered the inhibitory process, but that the target following the masker by 100 ms was somehow protected from its effect (Scharf *et al.*, 2002). Additionally, why should a 70 or an 80 dB SPL masker cause no loudness enhancement in a 60 dB SPL target, while Experiment 1 demonstrated that a 90 dB SPL masker does? The two-process hypothesis can resolve this puzzle, simply by assuming that for maskers 10–20 dB higher in level than a mid-level target, loudness recalibration and loudness enhancement cancel.

The results obtained in the present study demonstrate that the dependence of recalibration and enhancement on the masker-target level combination is not identical. For instance, a 90 dB SPL masker causes more enhancement than recalibration in a 60 dB SPL target. The two-process model can thus be refined as follows. The process causing loudness enhancement or loudness decrement (Elmasian *et al.*, 1980) is effective if two tones are presented within a temporal window of about 400 ms (Zwislocki and Sokolich, 1974; Arieh and Marks, 2003a), because loudness enhancement is observed with both forward and backward maskers (e.g., Elmasian and Galambos, 1975). The effect is maximal at intermediate masker-target level differences (Experiment 2; Zeng, 1994; Plack, 1996; Oberfeld, 2003). It can be assumed that loudness enhancement is effective only if the masker and the target are similar in frequency (Zwislocki and Sokolich, 1974). These effects cannot be explained by mechanisms located in the auditory periphery, because at early processing stages, forward maskers have been found to *reduce* rather than to enhance the neural response to a following test tone (e.g., Bauer *et al.*, 1975; Harris and Dallos, 1979; Relkin *et al.*, 1995), and because it does not seem possible that a backward masker following the target by 100 ms alters the representation of the target in the auditory nerve. A model based on more centrally located mechanisms was proposed by Elmasian *et al.* (1980). They suggested that the loudness representations of the masker and the target are merged automatically. Applied to the three-tone matching task, it follows that the initial value of target loudness is no longer available at the presentation of the comparison, but that the listener will instead compare a weighted average of the masker loudness and the target loudness with the loudness of the comparison. The mergence hypothesis can explain why the loudness is reduced if the masker is less intense than the target, while a more intense masker results in enhancement—in other words, why the loudness of the target always seems to be shifted towards masker loudness.[2] On the other hand, the mergence hypothesis alone cannot account for the mid-level hump in loudness enhancement (Zeng, 1994; Plack, 1996). Given a constant masker level of, e.g., 90 dB SPL, loudness enhancement should increase monotonically with decreasing target level if a simple weighted average between masker loudness and target loudness was used in the loudness match. The same argument applies to the mid-difference hump observed for three listeners in Experiment 2. Oberfeld (2005) proposed that it is possible to resolve these problems by assuming that the effect of the masker depends on the *perceptual similarity* between masker and target, that is, that the masker loudness will receive a smaller weight if the masker and the target differ strongly in, e.g., spectral content, duration, or loudness. Effects of the masker-standard similarity on intensity-difference limens were reported by Schlauch *et al.* (1997, 1999). The finding by Elmasian and Galambos (1975) that a diotic masker combined with a monaural target produced a smaller amount of loudness enhancement than an ipsilateral masker is also compatible with a similarity effect.

The second process, causing loudness recalibration, is assumed to have a fast onset, but a slow decay in the order of several seconds (Arieh and Marks, 2003a; Arieh *et al.*, 2005). It is effective only for tones similar in frequency (Marks, 1994), and only for masker durations longer than or equal to target duration (Nieder *et al.*, 2003). It reaches its maximum at masker levels 10–20 dB above target level and does not further increase with masker level (Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003). The effect is larger if the target is presented at intermediate rather than at low levels (Experiments 1 and 2; Mapes-Riordan and Yost, 1999). Loudness recalibration is viewed as a centrally based, adaptation-like process (Marks, 1996; Arieh and Marks, 2003b), although the exact nature of the mechanism remains unclear (cf. Wagner and Scharf, 2006).

Under the assumption that loudness enhancement decays within about 400 ms following masker onset, but that ILR remains constant for several seconds, it is possible to independently estimate the amount of loudness enhancement and ILR a masker causes in a proximal target. The estimate of ILR would be the loudness reduction measured with the masker-target ISI well above 500 ms in a three-tone loudness matching task presenting the comparison at a different frequency than the masker. In this case, loudness enhancement of the target can be assumed to be absent. The estimate of loudness enhancement would be based on the loudness change measured at the short masker-target ISI in the same three-tone loudness matching task. Because according to the two-process model the masker causes both loudness enhancement and ILR in a proximal target, the estimate of loudness enhancement would be the change in loudness level at the short interval *plus* the change in loudness level measured with the long masker-target ISI. According to this rationale, loudness matches (with masker and comparison differing in frequency) obtained by Arieh and Marks (2003a) for a masker-target ISI of 75 and 1650 ms indicate that an 80 dB SPL masker caused on average 11.1 dB of ILR in a 60 dB-target, and 11.2 dB of loudness enhancement. The latter value is roughly compatible with the change in loudness level in similar conditions found in experiments where masker and comparison shared the same frequency (e.g., Elmasian and Galambos, 1975). In fact, if the two-process model is valid, it would follow that the loudness matches obtained with the comparison and the masker presented at the same frequency are actually an estimate of loudness enhancement of the proximal target. According to the model, the match reflects loudness enhancement of the target, ILR of the target, and ILR of the comparison. Because the latter two effects are assumed to be identical in size, they should cancel. The necessary assumption would be that the target interpolated between masker and comparison does not influence the loudness recalibration induced in the comparison, however. Additional data are necessary to test this hypothesis.

ACKNOWLEDGMENTS

I am grateful to Armin Kohlrausch for pointing out the importance of measuring the reduction in comparison loudness with exactly the same two tones preceding it as in the three-tone loudness matching task. I thank Andrew Oxenham, Yoav Arieh, and an anonymous reviewer for helpful comments concerning an earlier version of this manuscript.

[1]In some other studies, the alternative terms *conditioner* or *inducer tone* have been used for the masker, and the terms *test tone* or *standard* for the target.

[2]The concept underlying the mergence hypothesis is similar to adaptation-level theory (Michels and Helson, 1954) and the sensation weighting model by Hellström (1977). These models explain the time-order error (Fechner, 1860) by assuming that if two sequentially presented sounds are compared for their loudness, the representations used in the comparison process are influenced by the context, for example, by the average loudness of all preceding stimulation. The representation of target loudness is therefore assumed to be a weighted average between the momentary sensation and "adaptation level." Hellström (1985) noted that loudness enhancement and decrement can be explained if one assumes that the masker influences the adaptation level effective for the target. If so, the remembered loudness of the target would drift towards masker loudness during the target-comparison interval. The explanation rests on the (reasonable) assumption that the adaptation level effective for the target is influenced more strongly by the masker than the adaptation level effective for the comparison, due to the relative temporal proximity between masker and target.

Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time to recovery after induced loudness reduction (L)," J. Acoust. Soc. Am. **117**, 3381–3384.

Arieh, Y., Mailloux, J. R., and Marks, L. E. (**2004**). "Loudness recalibration at short ISI: A closer look," J. Acoust. Soc. Am. **115**, 2600(A).

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

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

Bauer, J. W., Elmasian, R. O., and Galambos, R. (**1975**). "Loudness enhancement in man I. Brainstem-evoked response correlates," J. Acoust. Soc. Am. **57**, 165–171.

Buus, S., Müsch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

Carlyon, R. P., and Beveridge, H. A. (**1993**). "Effects of forward masking on intensity discrimination, frequency discrimination, and the detection of tones in noise," J. Acoust. Soc. Am. **93**, 2886–2895.

Elmasian, R., and Galambos, R. (**1975**). "Loudness enhancement: Monaural, binaural, and dichotic," J. Acoust. Soc. Am. **58**, 229–234.

Elmasian, R., Galambos, R., and Bernheim, A. (**1980**). "Loudness enhancement and decrement in four paradigms," J. Acoust. Soc. Am. **67**, 601–607.

Epstein, M., and Gifford, E. (**2006**). "A potential carry-over effect in the measurement of induced loudness reduction," J. Acoust. Soc. Am. **120**, 305–309.

Fechner, G. T. (**1860**). *Elemente der Psychophysik* (Breitkopf und Härtel, Leipzig).

Galambos, R., Bauer, J., Picton, T., Squires, K., and Squires, N. (**1972**). "Loudness enhancement following contralateral stimulation," J. Acoust. Soc. Am. **52**, 1127–1130.

Harris, D. M., and Dallos, P. (**1979**). "Forward-masking of auditory nerve fiber responses," J. Neurophysiol. **42**, 1083–1107.

Hellström, Å. (**1977**). "Time errors are perceptual. An experimental investigation of duration and a quantitative successive-comparison model," Psychol. Res. **39**, 345–388.

Hellström, Å. (**1985**). "The time-order error and its relatives: Mirrors of cognitive processes in comparing," Psychol. Bull. **97**, 35–61.

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

IEC 318 (**1970**). *An IEC artificial ear, of the wide band type, for the calibration of earphones used in audiometry* (International Electrotechnical Commission, Geneva).

Levitt, H. (**1971**). "Transformed up-down procedures in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.

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

Marks, L. E. (**1996**). "Recalibrating the perception of loudness: Interaural transfer," J. Acoust. Soc. Am. **100**, 473–480.

Michels, W. C., and Helson, H. (**1954**). "A quantitative theory of time-order effects," Am. J. Psychol. **67**, 327–334.

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

Oberfeld, D. (**2003**). "Intensity discrimination and loudness in forward masking: The effect of masker level," in *Fortschritte der Akustik: Plenarvorträge und Fachbeiträge der 29 Jahrestagung für Akustik DAGA '03, Aachen*, edited by Deutsche Gesellschaft für Akustik (Berlin: DEGA), pp. 606–607.

Oberfeld, D. (**2005**). *Alteration of Intensity Resolution and Loudness by Sounds Presented in Temporal Proximity*. Doctoral Dissertation, Technical University Berlin. Web site of the German National Library: http://nbn-resolving.de/urn:nbn:de:kobv:83-opus-10616.

Plack, C. J. (**1996**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024–1030.

Plack, C. J., and Carlyon, R. P. (**1995**). "Loudness perception and intensity coding," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 123–160.

Relkin, E. M., Doucet, J. R., and Sterns, A. (**1995**). "Recovery of the compound action potential following prior stimulation: Evidence for a slow component that reflects recovery of low spontaneous-rate auditory neurons," Hear. Res. **83**, 183–189.

Scharf, B. (**2001**). "Sequential effects in loudness," in *Fechner Day 2001: Proceedings of the 17th Annual Meeting of the International Society for Psychophysics*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Papst, Berlin), pp. 254–259.

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

Schlauch, R. S., Clement, B. R., Ries, D. T., and DiGiovanni, J. J. (**1999**). "Masker laterality and cueing in forward-masked intensity discrimination," J. Acoust. Soc. Am. **105**, 822–828.

Schlauch, R. S., Lanthier, N., and Neve, J. (**1997**). "Forward-masked intensity discrimination: Duration effects and spectral effects," J. Acoust. Soc. Am. **102**, 461–467.

Schlauch, R. S., and Wier, C. C. (**1987**). "A method for relating loudness-matching and intensity-discrimination data," J. Speech Hear. Res. **30**, 13–20.

Suzuki, Y., and Takeshima, H. (**2004**). "Equal-loudness-level contours for pure tones," J. Acoust. Soc. Am. **116**, 918–933.

Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **119**, 1012–1020.

Yost, W. A. (**2000**). *Fundamentals of Hearing. An Introduction* (Academic, San Diego).

Zeng, F.-G. (**1994**). "Loudness growth in forward masking: Relation to intensity discrimination," J. Acoust. Soc. Am. **96**, 2127–2132.

Zeng, F.-G., Turner, C. W., and Relkin, E. M. (**1991**). "Recovery from prior stimulation II: Effects upon intensity discrimination," Hear. Res. **55**, 223–230.

Zwislocki, J. (**1965**). "Analysis of some auditory characteristics," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. B. Bush, and E. Galanter (Wiley, New York), Vol. 3, pp. 3–97.

Zwislocki, J. J., and Sokolich, W. G. (**1974**). "On loudness enhancement of a tone burst by a preceding tone burst," Percept. Psychophys. **16**, 87–90.

The time required to focus on a cued signal frequency[a)]

Bertram Scharf,[b)] Adam Reeves, and John Suciu
Department of Psychology, Northeastern University, Boston, Massachusetts 02115

(Received 9 January 2007; accepted 16 January 2007)

How quickly can a listener focus on a single tonal cue that indicates the frequency of an upcoming signal? Initial measurements were made with frequency uncertainty (signal frequency varies randomly from trial to trial) and with certainty (same frequency on all trials). Measured by a yes–no procedure, thresholds for 40- and 20-ms signals presented in continuous broadband noise at 50 dB SPL were higher in uncertainty than in certainty; the difference decreased monotonically from 5 dB at frequencies below 500 Hz to under 3 dB above about 2500 Hz. This decrease in the detrimental effect from uncertainty, which comes about with increasing signal frequency, may result from preferential attention to higher frequencies. In a second experiment, frequency again varied randomly, but each trial now began with a cue at the signal frequency. The critical variable was the delay from cue onset to signal onset. A delay of 352 ms eliminated the detrimental effect of frequency uncertainty at all frequencies. At the shortest delays of 52 and 82 ms the detrimental effect was reduced primarily at lower frequencies. Our analysis suggests that shifting focus to a cued frequency region, under optimal stimulus conditions, requires less than 52 ms. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2537461]

PACS number(s): 43.66.Dc, 43.66.Mk, 43.66.Ba [GDK] Pages: 2149–2157

I. INTRODUCTION

The question we pose is simple. How quickly can a listener focus auditory attention on a given signal frequency? Put another way, how quickly can a listener go from frequency uncertainty, and presumably some sort of broadband or multiband listening, to frequency certainty and more focused listening? A number of previous papers (e.g., Green, 1961; Gilliom and Mills, 1976; Schlauch and Hafter, 1991; Hübner and Hafter, 1995; Green and McKeown, 2001) have shown that when listeners do not know at what frequency a tonal signal will be presented, the threshold for detection is higher than when they do know. Most researchers have informed listeners about the signal to come by presenting the same frequency throughout a block of trials and/or by presenting a cue at the beginning of each trial well before the signal and at the same frequency as the signal. Information about the signal frequency leads to at least a 3-dB improvement in threshold, equivalent to a 15% increase in performance on a 2IFC task. Although small, the improvement in detection afforded by certainty is robust. Yet, we know little about how long it takes to make use of the frequency information in going from uncertainty to certainty. We know only that a delay of 600 ms or so between cue onset and signal onset is long enough when the information is provided by an ipsilateral cue (Schlauch and Hafter, 1991) and 700 ms is long enough when the information is provided by a contralateral cue (Gilliom and Mills, 1976).

Given that switching from one auditory event, such as a speech token, to another is generally much faster than 600 ms, we expect that focusing attention will prove also to be much faster. Our approach to measuring the speed of auditory focusing is straightforward. Just prior to an observation interval, the listener hears an ipsilateral cue tone at the same frequency as the signal. Until the occurrence of the cue, the listener does not know what frequency to focus on because the frequency of the cue and signal are chosen at random from a large range of possible values. Determining the shortest delay that permits a listener to take advantage of a cue to improve detection begins to provide a measure of the time it takes to focus attention in the frequency domain. Before undertaking these measurements, we needed first to measure detection under the more usual conditions of frequency certainty and uncertainty in order to know just what the threshold limits were for our listeners over a wide range of frequencies. These measurements made apparent that uncertainty reduces detection more at low than at high frequencies, an effect sometimes tentatively noted earlier (e.g., Green, 1961); accordingly, this *frequency effect* became an important part of our study.

To meet the goals outlined above, we carried out two experiments. Experiment 1 determined the effect of frequency uncertainty on detection as a function of signal frequency for 40-ms signals over three frequency ranges and for 20-ms signals over a single range. Experiment 2 determined how much time the listener needed between the onset of an auditory cue and the onset of the signal to eliminate the effect of uncertainty. Features common to both experiments are described in Sec. II.

II. OVERALL METHOD

Listeners sat in a sound-isolated booth. A Tucker-Davis System III signal processor (RP2.1) generated all sounds,

[a)]Portions of this work were presented in "Focusing auditory attention from broadband to single tones: Faster at lower frequencies" (Reeves, Scharf, Suciu, and Jin) at the 46th. Annual Meeting of the Psychonomics Society, Toronto, November 2005.

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

TABLE I. Signal frequencies in the three frequency ranges tested.

Range	Signal frequencies (hertz)								
Low	265	326	391	465	698	1000	1163	1349	1581
Middle	570	700	840	1000	1500	2150	2500	2900	3400
High	1140	1400	1680	2000	3000	4300	5000	5800	6800

sampled at a rate of 48.83 kHz. A microcomputer (Dell Optiplex GX270) controlled the processor and collected data via a response box (TDT BBOX). Sounds were sent through a headphone driver (TDT HB7) to a single Sony MDR-V6 headphone. Wave forms, frequency content, and distortion were checked with a wave-analyzer (GRC 1900) and an oscilloscope (Tektronix TAS220). Background noise was generated digitally to resemble an analogue bi-quad bandpass filter with cutoff frequencies at 300 and 6000 Hz for signals in a middle range of frequencies from 570 to 3400 Hz.[1] The cut-off frequencies of the masking noise were halved to accommodate signals in a low frequency range from 265 to 1581 Hz and were doubled to accommodate signals in a high frequency range from 1140 to 6800 Hz. The noise was set to a spectrum level of 12.44 dB for the middle frequency range, 12.93 dB for the high frequency range, and some 10 dB higher, to 22.45 dB, for the low frequency range to keep the lowest-frequency signals well above absolute threshold.

All signals, including cues, were tone bursts. In most conditions, signal duration was 40 ms, not including the cosine squared rise time of 7 ms and fall time of 5.3 ms; the equivalent rectangular duration equaled 46 ms. In other conditions as noted in the following, the duration was 20 ms, again not including the cosine squared rise time of 7 ms and fall time of 5.3 ms, for an equivalent rectangular duration of 26 ms. Such short durations were used to prevent listeners from shifting attention while the signal was on. However, the attention band is wider for short-duration signals (Wright and Dai, 1994), which would reduce the distinction between broadband listening and focused listening. To help keep the attention band within reasonable limits, we set the continuous background of broadband noise to a low spectrum level between 12 or 13 and 24 dB. At such a relatively low noise level, Botte (1995) has shown that the attention band is narrower than at higher levels.

To provide reference levels for the signals, masked thresholds were measured at 26 frequencies from 265 to 6800 Hz, one frequency at a time, by a 2IFC adaptive procedure. Measurements converged on 79% correct. Table I lists the measured frequencies in three ranges, each of which was tested in a separate session. The nine frequencies in each range were at least one critical band or ERB (Moore and Glasberg, 1987) apart, except that on either side of the center of the range, the nearest frequency was over two critical bands away. (The center was approximately 10% above the geometric mean of the range.) These larger gaps around the center were inserted in anticipation of later experiments now in progress. The mean thresholds across listeners varied from 44.6 dB at 265 Hz to 47.1 dB at 1581 Hz (noise spectrum level: 22.45 dB), from 35.5 dB at 570 Hz to 38.4 dB at 3400 Hz (noise spectrum level: 12.44 dB), and from 36.5 dB at 1140 Hz to 41.0 dB at 6800 Hz (noise spectrum level: 12.93 dB). Based on these means, a smoothed set of signal levels was established for use over all the frequencies in each of the frequency ranges. Signals in all later measurements were set 2 dB above these levels, except in experiment 1 for certainty when they were set 1 dB higher.

For consistency, in both experiments 1 and 2, a single-interval yes–no procedure was used so as to have, in experiment 2, a unique and unequivocal measure of the time between the onset of a tonal cue and the onset of the observation interval. (Control measurements by an adaptive 2IFC procedure were also included in experiment 1 and are indicated where appropriate.) Each trial began always with a visual marker and usually also with a simultaneously presented auditory cue; an observation interval then followed after a predetermined delay. On half the trials a signal was presented during the observation interval, and on the other half no signal was presented. The listener pressed a button to indicate whether or not a signal was heard, upon which the correct answer was indicated. Listeners were encouraged to respond quickly, but no limit was placed on the response interval. Subsequent analyses showed no interactions between response times and any of the relevant stimulus variables. Following the response, the next trial began after 850 ms.

III. EXPERIMENT 1: FREQUENCY CERTAINTY VERSUS UNCERTAINTY

A. Method

Altogether 11 listeners served in these experiments. Seven listeners, four women and three men, participated in the measurements with 40-ms signals in the middle frequency range; five of them plus one new female listener participated in those with 20-ms signals. Six listeners served in the measurements in the low and high frequency ranges; three of those listeners had also served in measurements at signal frequencies from the middle range. Two, including J. S., were members of the laboratory; the others were students at Northeastern University who were paid for their time. Their ages ranged from 19 to 24 years, except for the other lab member who was 34 years old. For all listeners, the ear tested, which was always the left one, was normal as determined by audiometric tests at the NU hearing clinic.

Experiment 1 had two parts, the first with frequency uncertainty, and the second with frequency certainty. In the first part, the signal frequency changed from one signal trial to another; the probability was 0.11 on a given trial that the frequency would be at one of nine frequencies. The frequency was chosen randomly on each trial with no indication

to the listener of what it would be. Each trial began with a visual marker alone. A session, usually lasting approximately 1 h, comprised 12 blocks of 72 trials each (total of 864 trials). In each block, on 36 trials no signal was presented, and on the other 36 trials a signal was presented four times at each of the nine frequencies; accordingly, over the 12 blocks of a session, signals were presented a total of 48 times at each of the nine frequencies. Signal levels were 2 dB above those determined previously to yield 79% correct.

In the second part, for frequency certainty, the same frequency was presented on the 32 signal-present trials within a block of 64 trials. (Fewer trials were run in certainty than in uncertainty since thresholds were more stable for certainty.) Nine blocks were run in a single session to encompass all nine signal frequencies. Within each block, the same frequency was presented on every trial as a cue at a level 7 dB above the 79% threshold. The cue was followed after 410 ms (450 ms from cue onset to signal onset) by an observation interval; a signal occurred at the same frequency as the cue on all 32 randomly chosen signal trials. To avoid ceiling effects, the signal was set 1 dB above the 79% threshold instead of 2 dB as with uncertainty.

Parts 1 and 2 were initially carried out on separate days with signals at the nine middle frequencies listed in Table I. A few weeks after the two sessions had been completed with 40-ms cues and signals, they were replicated with 20-ms cues and signals. (The 20-ms signals came on 430 ms after cue onset.) Still later, both parts 1 and 2 were extended to signals (40-ms duration) in the two other frequency regions above and below the original middle frequency limits. Note that each of the three frequency ranges in Table I covers approximately 2.5 octaves.

Performance was assessed by calculating d'. Under frequency certainty, we obtained $d'=z(H)-z(F)$ for each listener and each frequency, where F is the false alarm rate and H the hit rate. Under frequency uncertainty, a false alarm could not be assigned to any particular signal frequency because a noise-only trial had no frequency signature. Accordingly, we used each listener's overall false alarm rate at all frequencies in a given session.[2] To facilitate comparisons among conditions, from each d' and associated signal level we calculated the level required for d' to equal 1.0. On the basis of data for short-duration signals in noise, plotted in Green and Swets (1988, p. 193), we assumed that d' changes with signal level at the rate of 1.0 unit d' per 3 dB.[3]

B. Results

We first give the results for the middle frequency range, which was the only range studied with both 20- and 40-ms signals in this experiment 1 and the only range studied in experiment 2. Figure 1 gives the results at these frequencies under uncertainty (circles) and certainty (squares) for the 40-ms signals in the upper panel and for the 20-ms signals in the lower panel. The mean SPL calculated to yield a d' of 1.0 is plotted as a function of signal frequency. In order to make clear the overall trends in these data and later in those of Figs. 2 and 3, the means shown are three-point moving averages which attenuate any random variations between adjacent values. However, all tests of statistical significance were based on the original means. Vertical bars are plus and minus one standard error of the mean. Although the size of the error bar did not vary consistently across frequency, for the 20-ms signals the average standard error was markedly larger with uncertainty than with certainty, 0.78 compared to 0.42 dB; on the other hand, for 40-ms signals they were the same, 0.58 and 0.57 dB, in both conditions.

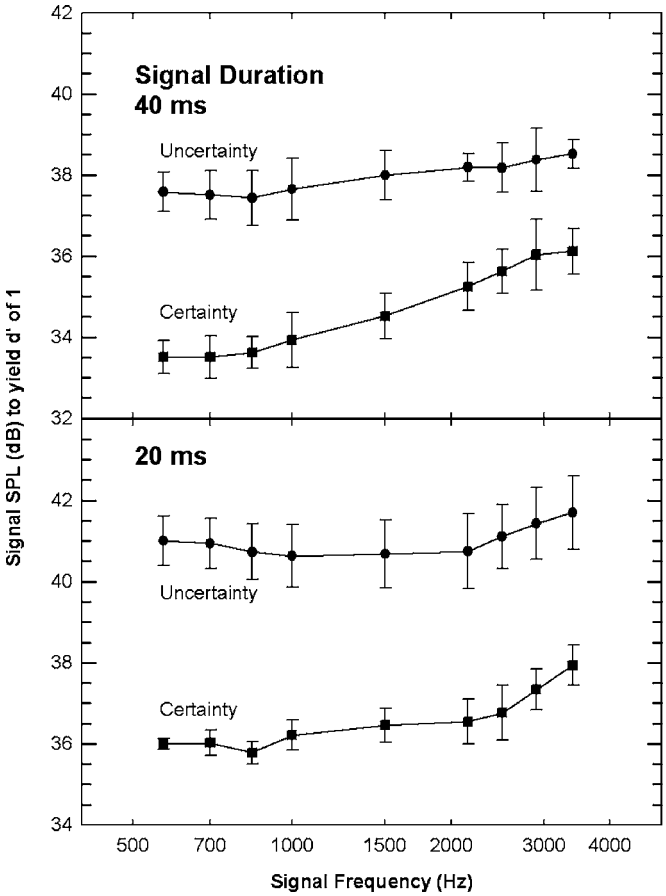


FIG. 1. Mean signal level calculated to yield a d' of 1 is plotted as a function of signal frequency for uncertainty (circles) and certainty (squares). The top panel is for signals with a duration of 40 ms (seven listeners), the bottom for signals with a duration of 20 ms (six listeners). Vertical lines through the data are plus and minus one standard error.

Owing to the shorter duration, thresholds for 20-ms signals are 2–4 dB higher than those for 40-ms signals. At both signal durations, thresholds under uncertainty are significantly higher than under certainty; averaged across all frequencies, the difference is 3.1 dB at 40 ms ($F_{1,6}=142.8$, $p<0.0001$) and 4.5 dB at 20 ms ($F_{1,6}=42.4$, $p=0.001$). Moreover, at both durations the detrimental effect of uncertainty is true for all listeners over all frequencies and for all but three combinations of listener and frequency. For the five listeners who provided thresholds at both signal durations, the effect of uncertainty is greater at 20 than at 40 ms, but not significantly so ($p=0.13$). The threshold for 40-ms signals increases significantly with frequency ($F_{8,48}=6.65$, $p=0.0001$ with certainty and $F_{8,48}=2.24$, $p=0.03$ with uncertainty, by repeated-measures ANOVA.) This increase is smaller than that measured with signals of longer duration, ones usually exceeding 100 ms (e.g., Green *et al.*, 1959; Schlauch and

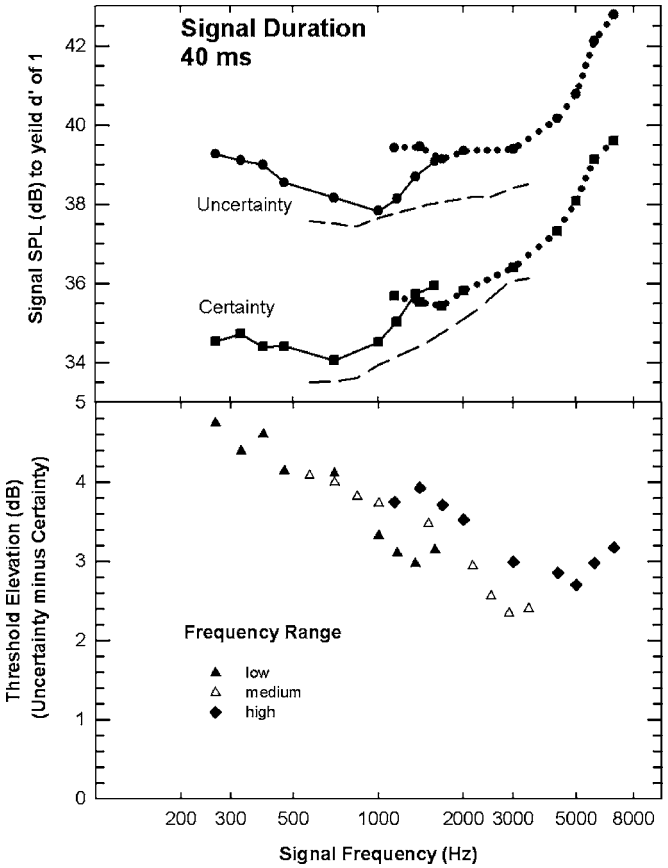


FIG. 2. The top panel plots the mean signal level calculated to yield a d' of 1 as a function of signal frequency. Signal duration was 40 ms. Data for six listeners were collected over a lower range of signal frequencies (solid line) and over a higher range (dotted line) with uncertainty (circles) or with certainty (squares). The dashed lines are data collected over a middle range of signal frequencies (from the upper panel of Fig. 1). The lower panel plots the threshold elevation caused by uncertainty over all three frequency ranges as a function of signal frequency. Each frequency range is indicated by a different symbol as indicated.

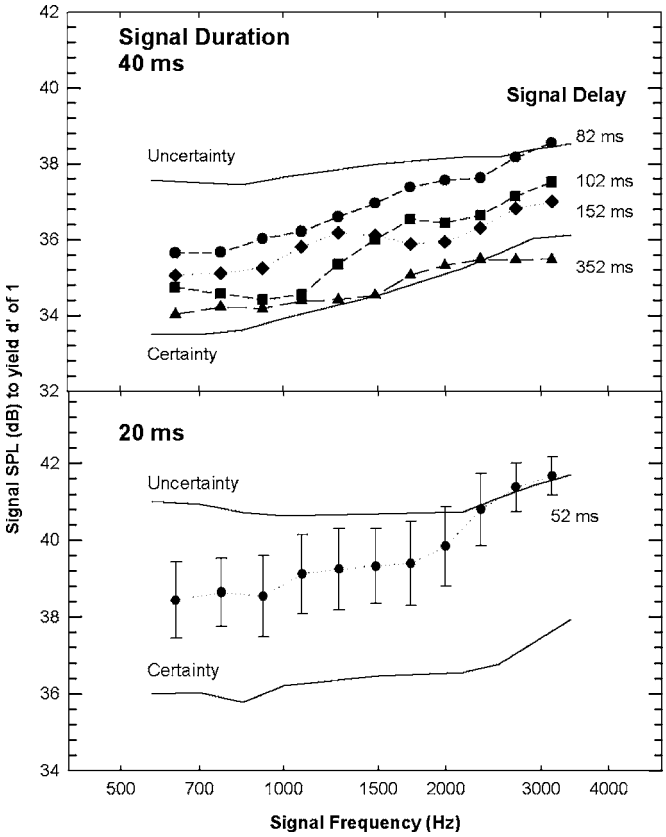


FIG. 3. Mean signal level calculated to yield a d' of 1 is plotted as a function of signal frequency for 40-ms signals in the top panel and for 20-ms signals in the lower panel. The solid curves are taken from Fig. 1. The parameter on the other curves is the delay of signal onset relative to cue onset. Results are from seven listeners.

Hafter, 1991). Although the increase with frequency appears to be smaller with uncertainty than with certainty, this difference was not statistically significant. For 20-ms signals, the threshold increase with frequency is so shallow under both certainty and uncertainty as not to reach statistical significance. A reduced dependence of masked threshold on signal frequency for very brief signals has been documented by Dai and Wright (1996) whose own function for signals of comparably short durations is very similar to our 20-ms function with certainty.

The results in Fig. 1 suggest that the threshold elevation caused by uncertainty is greater at low than at high frequencies. Although this relation between uncertainty and frequency was not statistically significant by ANOVA, it becomes quite clear when we consider the results for these 40-ms signals in the middle frequency range together with those in the lower and higher ranges.

Figure 2 plots, in the upper panel, in the same way as Fig. 1 the signal level required to achieve a d' of 1.0 as a function of signal frequency. Three curves are shown for certainty and three for uncertainty. The low-frequency and high-frequency series are depicted by circles for uncertainty and by squares for certainty; the middle series, from the top panel of Fig. 1, are represented by the dashed lines. (Levels for the low series are plotted 10 dB lower than measured to bring them in line with those for the middle series in which the signals were presented against a noise spectrum level 10 dB higher. Likewise, levels for the high series are plotted 0.49 dB lower.) The three curves overlap nicely under both conditions, except that the thresholds in the middle range (dashed line) were lower than at corresponding frequencies in the other two ranges; this difference probably arises from variability among the listeners as only two of them served in both the middle range and in the other two ranges. Of special note is the decreasing separation between thresholds under certainty and under uncertainty with increasing frequency. Because this divergence is seen within each of the three frequency ranges, it would appear that the greater effect of uncertainty at lower frequencies does not result from their relative position in a series of signals. The interaction between frequency and uncertainty/certainty is seen more clearly in the lower panel of Fig. 2.

The lower panel of Fig. 2 plots the elevation in threshold caused by uncertainty as a function of frequency; the elevation is the threshold with certainty minus the threshold with uncertainty, taken from the upper panel. The three frequency ranges are intermixed but indicated by different symbols. The elevation goes from near 5 dB at the lowest frequency to under 3 dB at the higher frequencies. Overall, the interaction between certainty/uncertainty and frequency was highly sig-

nificant by a repeated-measures ANOVA ($F_{17,85}=3.76$, $p<0.0001$) applied to the data of the six listeners who participated in both the low- and high-frequency series. (Because only two listeners served in the middle and the other two series, we did not do an ANOVA for all three series.) Particularly striking is the nearly monotonic decline in the detrimental effect of uncertainty with increasing signal frequency from 265 to around 2500 Hz; between 2500 and 6800 Hz the detrimental effect appears to remain approximately constant.

C. Discussion

Our results are in line with most of those in the literature as referenced in Sec. I; uncertainty about signal frequency raises threshold 3–5 dB. Although uncertainty appears to be more detrimental for 20-ms signals than for 40-ms signals, Green (1961) reported an opposite result for two listeners who showed a greater uncertainty effect for 1-s than for 10- or 100-ms signals.

Our finding that the uncertainty effect is greater at low than at high frequencies agrees with a similar finding by Green (1961), who found that the uncertainty effect increased by 1 dB as the center frequency of the range of possible signals (all, apparently, 100 ms in duration) decreased from 3200 to 800 Hz. Dai (1994), in his study of the effect of frequency uncertainty on profile analysis, showed larger effects for frequencies below about 1000 Hz than for those above. [Support for the frequency effect, with 100-ms tone bursts, also comes from a report by Gilliom and Mills (1976), albeit in a footnote.] In a related study, of informational masking with 51-ms tone bursts, Richards and Neff (2004) suggested that when uncertain about signal frequency, "observers appear to" pay "somewhat more attention to … higher signal frequencies" (p. 298). Accordingly, it may be that uncertainty is more detrimental at low than at high frequencies because in the absence of any frequency information, listeners focus more at higher frequencies. Despite uncertainty, when a high-frequency signal arrives, processing begins immediately and so even very brief signals can be adequately processed. In contrast, when a low-frequency signal arrives, attention must shift toward that frequency thereby shortening the effective signal duration. In other words, the time spent shifting to the frequency region of the signal is not available for signal processing. This analysis is buttressed by our finding that the frequency effect appears to be somewhat greater for 20-ms signals than for 40-ms signals. Moreover, according to this interpretation, for signals longer than the temporal integration time of approximately 200 ms, the frequency effect should begin to diminish, i.e., uncertainty should begin to raise threshold as much for high as for low frequencies; such is the case as reported by Buus *et al.* (1986, Figs. 3 and 6) for signals 450 ms in duration. It is to be noted that our study as well as all the earlier ones that reported a frequency effect employed signals with durations of 100 ms or less (e.g., Dai, 1994; Gilliom and Mills, 1976; Green, 1961; Richards and Neff, 2004).We return to this point in Sec. IV C.

IV. EXPERIMENT 2: SWITCHING FROM UNCERTAINTY TO CERTAINTY

When a listener has no prior knowledge about signal frequency, how quickly can he or she take advantage of a cue to detect a closely following signal of the same frequency? To ascertain this time, we measured the threshold for a 40- or 20-ms tone burst as a function of its delay relative to a preceding cue.

A. Method

Seven of the listeners from experiment 1 served also in most conditions of this experiment in addition to one new female listener.

All measurements were made against the same 50-dB broadband noise (spectrum level of 12.44 dB), 300–6000 Hz, as in most of experiment 1. The signal, which occurred randomly in half the trials, always had the same duration and frequency as the cue. The frequency was chosen from a logarithmic distribution of frequencies from 570 to 3400 Hz divided into 11 groups each of which was represented in the tests an equal number of times. The step size between adjacent frequencies was equal to a ratio of 1.0053 for a minimum frequency separation, at the lowest frequency, of 3 Hz. The selection was random except that the frequency on a given trial had to be from a different group from that on the previous trial and had to be at least one critical band away. This very large array of frequencies was used because the cues, which were 8 dB above threshold, were readily audible. Hence, had the cues (and signals) been selected from only nine frequencies as in experiment 1, listeners may well have begun to monitor and focus on some of them (cf. Schlauch and Hafter, 1991).

A single session comprised nine blocks of 66 trials each. In each block, a cue from each of the 11 frequency groups was presented six times. On the 33 signal trials, a signal at the same frequency as the cue was presented three times, for a total, over the nine blocks, of 27 trials for each frequency group. In the first four sessions, run on separate days, all the cues and signals were 40 ms in duration. In a given session the delay from cue onset to signal onset or signal onset asynchrony, SOA, was set to 82, 102, 152, or 352 ms. (These delays were measured at the onsets of the flat amplitude portions of the cue and signal.) In a later and separate series of measurements, signals were 20 ms in duration and the onset asynchrony was 52 ms.

B. Results

As in experiment 1, for each listener under each condition we calculated the signal level at which d' would be expected to equal 1.0. Results for all frequencies in a given frequency group were treated together.[4] Unlike the case for the uncertainty condition in experiment 1, the false-alarm rate could now be determined independently for each frequency group, on the basis of the cue frequency. For the 40-ms signals, the false-alarm rate did not vary with signal delay. As a function of frequency, the rate varied from a low of 0.10 to a high of 0.18 with some tendency to increase with frequency up to the middle frequencies and then to decrease

again; overall, the false-alarm rate was 0.14. For the 20-ms signals, only one signal delay was tested so only the effect of frequency could be analyzed. The mean false-alarm rate did not vary in any consistent manner with frequency, ranging from 0.14 to 0.32 for an overall mean of 0.23. It is to be noted that in calculating d' for individual listeners, we replaced 0% false-alarm rates in n trials by $1/(2n)$ and 100% hit rates by $1-1/(2n)$. The effect was to cap d' at 4.1. Such replacements occurred 9% of the time. (No replacements of this kind were required in experiment 1.)

Figure 3 gives the results for the 40-ms signals in the upper panel, from seven listeners all of whom had served in experiment 1, and in the lower panel for the 20-ms signals, from six listeners including five who served in the 40-ms measurements. The mean SPL calculated to yield a d' of 1.0 is plotted as a function of signal frequency with signal delay the parameter on the curves. Also shown, as solid lines, are the uncertainty and certainty curves from Fig. 1.

We describe first the 40-ms results. Overall, the longer the delay of the signal, the lower the threshold. This effect is highly significant ($F_{3,18}=5.76$, $p=0.006$, by repeated-measures ANOVA). With a 352-ms delay, performance was much like that under frequency certainty at all frequencies. (Recall that the results under certainty were collected in similar fashion insofar as a cue was presented, on every trial, 450 ms before signal onset. However, the certainty results were collected with the *same* frequency on every trial in a block.) With shorter delays, threshold rose at all frequencies so that the curves remained roughly parallel to the certainty curve. Given this parallel relation, delay and frequency do not show a significant interaction (by repeated-measures ANOVA, $F_{30,180}=1.24$, $p=0.2$).To avoid confusion, standard errors are not shown. The mean standard error was 0.87 dB across all frequencies and onset asynchronies. It varied inconsistently across frequency from 0.74 to 1.04 dB; collapsed over frequency, it increased from 0.65 dB at a delay of 352 ms to 1.11 dB at a delay of 82 ms. These standard errors are large relative to the effects under study. Indeed, although highly statistically significant for the group, the increase in threshold with decreasing signal delay was not true for two of the seven listeners. Moreover, one listener had higher thresholds with an 82-ms delay than with no cue at all; in later experiments, despite extra training, she continued to do very poorly at the shortest delay.

Compared to uncertainty, performance with the shortest delay of 82 ms was markedly better at the lower frequencies but was no better at the highest frequencies. Since the delay curves are roughly parallel to each other and to the certainty curve, the threshold elevation caused by uncertainty depends on frequency in the same way whether referred to the delay curves or to the certainty curve, as in the lower panel of Fig. 2. In both cases, the amount of threshold elevation caused by uncertainty increases by about 2 dB as frequency decreases from around 2500 to around 600 Hz.

With the signal duration reduced to 20 ms, measurements were made only with a delay of 52 ms. Thus, space was available in the lower panel of Fig. 3 for the plotting of the average standard error at each frequency. Threshold increased with signal frequency in much the same way for these 20-ms signals when presented after a 52-ms delay as for the 40-ms signals presented after an 82-ms delay (upper panel); the standard errors were also similar averaging 0.93 dB at 20 ms and 1.10 dB at 40 ms. At most frequencies, all listeners had higher thresholds with the 52-ms delay than under certainty. These delays of 52 and 82 ms result in better detection—relative to uncertainty—at lower frequencies but not at the highest frequencies. This interaction is significant by repeated-measure ANOVA (for the 52-ms delay, $F_{4,20}=3.34$, $p=0.03$, and for the 82-ms delay, $F_{4,20}=6.54$, $p<0.01$).

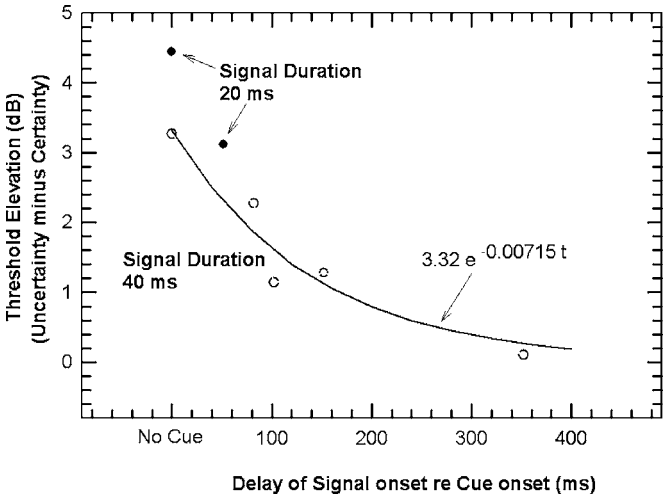


FIG. 4. Averaged across frequency, the threshold elevation or difference between mean threshold with frequency uncertainty and with frequency certainty is plotted as a function of signal delay. Closed circles are for 20-ms signals, open circles are for 40-ms signals. The exponential function fitted to the 40-ms data has a time constant of 140 ms.

The dependence of signal detection on signal delay can be summarized by averaging across all signal frequencies at each delay, since frequency and delay did not interact. Figure 4 plots threshold elevation as a function of signal delay. Here, threshold elevation is the difference between the threshold when signal frequency was initially uncertain on each trial (i.e., no cue was presented with the randomly varying signal or the frequency of the cue and delayed signal varied randomly from trial to trial) and the threshold when frequency was certain (i.e., signal frequency remained the same throughout a block of trials). Under all conditions, thresholds were taken as the mean SPLs calculated to yield a d' of 1.0. Closed circles are the 20-ms data, and open circles are the 40-ms data. The detrimental effect of frequency uncertainty declines from a maximum between 3 and 5 dB when no cue was presented to 0 dB when the cue preceded the signal by over 300 ms. The lack of any threshold elevation at a delay of 352 ms means the effect of uncertainty was eliminated. Accordingly, no further improvement with longer delays is to be expected and an exponential function provides a reasonable fit to the 40-ms data; the one shown has a time constant of 140 ms.

C. Discussion

We assume that in experiment 2, with a new frequency chosen at random on every trial from a wide range of pos-

sible values, listeners attend, prior to trial onset, over a broad range of frequencies.[5] A cue at the beginning of a trial directs attention to the cue's frequency, which was always the same as the signal frequency. The listener is assumed to attempt to focus on a single attention band centered on that frequency. He or she is successful when the cue comes on 352 ms prior to signal onset; the effect of uncertainty is essentially eliminated at all frequencies. Cues that come on 152 and 102 ms earlier than the signal result in a lowering of the threshold to within 1 or 2 dB of the threshold under certainty. Cues that come on 82 ms earlier (or 52 ms earlier with 20-ms signals) put threshold to within 2–3 dB of the threshold under certainty at all frequencies. Accordingly, the functions relating threshold to frequency at the various signal delays are all roughly parallel to the threshold with certainty.

These results show that at all frequencies listeners can shift from broadband or multiband to focused listening in less than 100 ms although focusing may not be complete until the delay approaches 300 ms. Richards and Neff (2004) also varied the delay between cue onset and signal onset, but their study of informational masking differed too much from ours for a detailed comparison. Nonetheless, their findings are in line with ours in that they reported thresholds that were much higher with a delay of 56 ms than with delays of a few hundred milliseconds.

The picture in Fig. 3 is very different relative to uncertainty. The shortest delays result in a lowering of thresholds from uncertainty by as much as 2 dB at the lowest frequencies but not at all at the highest frequencies. Thus, no more than 52 ms is needed to make the detrimental effect of uncertainty the same at all frequencies. This result is fully in accord with the hypothesis we offered in the discussion of experiment 1. In the absence of frequency information, observers, although listening initially in broadband mode, do not attend equally at all frequencies but concentrate more on higher than on lower frequencies, thereby facilitating detection at high frequencies. The difference between the threshold with uncertainty and that with the shortest signal delays provides a measure of the disadvantage for low frequencies of listening "high." The disadvantage, which decreases monotonically with increasing frequency up to around 2500 Hz, is eliminated by a frequency cue which occurs long enough before the signal so that attention can be directed to the appropriate listening band. Since the signal delay can be as short as 52 ms, it would appear that attention shifts at least that quickly to lower frequencies. Processing can then begin at signal onset. But why is threshold after a short signal delay still higher than with certainty? According to our current interpretation, the listener, once cued to the correct frequency region, is no longer uncertain about location. However, enough time and energy must be available in the signal so that the listener can reduce the uncertainty about the precise nature of the signal. This uncertainty is the same at all frequencies; hence the parallel functions in Fig. 3. It may well be this *qualitative* uncertainty is present even when the signal frequency is known ahead of time in so-called certainty conditions.

The question arises as to whether the exclusive cause for poorer detection at shorter delays is the listener's failure to focus quickly enough on the signal frequency. Could a cue, even if only 8 dB above threshold but very close in time to the signal, interfere with signal processing, thereby countering, to some extent, the information it carries? This possibility was checked by measuring the detection of 40-ms signals without a cue, with a cue presented 352 ms before the signal, and with a cue presented 82 ms before the signal; cue and signal frequency were the same and constant at 500, 1500, or 3400 Hz throughout a block of trials. Since frequency was certain, any increase in threshold with the introduction of a temporally close cue would reveal some interference from the cue. For seven listeners, thresholds at all three tested frequencies were 1 dB lower with the 352-ms cue than with either the 82-ms delay or with no cue. Accordingly, the increase in the threshold at shorter delays seen in Fig. 3 may reflect some interference from a proximate cue as well as insufficient time for focusing on the target frequency.[6]

V. GENERAL DISCUSSION

Our measurements of the effect of frequency uncertainty on the detection of tones in noise over a broad range of frequencies, from 265 to 6800 Hz, confirm earlier findings that uncertainty raises threshold approximately 3–5 dB. In addition, they show clearly that the detrimental effect of uncertainty becomes smaller as signal frequency increases, a tendency first noted by Green (1961). The present results suggest further that the detrimental effect of uncertainty is greater for 20-ms signals than for 40-ms signals. Of most importance, these experiments are the first, to our knowledge, to measure the time course of detection as a listener moves from uncertainty to certainty. They show that to overcome fully the detrimental effect of uncertainty, a cue to the signal frequency must come approximately 300 ms before the signal. Delays as short as 52 ms lead to thresholds better than in uncertainty, but the improvement diminishes with increasing signal frequency so that, at around 2500 Hz and higher, delays of 52 and 82 ms yield thresholds no better than under full uncertainty. As noted earlier and in footnote 6, elevated thresholds at short SOAs may also reflect interference from the temporally proximate cue.

The results of these experiments lead us to hypothesize that a cue overcomes two kinds of uncertainty. The initial uncertainty, in the absence of any prior information about signal frequency, is about the spectral locus of the signal and in which critical band to listen. This uncertainty is accompanied by a bias toward listening more for frequencies around and above 2500 Hz than for lower frequencies. It is this initial uncertainty and bias that result in the frequency effect, i.e., the decrease in the detrimental effect of uncertainty with increasing signal frequency. The initial uncertainty is overcome very quickly and probably completely, in less than 52 ms, by a frequency cue so that the frequency effect disappears and detection is impaired by uncertainty no more at low than at high frequencies.

The second uncertainty is always present when attempting to detect a signal in noise (and no doubt also in the quiet) whether information about the signal frequency is available or not. This uncertainty is about the qualitative nature of the

signal. The cues in our experiment 2 cannot eliminate this uncertainty but only reduce it so that threshold, after a long enough signal delay, is the same as with certainty. Figure 4 suggests that approximately 300 ms are needed to reduce, as much as possible, this qualitative uncertainty.

While these results shed new light on how detection in frequency uncertainty varies with signal frequency and duration and on how its improvement depends on the delay of the signal relative to a preceding cue, they add only indirectly to our understanding of just how uncertainty affects detection. In particular, there remains the old question as to why detection is not more adversely affected by uncertainty than it is. Energy-detection models lead to a prediction that threshold in frequency uncertainty would be 10 dB or more above the threshold in certainty. Few attempts [for an exception, based on consideration of the role of internal noise, see Dai (1994)] seem to have been made to come up with a better solution to the apparent discrepancy than the one proposed by Green (1961). Green suggested that even when the same signal is presented on every trial (or, we add, is preceded sufficiently early by a cue), the listener is uncertain about the signal's characteristics. According to Green, the threshold increase of 3 dB he measured in uncertainty does not reflect a change from certainty to uncertainty but from less uncertainty to more uncertainty. The implication seems to be that even with full information about the signal frequency, a listener is still uncertain about what frequency to focus on and listens for more than one frequency over a wider band or in multiple bands. According to our hypothesis, this residual uncertainty is about the qualitative nature of the signal not about its locus. However, because the present experiments did not directly address this question, we leave to an Appendix a somewhat different approach, a decision model that is compatible with an uncertainty effect as small as 3 dB.

Our interpretation of the frequency effect in uncertainty can help explain a corresponding frequency effect in the relation between detection and signal duration. As noted in Sec. III B, Dai and Wright (1996) summarized many data that show that as signal duration is shortened below 80 ms or so, the increase in the masked threshold is greater at low than at high frequencies. Thus, the whole function that relates threshold to signal frequency is flatter at shorter than at longer signal durations. The greater increase in threshold at lower frequencies with decreasing signal duration may mean that even with frequency certainty, listeners focus preferentially on high frequencies. Accordingly, the effective duration of signals at lower frequencies is shortened because additional time is needed after signal onset to first move attention to lower frequencies. This additional time would be of the order of 25 ms, the signal duration at which, according to Dai and Wright (1996), the low-frequency disadvantage becomes most apparent. This estimate is entirely compatible with our estimate that the frequency effect in uncertainty is overcome in less than 52 ms; determining just how much less requires measurements with briefer signals and delays.

As expected, we have shown that listeners can take advantage very rapidly of a frequency cue. At low frequencies, 52 ms after cue onset, threshold for a 20-ms signal at the cued frequency is 2 dB lower than with no cue, i.e., with no prior frequency information. Another 100 to 150 ms are needed to reduce uncertainty (and/or forward interference from the cue) enough so that detection comes close to the same level as under certainty. Figure 4 suggests that this additional time is the same at all frequencies. These results lead us to postulate two kinds of frequency uncertainty. One kind concerns the spectral locus, especially at low frequencies, of the signal. That uncertainty is overcome well within 52 ms by a cue to the signal frequency. A second kind of uncertainty concerns the qualitative nature of the signal. That uncertainty, which is always present, can only be reduced, not eliminated, by a frequency cue, and the maximum reduction requires approximately 150 ms.

ACKNOWLEDGMENTS

This research was supported by AFOSR Grant No. FA9550-04-1-0244 to A.R. and B.S. Zhenlan Jin programmed the experiments, and Katherine Flora helped run subjects. We thank Huanping Dai for comments on an earlier version of the manuscript. We also greatly benefited from comments and suggestions by two anonymous reviewers.

APPENDIX

The principal purpose of this research was to determine how quickly a listener can focus attention on a designated frequency starting from initial uncertainty. We did not seek to study uncertainty per se. Nevertheless, we suggest two different approaches to explaining why the measured uncertainty effect is not as large as the approximately 10 dB predicted by an energy-detection model (Green, 1961).

In our first approach, we follow Green (1961) in assuming that the listener is uncertain about which critical band is likely to contain a signal even when measurements are made with apparent frequency certainty. However, unlike Green, we suggest that the residual uncertainty is not about which critical band to attend but about the precise nature of the change occasioned by a signal *within* the targeted critical band. The listener may easily confuse a change caused by the signal with changes inherent in the noise or in the auditory system. Hence knowing which critical band to focus on does not suffice for signal detection because it is not a change in the overall energy in the targeted band that determines detection but the occurrence of a particular sensory experience. Thus the uncertainty effect may well differ from that predicted by energy detection. This line of reasoning is supported by introspection and, quite tentatively, by supplementary measurements by Green (1961). Green showed that although listeners did not know what frequency to expect, once they detected a tonal signal, they needed only an additional 1 or 2 dB above threshold to report whether the signal was high or low in pitch. This information could be interpreted to show that listeners were detecting signals on the basis of more than overall energy, although it could mean that they are simply able to identify roughly the locus of the critical band containing the signal. In summary, in this first approach we accept Green's hypothesis that listeners are always uncertain in a detection task, but the initial uncertainty, given foreknowledge about the signal frequency, is confined

to the targeted critical band; in other words, listeners are always uncertain near threshold about just *what* to listen for even when they know *where* (spectrally) to listen.

In our second approach, we assume the existence of M independent channels, with one channel containing the signal. In uncertainty, the listener does not know which channel will contain the signal and so attends to all M channels. In certainty, he or she focuses on the channel containing the signal. We identify the channels with distinct combinations of observation intervals and critical bands. Such channels can be treated as statistically independent if each observation interval is at least a critical period from the next, and each band is spaced at least one critical band apart from the nearest band. The probability of detecting the target is the probability that the sample value from the signal distribution exceeds all of the sample values from the $M-1$ identical and independent noise distributions. Smith (1982) showed that to within ±0.1 units, the sensitivity (d') per channel needed to attain an accuracy of p is

$$d'_M \sim [(16+25m)^{0.5}-4]/3 + z(p)[(m+2)/(m+1)]^{0.5}, \tag{A1}$$

where $m=\log_e(M-1)$, and $z(p)$=the z-score of p, with $(0.5=0)$

To illustrate, given that the auditory system comprises 24 critical bands, an accuracy of $p=0.76$ requires that the d' per channel is 2.69 with complete uncertainty (24 channels) and 1.00 with perfect certainty (2 channels). (Note that $M=2$, not 1, in certainty, because a listener who focuses on a single critical band compares two presumably successive samples, one containing the target, and one not; this strategy in uncertainty would yield $M=48$ for a d' per channel of 2.97, which is not much more than 2.69.) The difference of 1.69 d' units corresponds to a benefit of perfect certainty of 5.1 dB, assuming 3 dB per d' unit. In our experiment, listeners knew the signals would not fall outside the noise, which covered 17 critical bands. Accordingly, $d'_{17}=2.54$ in uncertainty, so the benefit of perfect certainty is 4.6 dB [i.e., $3(2.54-1.0)$ dB]. The benefit of partial certainty is less; e.g., focusing on the signal band and two other bands, given that $d'_3=1.48$, is $3(2.54-1.48)=3.2$ dB. These benefits (4.6 and 3.2 dB) are comparable to those seen in Fig. 1 for low and high frequencies, respectively.

[1]The noise wave form was resampled once at the start of each trial and then frozen; otherwise, resampling wide-band noise during the trial grossly distorted the noise wave form when the BBOX was on, owing to a "bug" in the TDT System III.

[2]As a check on the validity of this approach, we repeated the measurements with an adaptive 2IFC procedure. The dependence of detection on signal frequency for uncertainty was essentially the same as measured with the yes/no procedure. We note that Richards and Neff (2004, footnote 4) measured the same thresholds whether they used separate or aggregate false alarms.

[3]Except in the tails (above 95% and below 55%), a slope of 1.0 d'/3.0 dB for unbiased observers corresponds almost exactly to the increase of 5%/dB reported for detection of tones in broadband noise under both frequency certainty and uncertainty (e.g., Green, 1961; Buus *et al.*, 1986). Assuming instead that d' is linear with intensity would make no practical difference, given the restricted range of measured d'.

[4]Values of the criterion $c=-[z(H)+z(F)]/2$ in experiment 2 ranged from −0.19 to 0.58, the typically positive values indicating a slightly greater likelihood to report No than Yes. Averaged across SOA, values of c were near zero (unbiased) for the frequency group centered at 1720 Hz, but increased slightly to 0.48 at the lowest and 0.26 at the highest frequencies ($F_{10,60}=4.64$, $p<0.01$). Averaged across frequency, c decreased from 0.34 at a SOA of 82 ms to 0.11 at a SOA of 352 ms ($F_{3,18}=4.87$, $p=0.012$). The interaction was not significant ($F_{30,180}=1.09$, $p=0.35$). These small variations in criterion were uncorrelated with variations in d' ($r^2=0.012$), and so are unlikely to have biased the calculated values of d'.

[5]We cannot exclude the possibility that listeners were, in fact, in single-band mode, still focused on any given trial on the cue from the previous trial. Evidence for such an influence from preceding trials was offered by Green and McKeown (2001). However, sequential analyses of the present data showed no influence from cues on previous trials.

[6]Experiments carried out since the completion of this paper suggest that the interference may be closer to 2 dB than to 1.

Botte, M. Cl. (**1995**). "Auditory attentional bandwidth: Effect of level and frequency range," J. Acoust. Soc. Am. **98**, 2475–2485.

Buus, S., Schorer, E., Florentine, M., and Zwicker, E. (**1986**). "Decision rules in detection of simple and complex tones," J. Acoust. Soc. Am. **80**, 1646–1657.

Dai, H. (**1994**). "Signal-frequency uncertainty in spectral-shape discrimination: Psychometric functions," J. Acoust. Soc. Am. **96**, 1388–1396.

Dai, H., and Wright, B. A. (**1996**). "The lack of frequency dependence of threshold for short tones in continuous broadband noise," J. Acoust. Soc. Am. **100**, 467–472.

Gilliom, J. D., and Mills, W. M. (**1976**). "Information extraction from contralateral cues in the detection of signals of uncertain frequency," J. Acoust. Soc. Am. **59**, 1428–1433.

Green, D. M. (**1961**). "Detection of auditory sinusoids of uncertain frequency," J. Acoust. Soc. Am. **33**, 897–903.

Green, D. M., McKey, M. J., and Licklider, J. C. R. (**1959**). "Detection of a pulsed sinusoid in noise as a function of frequency," J. Acoust. Soc. Am. **31**, 1446–1452.

Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics*, revised ed. (Peninsulat, Los Altos, CA).

Green, T. J., and McKeown, J. D. (**2001**). "Capture of attention in selective frequency listening," J. Exp. Psychol. Hum. Percept. Perform. **27**, 1197–1210.

Hübner, R., and Hafter, E. R. (**1995**). "Cuing mechanisms in auditory signal detection," Percept. Psychophys. **57**, 197–202.

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.

Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.

Schlauch, R. S., and Hafter, E. R. (**1991**). "Listening bandwidths and frequency uncertainty in pure-tone signal detection," J. Acoust. Soc. Am. **90**, 1332–1339.

Smith, J. E. K. (**1982**). "Simple algorithms for M-alternative forced-choice calculations," Percept. Psychophys. **31**, 95–96.

Wright, B. A., and Dai, H. (**1994**). "Detection of unexpected tones with short and long durations," J. Acoust. Soc. Am. **95**, 931–938.

The measurement problem in level discrimination

Daniel Shepherd[a] and Michael J. Hautus
The University of Auckland, Auckland, New Zealand

(Received 21 March 2006; revised 21 January 2007; accepted 24 January 2007)

There is disagreement among theorists over the exact measure to be used to quantify auditory level discrimination. It has been proposed that, for level discrimination tasks, the measure that is most linearly related to the sensitivity index, d', will be the correct measure. The level difference (ΔL) and the Weber fraction (Θ) are both candidates, though the latter is sensitive to the physical unit in which it is expressed (e.g., pressure or intensity) while the former is not. Psychometric functions for level discrimination were obtained at a number of pedestal levels for 10-ms sinusoids (either 1000 or 6500 Hz) and broadband noise bursts. These functions were used to assess which of three measures: ΔL, $\Theta=\Delta p/p$, or $\Theta=\Delta I/I$, is most nearly linearly related to d'. The results suggest that $\Delta p/p$ is the measure that comes closest to being linearly related to d'. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2697628]

PACS number(s): 43.66.Fe, 43.66.Dc, 43.66.Ba, 43.66.Yw, 43.66.Cb [JHG] Pages: 2158–2167

I. INTRODUCTION

Auditory level discrimination refers to the ability of an observer to distinguish between two acoustic waves differing in amplitude. Assume two stimuli are presented to an observer who is to judge which has the greater level. The first is a standard stimulus of magnitude A (the pedestal) and the second a comparison stimulus of magnitude A plus an increment ΔA. Researchers seek to find the smallest value of ΔA that allows the observer to reliably identify the stimulus that contains the increment. This minimum difference is termed the "just noticeable difference" (jnd), and is synomonous with the difference limen (DL). Among theorists, however, there exists no consensus on how the jnd should be measured, and we have termed this difficulty the *measurement problem in level discrimination*.

The measurement problem centers on how the auditory system's discriminatory capabilities should be modeled. More specifically, in an experiment probing how well an observer can discriminate auditory stimuli differing in amplitude, how should a researcher define the dependent variable? For example, should the nature of the measurement be absolute, that is $(A+\Delta A)-A$, or relative, that is, $\Delta A/A$? One absolute measure, called the *level difference* by Buus (1990), and denoted ΔL, is commonly calculated as

$$\Delta L = 20 \log_{10}\left(\frac{p+\Delta p}{p}\right), \tag{1}$$

where p indicates pressures are being used. This measure simply reflects the difference, in decibels, between the pedestal and increment, $p+\Delta p$, and the pedestal alone, p. In units of intensity (I) the level difference is sometimes known as "ΔI in dB" (Grantham and Yost, 1982). Note that the value of ΔL when pressures are used is equal to the value of ΔL when intensities have been selected (Green, 1993). A relative measure is the ubiquitous Weber fraction:

$$\frac{\Delta X}{X} = \Theta, \tag{2}$$

where X can be in units of pressure, or intensity. The Weber fraction, Θ, differs between individuals and sensory dimensions. This measure simply reflects the proportional increase in magnitude, ΔX, needed for a change in level to be detected for a given pedestal value, X.

The measurement problem in audition exists because the candidate jnd metrics, ΔL, $\Delta I/I$, and $\Delta p/p$, are proportional to one another within the typical range of human discriminatory performance. Grantham and Yost (1982) demonstrate this proportionality and Green (1988, 1993) offers approximations between ΔL, $\Delta I/I$, and $\Delta p/p$, and the latter two measures presented in decibels: $10\log(\Delta I/I)$ and $20\log(\Delta p/p)$. If the measures are simply transformations of one another then which is correct, and pertinently, is the choice of measure of consequence? Most Weber fractions are small, with $\Delta I/I$ typically between 0.21 and 0.73 (Green, 1993). Therefore it is difficult to distinguish between Eqs. (1) and (2), because $\ln(1+\varepsilon)\approx\varepsilon$, for small ε (Raney *et al.*, 1989). To circumvent this difficulty, experiments must be designed to deliberately inflate jnd values to a region where a nonlinear relationship exists between the jnd measures.

A common psychophysical construct applied in the measurement of discriminatory performance is the psychometric function, which plots the magnitude of an increment normalized to pedestal level (i.e., $\Delta X/X$) as a function of some performance criterion, for example, proportion correct or the sensitivity index d'. The value of ΔX that yields a jnd satisfying some predetermined performance criterion (e.g., 75% correct) and the corresponding value of X are substituted into Eq. (2) to yield a Weber fraction. In relation to the psychometric function, the measurement problem in level discrimination manifests itself as to which of ΔL, $\Delta I/I$, or $\Delta p/p$ should the performance measure be a function of. A solution to the measurement problem in level discrimination is vitally important on theoretical grounds (Doble *et al.*, 2003) because different models of auditory level discrimination pre-

[a]Electronic mail: daniel.shepherd@aut.ac.nz

 0001-4966/2007/121(4)/2158/10/$23.00

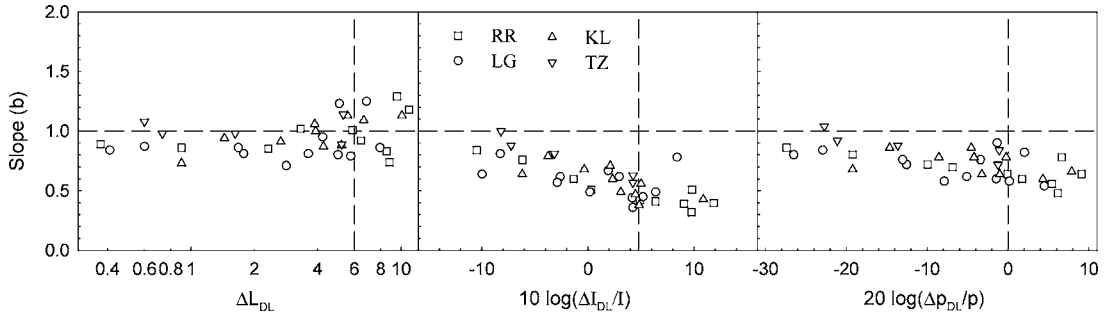


FIG. 1. Slope estimate, b, plotted as a function of the jnd expressed in terms of either ΔL_{DL} (left panel); $10\log(\Delta I_{\mathrm{DL}}/I)$ (center panel), or $20\log(\Delta p_{\mathrm{DL}}/p)$ (right panel). The dashed horizontal line is $b=1$; vertical lines demarcate proportionality (right of line) or not (left of line) between the jnds. Data, for four observers, from Buus and Florentine (1991), Table I (p. 1374).

dict different measures. Furthermore, Buus and Florentine (1991) argue that measures of the jnd based on Eq. (2) distort the relationship between stimulus magnitude and sensitivity, and hence misrepresent the sensitivity of the auditory system to changes in stimulus parameters.

Previous attempts to solve the measurement problem (Buus and Florentine, 1989; Raney *et al.*, 1989; Moore *et al.*, 1999) have focused on a popular model of the psychometric function (Egan *et al.*, 1969):

$$d' = aX^b, \tag{3}$$

where d' is the detection-theory index of discriminability (Green and Swets, 1966), a is a scalar that accounts for individual differences, and X can be ΔL, $\Delta I/I$, or $\Delta p/p$. The exponent, b, determines the slope of the psychometric function, though on occasion it is thought of as describing the *shape* of the function (Moore *et al.*, 1999). It has been proposed that the measure of X that exhibits linearity with d' is the correct metric in which to judge auditory level resolution (Buus and Florentine, 1991; Moore *et al.*, 1999). However, for stimuli with small difference limens there already exists a proportionality between the three candidates for X. For stimuli that do not afford such high sensitivity, however, the proportionality between the measures diminishes, and they can be pitted against one another. The region of proportionality is approximately below $\Delta I/I=3(\Delta L=6.02;\Delta p/p=1)$ and so emphasis should be placed on stimuli producing difference limens beyond these values when judging if X is linearly related to d'.

Three previous studies have attempted to determine which metric is linearly related to d' (Buus and Florentine, 1989; Raney *et al.*, 1989; Moore *et al.*, 1999). When Eq. (3) is plotted on log coordinates a value of b close to unity indicates linearity, which is Weber's law. Figures 1 and 2, presenting data from Buus and Florentine (1991) and Moore *et al.* (1999), respectively, plot values of b as a function of jnd: ΔL_{DL}, $10\log(\Delta I_{\mathrm{DL}}/I)$, or $20\log(\Delta p_{\mathrm{DL}}/p)$. The dashed horizontal lines are $b=1$, while the dashed vertical lines discern the zone of proportionality (to the left of the line) or nonproportionality (to the right of the line).

Inspection of Figs. 1 and 2 reveals conflicting conclusions, with Buus and Florentine (Fig. 1) concluding that d' is linearly related to ΔL, while the data of Moore *et al.* (Fig. 2) suggest that d' is linearly related to $\Delta I/I$. The stimulus configurations utilized by the two studies were, however, different, and may account for the lack of agreement between the studies (Laming, 1986). Buus and Florentine (1991) employed a difference discrimination task, while Moore *et al.* (1999) used stimuli more consistent with an increment detection task. Difference discrimination involves discriminating between two stimuli (X and $X+\Delta X$) separated either in space or time. Increment detection involves a continuous uniform stimulus (X) above the level of background noise and the addition of an increment (ΔX). The possibility that different metrics underlie these two tasks is of fundamental importance to those modeling the auditory system.

The findings of Raney *et al.* (1989), using complex profile stimuli, were inconclusive, and their data failed to indicate which of ΔL or $\Delta p/p$ was linearly related to d'. It is clear that further evidence is required to resolve this problem. Consequently, we conducted three difference discrimination experiments in which the stimuli were specifically chosen to yield relatively large jnds.

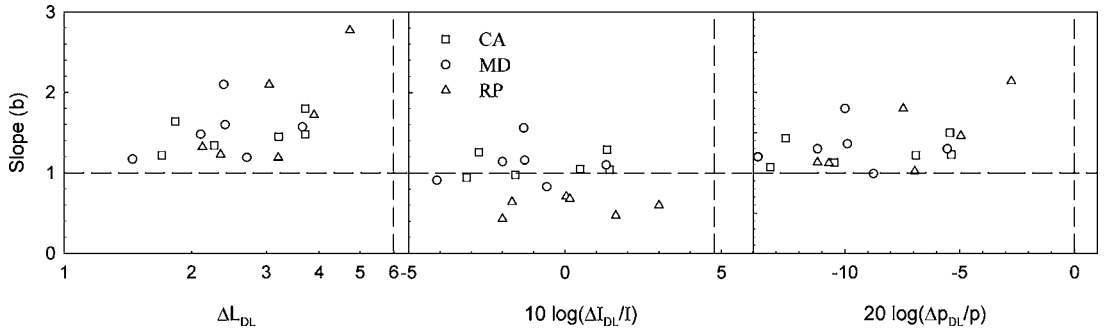


FIG. 2. As for Fig. 1, data, for three observers, from Moore *et al.* (1999).

II. EXPERIMENT 1: 1000 Hz SINSUOIDS IN GATED NOISE

A. Introduction

Among the stimulus configurations employed by Buus and Florentine (1991) were 10-ms 1000-Hz sinusoids. Experiment 1 adopts the same sinusoidal dimensions but additionally includes a broadband-noise background, the purpose of which was to produce larger difference limens through direct masking of the stimuli. The noise also serves to mask spectral splatter associated with short-duration stimuli, a known facilitator of the detection process.

B. Method

1. Observers

Four males: WC (aged 24), MK (aged 29), EL (aged 33), and DS (aged 30) served as observers. All had extensive experience in psychoacoustic tasks and had normal audiometric thresholds in the ear to be tested. All but the author, DS, received a monetary incentive to participate in the experiment.

2. Stimuli

Stimuli were 1000-Hz sinusoids temporally centered in gated broadband noise. The sinusoid had a duration of 10 ms with 1-ms rise and fall times ($\cos^2$) and was generated at a sampling rate of 44.1 kHz. Ten pedestal levels were employed, ranging from 15 to 60 dB sound pressure level (SPL) in 5-dB SPL steps. The waveform of the pedestal, A, differed from the pedestal-plus-increment, $A+\Delta A$, only in the amount of attenuation it was subjected to. The broadband noise (N_0 =35 dB SPL) was 200 ms in duration with 10-ms rise and fall times ($\cos^2$).

3. Apparatus

The gated noise and 1000-Hz sinusoid were generated independently using National Instruments LABVIEW 6.1, and then converted from a digital to an analog representation (NI PCI-6052E). The noise was directed to a pair of static attenuators (TDT, PA4) whose level of attenuation remained constant across the experimental block. The sinusoids were directed through two programmable attenuators (TDT, PA5), set up in series, and then added to the noise in a signal mixer (TDT, SM5). Once combined, the noise and sinusoid were delivered to a headphone buffer (TDT, HB7) and from there to an earphone (TDH 49P, No. 30195). All stimuli were presented monaurally to the observer's left ear.

In addition to generating the stimuli, the LABVIEW software also controlled the programmable attenuators, presented instructions to the observer on a terminal positioned within the sound-attenuating chamber (Amplaid Model E) housing the observer, and, through an auxiliary keyboard, recorded the observer's responses, and controlled feedback lights.

4. Procedure

A two-alternative forced-choice (2-AFC) adaptive three-down, one-up staircase procedure (Levitt, 1971) was used to measure difference limens for each of the ten pedestal levels. This provided the observers practice as well as facilitating in the selection of increments to be used in a subsequent difference discrimination task. Each difference limen was based upon three blocks of trials, each consisting of 15 reversals. The first three reversals changed the pedestal-plus-increment level by ±3 dB SPL, while for subsequent reversals this change was ±1 dB SPL. Any single block returning a standard deviation greater than 2 dB was discarded and repeated. The adaptive procedure initially estimated the difference limen (DL) expressed in terms of the level difference: $\Delta L_{\mathrm{DL}}=20\log[(p+\Delta p_{\mathrm{DL}})/p]$, where Δp_{DL} is the sound pressure increment (Buus and Florentine, 1991). To calculate difference limens in pressure, Δp_{DL}, the mean of the last 12 reversals for each block were averaged and then converted to the difference limen by solving Eq. (1) for Δp.

Once difference limens had been estimated from the adaptive data, a variation of the method of constant stimuli (Moore *et al.*, 1999) was used to collect psychometric functions for ten pedestal levels. Observers were presented two intervals per trial with one of those intervals containing a pedestal, and the other a pedestal plus an increment. The interval containing the increment was determined randomly with an equal *a priori* probability (i.e., $p=0.5$). The observer was instructed to indicate on a keypad located in the experimental chamber the interval that contained the increment. Trial-by-trial feedback was provided contingent on response.

Empirical psychometric functions were collected with each function based on five increment levels that ranged from −10 to +10 dB SPL in 5-dB SPL steps with reference to the observer's difference limen (i.e., Δp_{DL}). Each psychometric function was based on five blocks of trials, with each of the five points based on 105 trials. The first ten trials of any block were designated practice trials and were omitted from the final analyses.

5. Data analysis

Empirical psychometric functions were constructed for each subject by plotting log d' vs log ΔL, $10\log(\Delta I/I)$, and $20\log(\Delta p/p)$ for each of the ten pedestal levels, yielding 30 functions in all. Values of d' were derived from percentage correct scores using an approximation developed by Hacker and Ratcliff (1979). Equation (3) is a basic power function, and linearity between d' and X occurs when the exponent, b, equals one. It is customary to represent Eq. (3) on double-logarithmic coordinates, where the family of power functions is transformed to a family of straight lines. The exponent, b, is the slope of the line, and the scalar, a, is the intercept. Consequently, b is commonly referred to as the *slope* parameter. The parameter, a, is generally not of interest; suffice it to say that it reflects differences in sensitivity between observers.

The psychometric functions were fitted with Eq. (3) using the method of least squares. This provided two parameter estimates for each function: a and b. Next, the value of the jnd was estimated by substituting the estimates of a and b into Eq. (3) and setting d' equal to unity. The value $d'=1$ is conventionally regarded as performance at threshold and is equal to 76% correct in a 2-AFC task (Green and Swets,

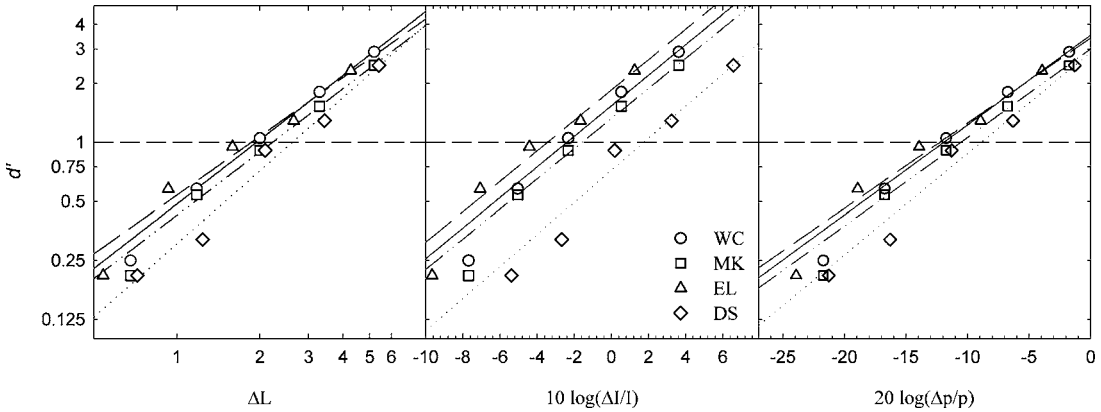


FIG. 3. Psychometric functions plotting the sensitivity index d' as a function of either ΔL (left), $10\log(\Delta I_{DL}/I)$ (center), or $20\log(\Delta p_{DL}/p)$ (right) for four observers. The pedestal was a 1-kHz, 10-ms pedestal of 30 dB SPL. The dashed horizontal lines represent performance at threshold (i.e., $d'=1$). The dashed diagonal lines are the best-fitting forms of Eq. (3) for each observer.

1966). Solving Eq. (3) for X provided an estimate of the jnd, against which the slope parameter, b, was subsequently plotted.

C. Results

Figure 3 shows psychometric functions for a single pedestal level: 30 dB SPL. The best-fitting lines, from Eq. (3), sufficiently accounted for data across all three of the experiments, with the goodness-of-fit statistic, R^2, being greater than 0.9 for each psychometric function. Table I lists mean parameter estimates and goodness-of-fit statistics for each observer obtained from fitting Eq. (3) to the data obtained at each pedestal level. The high values of R^2 show that the data were well accounted for by the equation. A one-sample t-test on the data from each subject showed that all estimates of b were significantly different from one ($p<0.05$) with the exception of two cases: observers MK $[t(9)=0.852, p=0.416]$ and DS $[t(9)=0.69, p=0.508]$ did not have estimates of b different from unity when $X=\Delta p/p$.

Figure 4 plots b as a function of jnd, with each plot consisting of 40 points (four observers by ten pedestal levels). The slope, b, is different across the three jnd measures $[F(2,119)=93.23, p<0.001]$ with mean values, across observers and pedestals, being 1.22 (s.d.=0.2) for ΔL, 0.75 (s.d.=0.14) for $\Delta I/I$, and 0.97 (s.d.=0.12) for $\Delta p/p$. The slope estimates for $\Delta L[t(39)=7.26, p<0.001]$ and $\Delta I/I[t(39)=-11.24, p<0.001]$ were significantly different from unity, but the slope parameter for $\Delta p/p[t(39)=-1.56, p=0.126]$ was not. These results indicate that the best measure of level discrimination for brief 1000-Hz sinusoids is the Weber fraction expressed in units of pressure.

III. EXPERIMENT 2: BROADBAND NOISE WITH 3-AFC

A. Introduction

Buus and Florentine (1991) utilized noise that covered the audible frequency range ($f_c=22$ kHz), had a noise power density of 20 dB SPL, and was 500 ms in duration. A direct comparison between the limens they obtained with noise to

TABLE I. Estimates of best-fitting parameters and of the corresponding fit statistic, R^2, for equations (a) $d'=a\Delta L^b$; (b) $d'=a(\Delta I/I)^b$, and; (c) $d'=a(\Delta p/p)^b$ in Experiment 1. Asterisks signify the means differ significantly from unity ($p<0.05$).

	a		b		R^2	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
(a) $d'=a\Delta L^b$						
WC	0.311	0.151	1.164*	0.173	0.988	0.009
EL	0.356	0.135	1.096*	0.104	0.982	0.012
MK	0.276	0.148	1.285*	0.249	0.981	0.016
DS	0.164	0.084	1.350*	0.135	0.984	0.001
(b) $d'=a(\Delta I/I)^b$						
WC	1.109	0.666	0.706*	0.099	0.990	0.005
EL	1.257	0.524	0.712*	0.122	0.987	0.008
MK	1.111	0.627	0.811*	0.165	0.982	0.009
DS	0.782	0.345	0.778*	0.153	0.988	0.006
(c) $d'=a(\Delta p/p)^b$						
WC	2.312	0.786	0.923*	0.071	0.990	0.006
EL	2.536	0.876	0.894*	0.074	0.980	0.018
MK	2.418	0.764	1.039	0.146	0.982	0.010
DS	1.854	0.805	1.024	0.110	0.988	0.007

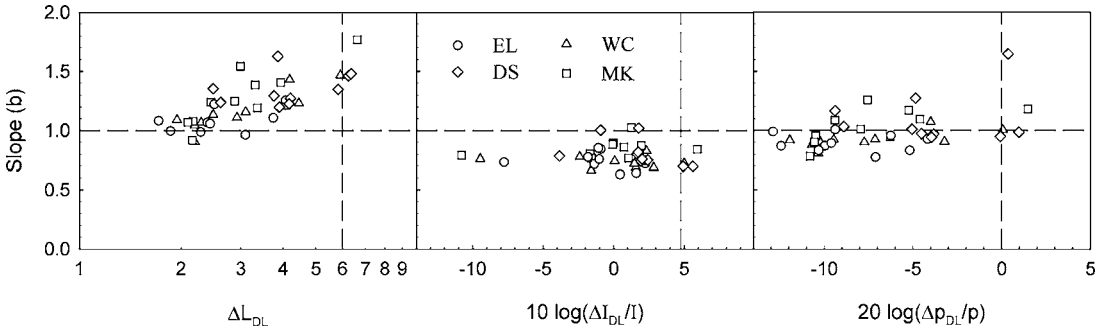


FIG. 4. Slope estimate, b, plotted as a function of the jnd expressed in terms of either ΔL_{DL} (left panel); $10\log(\Delta I_{\mathrm{DL}}/I)$ (center panel); or $20\log(\Delta p_{\mathrm{DL}}/p)$ (right panel). The dashed horizontal line is $b=1$; vertical lines demarcate proportionality (right of line) or not (left of line) between the jnds. Stimuli were 10-ms bursts of 1000-Hz sinusoids presented in noise. Data for four observers.

those they derived using 500-ms 1000-Hz sinusoids showed the noise thresholds were higher than the sinusoid thresholds. We hypothesize that such a difference will also hold between noise and sinusoids of 10-ms duration. Moore *et al.* (1999) implied that difference limens could also be increased by making the observer's task more difficult. The addition of an extra observation interval should increase the uncertainty inherent in the observer's decision, which has long correlated with task difficulty (Green and Swets, 1966). Confusion over when a signal occurs forces an observer to attend to an increased number of noise (i.e., pedestal alone) channels in an effort to detect the increment (Nachmias and Kocher, 1970). As the number of noise channels needing to be monitored increases, the psychometric function becomes shallower, increasing the difference limen. For example, uncertainty about stimulus duration (Dai and Wright, 1998) and temporal occurrence (Watson and Nichols, 1975) serve to introduce nonlinearities into the psychometric function. Consequently, in addition to adopting 10-ms bursts of Gaussian noise as stimuli, we increased the difficulty of the task by employing a 3-AFC procedure.

B. Method

1. Observers

Three observers undertook Experiment 2; two had participated in Experiment 1 (EL and DS), while one had not (MF, a 37 year old female). MF had extensive experience in auditory psychophysical tasks. MF had normal audiometric thresholds (re: ISO Standard, 1975) at all frequencies tested, bar 6000 Hz. MF's audiogram exhibited no thresholds greater than 20 dB HL. Only EL received a financial incentive to participate.

2. Stimuli

Stimuli were 10-ms broadband noises low-pass filtered at 8000 Hz. Filtering was undertaken with a fourth-order Butterworth filter. The ten noise pedestals had spectrum levels of −15, −10, −5, 0, 5, 10, 15, 20, 25, and 30 dB SPL. The noise had rise and fall times ($\cos^2$) of 1 ms. Masking noise was absent throughout.

3. Apparatus

The apparatus was identical to that employed in Experiment 1. Noise pedestals and increments were generated independently using National Instruments LABVIEW 6.1. The pedestals and increments were directed to the pair of static attenuators (TDT, PA4) and the pair of programmable attenuators (TDT, PA5), respectively, and they were combined at the signal mixer (TDT, SM5).

4. Procedure

The procedure was identical to that employed in Experiment 1 except that there were, on any one trial, three, as opposed to two, observation intervals. The adaptive three-down, one-up 3-AFC procedure located the difference limen that corresponds to 79% correct for each of the ten pedestals (Levitt, 1971). The variant of the Method of Constant Stimuli employed also had three observation intervals. Five increment levels were employed, defined with respect to the difference limen estimates obtained with the adaptive procedure. They ranged from −10 to +10 dB re: Δp_{DL}, in 5-dB steps.

C. Results

Estimates of a and b, averaged across the ten pedestal levels for each observer, are presented in Table II for each jnd measure. The goodness-of-fit statistics, R^2, again indicate that Eq. (3) provides an acceptable fit to the data ($R^2 > 0.97$). A one-sample t-test on the data from each observer showed that, for ΔL and $\Delta I/I$, all mean estimates of b were significantly different from one ($p<0.001$). For $\Delta p/p$, results for two observers [MF($t(9)=-0.153, p=0.882$]; EL[$t(9)=-0.209, p=0.839$)] were not significantly different from one, whereas that for DS was [$t(9)=4.967, p<0.001$].

Inspection of Fig. 5 suggests that the slope parameter, b, depends on the jnd measure. Analysis-of-variance confirms that this is the case [$F(2,89)=253.8, p<0.001$], with a *post hoc* test (Bonferonni) indicating that all three means were significantly different from each other ($p<0.001$). These mean estimates, obtained by averaging across both observer and pedestal level, are $b=1.694$ (s.d.$=0.163$) for ΔL, $b=0.769$ (s.d.$=0.0689$) for $\Delta I/I$, and $b=1.040$ (s.d.$=0.1$) for $\Delta p/p$. These means were significantly different from unity for $\Delta L[t(29)=15.82, p<0.001]$ and $\Delta I/I[t(9)=$

TABLE II. Estimates of best-fitting parameters and of the corresponding fit statistic, R^2, for the equations (a) $d'=a\Delta L^b$; (b) $d'=a(\Delta I/I)^b$, and; (c) $d'=a(\Delta p/p)^b$ in Experiment 2. Asterisks signify the means differ significantly from unity ($p<0.05$).

	a		b		R^2	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
(a) $d'=a\Delta L^b$						
MF	0.085	0.037	1.498*	0.164	0.983	0.011
EL	0.113	0.051	1.420*	0.184	0.988	0.006
DS	0.126	0.043	1.491*	0.142	0.993	0.006
(b) $d'=a(\Delta I/I)^b$						
MF	0.507	0.099	0.702*	0.084	0.973	0.019
EL	0.546	0.104	0.763*	0.069	0.983	0.009
DS	0.619	0.141	0.842*	0.050	0.985	0.010
(c) $d'=a(\Delta p/p)^b$						
MF	1.135	0.278	0.995	0.099	0.979	0.015
EL	1.294	0.212	1.019	0.104	0.987	0.006
DS	1.688	0.381	1.104*	0.066	0.989	0.008

$-14.27, p<0.001$], but not significantly different for $\Delta p/p$[$t(29)=1.82, p=0.079$]. Thus, of the three candidate measures: ΔL, $\Delta I/I$, or $\Delta p/p$, the evidence again favors $\Delta p/p$.

There is evidence that employing noise and increasing task complexity increased ΔL's, possibly through the added uncertainty in the decision-making process. The difference in ΔL's, averaged across pedestal levels and observers, between Experiment 1 ($\bar{x}=3.38$, s.d. $=1.33$, $n=40$) and Experiment 2 ($\bar{x}=4.94$, s.d. $=1.11$, $n=30$) are significantly different ($\alpha=0.05$, one-tailed) from those estimated in Experiment II [$t(68)=-5.21$, $p<0.001$]. For the two observers that participated in both experiments, DS and EL, the difference between Experiment 1 ($\bar{x}=3.58$, s.d. $=1.37$) and Experiment 2 ($\bar{x}=4.59$, s.d. $=0.8$) is not as pronounced but is still significantly different ($t(38)=-2.84, p=0.004$).

IV. EXPERIMENT 3: 6500 Hz SINUSOIDS IN BANDSTOP NOISE

A. Introduction

Buus and Florentine (1991) demonstrated that for sinusoids between 1 and 14 kHz the difference limen increases with frequency. Thus, larger jnds are obtained for high frequency sinusoids. Additionally, the severe departure to Weber's law manifests itself as inflated difference limens for midlevel ($\approx$35–55 dB SPL) high-frequency sinusoids (>5000 Hz).

Carlyon and Moore (1984) reported that the addition of bandstop noise boosted the difference limen for a sinusoidal signal centered in the spectral notch, expressed in units of intensity, by approximately $\Delta I_{DL}=5$ dB. Because this increase in the difference limen was evident only for midlevel pedestals, it appears the addition of bandstop noise serves to enhance the severe departure to Weber's law.

These findings suggest that relatively large jnds can be attained for the discrimination of high-frequency sinusoids in bandstop noise. Our choice of a 6500-Hz sinusoid was, in part, determined by the response characteristics of the headphones, which effectively acted as a low-pass filter with a cut-off of 8000 Hz. Additionally, Carlyon and Moore (1984, 1986) utilized sinusoids at this frequency, and both studies utilized identical bandstop masking noise.

B. Method

1. Observers

There were two participants, WC and DS, both of whom had also participated in Experiment 1. WC received financial compensation.

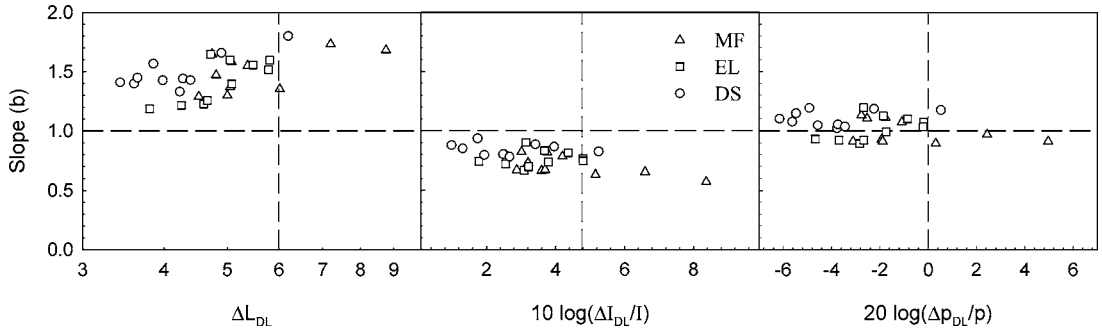


FIG. 5. As for Fig. 4. Stimuli were 10-ms broadband noise bursts. Data for three observers.

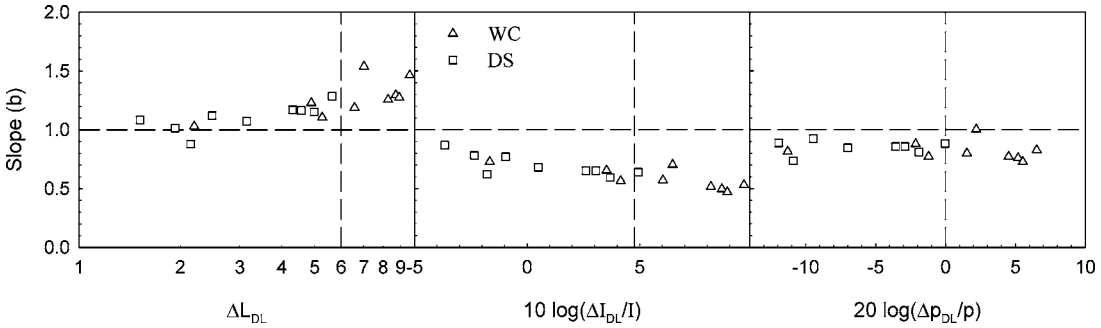


FIG. 6. As for Fig. 4. Stimuli were 10-ms bursts of 6.5-kHz sinusoids presented in bandstop noise. Data for two observers.

2. Stimuli

Stimuli were 6500-Hz sinusoids embedded in a bandstop noise background. Two bands of noise ($W=1950$ Hz), one centered on 4875 Hz and the other on 8125 Hz were produced with fourth-order Butterworth filters, and then added. The notch width was therefore 1300 Hz, and the 6500-Hz sinusoid fell in the middle of the notch. The extremities of the higher frequency band of noise extended beyond the frequency response of the observer's ear piece ($\approx$8000 Hz). The nine pedestals had levels of: 20–60 dB SPL in 5-dB steps. The five increment levels ranged from −10 to +10 dB SPL in 5-dB steps, with reference to the observer's difference limen, Δp_{DL}. Both the noise and the sinusoids were of 10-ms duration, including 1-ms onsets and offsets ($\cos^2$).

3. Apparatus and procedure

The apparatus and procedures used were identical to those employed in Experiment 1.

C. Results

Figure 6 plots the exponent b as a function of jnd for the two observers. Table III provides the mean parameter estimates for each observer and each measure. The goodness-of-fit is again very good, with Eq. (3) sufficiently accounting for the data ($R^2>0.97$). Two one-sample t-tests performed on each subject's data showed that values of b were significantly different from unity: ($p<0.001$) regardless of the measure used to represent the jnd. Examination of the data for these two observers shows that, for the 36 estimates of b associated with $\Delta I/I$ and $\Delta p/p$, only one was greater than unity. For the 18 estimates of b associated with ΔL, only one was less than unity. This latter finding reflects the results obtained for ΔL with 1000-Hz sinusoids (see Experiment 1) and noise (Experiment 2).

The mean estimates of b, obtained by averaging across both observer and pedestal level, were $\Delta L=1.25$ (s.d. $=0.22$), $\Delta I/I=0.655$ (s.d. $=0.09$), and $\Delta p/p=0.881$ (s.d. $=0.71$), and are significantly different [$F(2,55)=58.16, p<0.001$]. Bonferonni *post hoc* analyses indicate that all three means are different from each other ($p<0.001$). Additionally, these mean slope estimates are all significantly different from unity [$\Delta L(t(17)=5.101, p<0.001$]; $\Delta I/I[t(17)=-17.063, p<0.001$]; $\Delta p/p[t(17)=-7.53, p<0.001)$]. From this analysis it must be concluded that none of the three candidate measures, ΔL, $\Delta I/I$, and $\Delta p/p$, obtained strict linearity with d'. All measures produced slope estimates significantly different from unity, with the measures based on the Weber fraction ($\Delta I/I$ and $\Delta p/p$) yielding slopes less than unity, while ΔL produced slopes greater than unity. Of the three measures, $\Delta p/p$ is the closest to unity.

V. DISCUSSION

In order to ascertain which of ΔL, $\Delta I/I$, or $\Delta p/p$ achieves a linear relationship with d', a series of three experiments were undertaken using 1000-Hz sinusoids (Experiment 1), broadband noise (Experiment 2), or 6500-Hz sinusoids in bandstop noise (Experiment 3). All stimuli were presented monaurally for a duration of 10 ms. Figure 7 illus-

TABLE III. Estimates of best-fitting parameters and of the corresponding fit statistic, R^2, for the equations (a) $d'=a\Delta L^b$, (b) $d'=a(\Delta I/I)^b$, and (c) $d'=a(\Delta p/p)^b$ in Experiment 3. Asterisks signify the means differ significantly from unity ($p<0.05$).

	a		b		R^2	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
(a) $d'=a\Delta L^b$						
WC	0.114	0.133	1.367*	0.213	0.994	0.006
DS	0.329	0.187	1.091*	0.119	0.990	0.007
(b) $d'=a(\Delta I/I)^b$						
WC	0.498	0.331	0.616*	0.068	0.981	0.011
DS	1.041	0.508	0.687*	0.104	0.973	0.016
(c) $d'=a(\Delta p/p)^b$						
WC	1.0102	0.750	0.878*	0.065	0.989	0.009
DS	2.329	1.306	0.881*	0.090	0.986	0.008

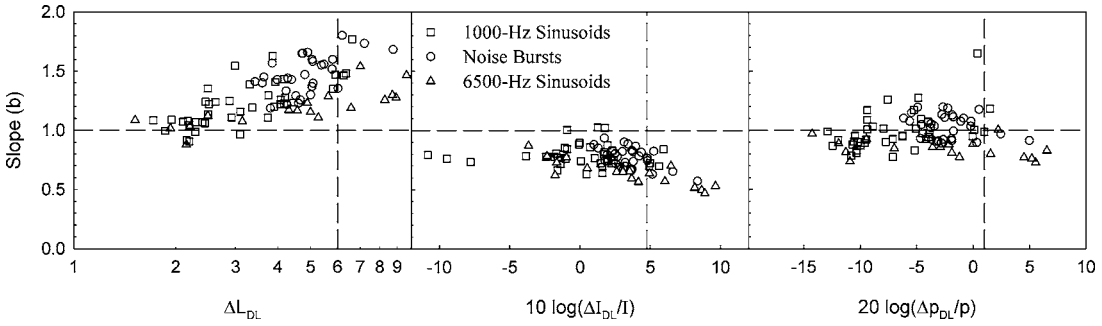


FIG. 7. Slope estimate, b, plotted as a function of the jnd expressed in terms of either ΔL_{DL} (left panel), $10\log(\Delta I_{\mathrm{DL}}/I)$ (center panel), or $20\log(\Delta p_{\mathrm{DL}}/p)$ (right panel). The symbols represent three different configurations of stimuli.

trates the slope estimates, b, obtained in all three experiments, for each of the jnd measures. Data points clustering around the dashed horizontal lines indicate linearity between d' and the jnd. From these data there is evidence that, when the jnd is represented as the Weber fraction, $\Delta X/X$, d' is linear to pressure ($\Delta p/p$). If $\Delta I/I$ is selected as the jnd, then b appears to be consistently lower than unity. When the measure of the jnd is taken to be ΔL, the slope exponent b progressively increases as ΔL increases.

Further support for the claim that d' is linearly related to $\Delta p/p$ comes from analysis of the entire data set. Averaging all estimates of b across the three experiments gives the following means: $\Delta L=1.312$ (s.d. $=0.22$), $\Delta I/I=0.736$ (s.d. $=0.12$), and $\Delta p/p=0.977$ (s.d. $=0.12$). Only the mean for $\Delta p/p$ is not significantly different from unity [$\Delta p/p(t(89)=-1.873, p=0.064$]; $\Delta L[t(89)=13.457, p<0.001]$; $\Delta I/I[t(89)=-20.587, p<0.001)]$. This conclusion is consistent with Laming's (1986) sensory analytical model that predicts $\Delta p/p$ to be the correct measure of auditory level discrimination.

The relationship among ΔL, $\Delta I/I$, and $\Delta p/p$ was discussed in Sec. I. It was stressed that for $\Delta I/I<3$, a proportionality exists among ΔL, $\Delta I/I$, and $\Delta p/p$. Beyond this region, however, the proportionality no longer holds, and it is argued (e.g., Buus and Florentine, 1991; Moore *et al.*, 1999) that stimuli falling into this region (i.e., $\Delta I/I>3$) should be given heavier weighting when judging the measure obtaining linearity with d'. The dashed vertical lines in Fig. 7 represent the value of the jnd where $\Delta I/I$ equals three. This occurs at $\Delta L=6.02$, $10\log(\Delta I/I)=4.77$, and $20\log(\Delta p/p)=0$. Adopting this criterion and examining the data to the right of the vertical lines in Fig. 7, it is again apparent that of the three measures $\Delta p/p$ is the candidate that obtains the most convincing linearly relationship with d'.

In contrast Buus and Florentine (1991), on the basis of their data (see Fig. 1), concluded that d' is linearly related to ΔL. They did, however, report departures from linearity, notably for stimuli possessing large jnds, where b consistently exceeded unity. They explained this inflation of b in terms of bias inherent both in data analytical procedures and attention lapses on the part of the observer. An increase in b with higher values of ΔL_{DL} has been found consistently during the course of the current investigation. Given the value placed upon stimuli producing large jnd measures in the elucidation of the correct measure to employ in level discrimination, Buus and Florentine's conclusion is in need of further empirical support. The data and subsequent interpretations presented by Moore *et al.* (1999) are also not reflected in the present study. They found that when ΔL and $\Delta p/p$ were selected as the jnd all estimates of b were above unity (see Fig. 2). In contrast the estimated exponents for $\Delta I/I$ fell on either side of unity. However, a comment on stimulus context is warranted. The data reported by Buus and Florentine (1991) and the current study were obtained from a traditional difference discrimination task, while that of Moore *et al.* (1999) are reported to have come from an increment detection task. Thus the evidence suggests that it is unlikely that a single jnd metric will be able to account for data obtained using both forms of stimulus configuration. This fundamental difference in the way the auditory system resolves level with respect to stimulus context has also been found with psychometric functions (Green and Sewall, 1962; Laming, 1986).

Estimates of b as a function of pedestal level for each of the three experiments are displayed in Fig. 8. Inspection of these figures reveals that the metric associated with the greatest amount of variability in b is ΔL (Fig. 8, left columns), though the data did not permit meaningful significance testing to be undertaken. However, it does appear that, for jnds expressed in terms of $\Delta p/p$ (Fig. 8, right columns) and $\Delta I/I$ (Fig. 8, center columns), b is relatively stable across pedestal levels.

VI. CONCLUSION

Three experiments employing 10-ms 1000-Hz sinusoids, broadband noise, and 6500-Hz sinusoids indicate that d' is most linearly related to $\Delta p/p$. These results differ fundamentally to those described by Buus and Florentine (1991) and Moore *et al.* (1999). That these differences exist among the three independent studies is of interest, and it is clear that further investigation is called for. One promising direction is suggested by Ward and Davidson (1993), who showed that large Weber fractions can be obtained from pedestals of low frequency and level.

ACKNOWLEDGMENTS

The authors would like to thank B. C. J. Moore for making data available for further analysis. This research was supported by grants from the University of Auckland Department of Psychology Research Expenses Fund and

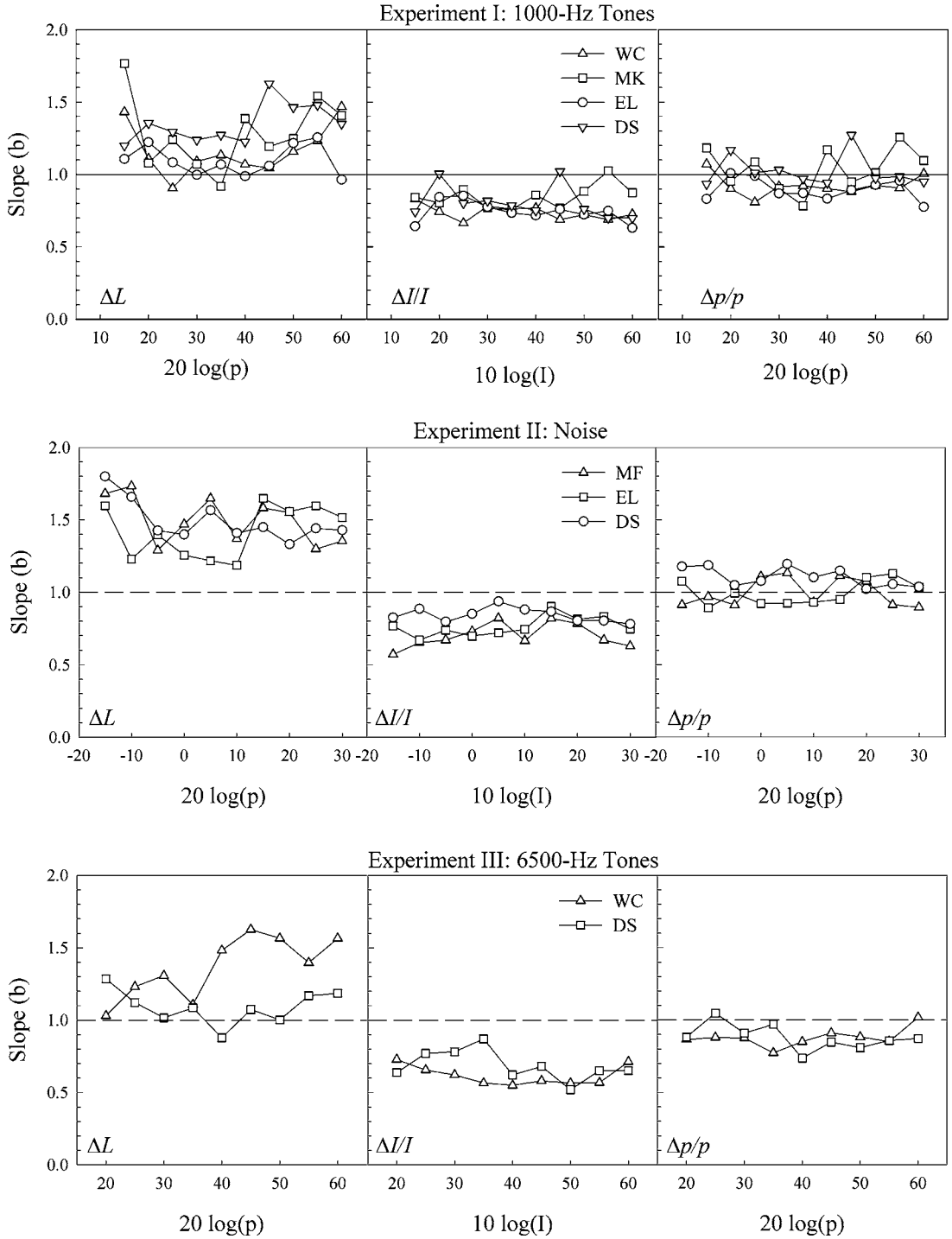


FIG. 8. The slope parameter, b, as a function of pedestal level for jnds expressed as ΔL (left), $10\log(\Delta I_{DL}/I)$ (center), and $20\log(\Delta p_{DL}/p)$ (right). The top (Experiment 1), middle (Experiment 2), and bottom (Experiment 3) panels represent the three stimulus configurations used, while the legends indicate the observers.

Conference Travel Fund, and also from the University of Auckland Research Committee. The authors would like to dedicate this paper to the memory of Soren Buus, who kindly provided valuable comments on experimental design and data analysis during its formative stage.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.

Carlyon, R. P., and Moore, B. C. J. (**1984**). "Intensity discrimination: A severe departure from Weber's law," J. Acoust. Soc. Am. **76**, 1369–1376.

Carlyon, R. P., and Moore, B. C. J. (**1986**). "Detection of tones in noise and the 'severe departure' from Weber's law," J. Acoust. Soc. Am. **79**, 461–464.

Dai, H., and Wright, B. A. (**1998**). "Predicting the detectability of tones with unexpected durations," J. Acoust. Soc. Am. **105**, 2043–2046.

Doble, C. W., Falmagne, J., and Berg, B. G. (**2003**). "Recasting Weber's law," Psychol. Rev. **110**, 365–375.

Egan, J. P., Linder, W. A., and McFadden, D. (**1969**). "Masking level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

Grantham, D. W., and Yost, W. A. (**1982**). "Measures of intensity discrimi-

nation," J. Acoust. Soc. Am. **72**, 406–410.
Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, New York).
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).
Green, D. M., and Sewall, S. T. (**1962**). "Effects of background noise on auditory detection of noise bursts," J. Acoust. Soc. Am. **34**, 1207–1216.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.
Laming, D. (**1986**). *Sensory Analysis* (Academic London).
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
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.
Nachmias, J., and Kocher, E. C. (**1970**). "Visual detection and discrimination of luminance increments," J. Opt. Soc. Am. **60**, 382–389.
Raney, J. J., Richards, M., Onsan, Z. A., and Green, D. M. (**1989**). "Signal uncertainity and psychometric functions in profile analysis," J. Acoust. Soc. Am. **86**, 954–960.
Watson, C. S., and Nichols, T. L. (**1975**). "Detectability of auditory signals presented without defined observation intervals," J. Acoust. Soc. Am. **59**, 655–668.
ISO Standard (1975).
Ward, L. M., and Davidson, K. P., (**1993**), "Where the action is: Weber fractions as a function of sound pressure at low frequencies," J. Acoust. Soc. Am. **94**, 2587-2594

Comparison of level discrimination, increment detection, and comodulation masking release in the audio- and envelope-frequency domains

Paul C. Nelson
Department of Biomedical and Chemical Engineering and Institute for Sensory Research, Syracuse University, Syracuse, New York 13244

Stephan D. Ewert
Centre for Applied Hearing Research, Technical University of Denmark, Lyngby, Denmark

Laurel H. Carney
Department of Biomedical and Chemical Engineering and Institute for Sensory Research, Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, New York 13244

Torsten Dau
Centre for Applied Hearing Research, Technical University of Denmark, Lyngby, Denmark

(Received 3 May 2006; revised 20 December 2006; accepted 10 January 2007)

In general, the temporal structure of stimuli must be considered to account for certain observations made in detection and masking experiments in the audio-frequency domain. Two such phenomena are (1) a heightened sensitivity to amplitude increments with a temporal fringe compared to gated level discrimination performance and (2) lower tone-in-noise detection thresholds using a modulated masker compared to those using an unmodulated masker. In the current study, translations of these two experiments were carried out to test the hypothesis that analogous cues might be used in the envelope-frequency domain. Pure-tone carrier amplitude-modulation (AM) depth-discrimination thresholds were found to be similar using both traditional gated stimuli and using a temporally modulated fringe for a fixed standard depth (m_s=0.25) and a range of AM frequencies (4–64 Hz). In a second experiment, masked sinusoidal AM detection thresholds were compared in conditions with and without slow and regular fluctuations imposed on the instantaneous masker AM depth. Release from masking was obtained only for very slow masker fluctuations (less than 2 Hz). A physiologically motivated model that effectively acts as a first-order envelope change detector accounted for several, but not all, of the key aspects of the data. [DOI: 10.1121/1.2535868]

PACS number(s): 43.66.Mk, 43.66.Dc [JHG] Pages: 2168–2181

I. INTRODUCTION

When discussing temporal cues in sounds, it is necessary to differentiate features that change on different time scales. On a relatively short time scale, a sound's pressure waveform fluctuates about zero; these variations are referred to as fine structure and are determined by the instantaneous audio frequency of the stimulus. In contrast, a signal's temporal envelope changes on a longer time scale and is always positive; the envelope is a description of the slow variations in overall level that define the instantaneous amplitude. It is useful to describe complex temporal amplitude modulations in the envelope-frequency domain. Virtually all natural sounds have complex audio-frequency domain spectra and complex envelope-frequency domain spectra.

A variety of fundamental experimental paradigms originally used in audio-frequency psychoacoustics have recently been translated into their envelope-frequency equivalents. In the process, certain parallels have emerged between the effective signal processing that is inferred to take place in the two domains. Masked tone-detection experiments that compare the amount of masking with different spectral relationships between the signal and masker indicate perceptual frequency tuning in both audio frequency (e.g., Wegel and Lane, 1924) and envelope frequency (e.g., Houtgast, 1989). Also, the "asymmetry of masking" has been observed in both domains [e.g., Moore *et al.*, 1998; Derleth and Dau, 2000 (audio frequency); Ewert *et al.*, 2002 (envelope frequency)], with tones proving to be relatively ineffective maskers of noise signals compared to the masking effect on tonal signals produced by a noise masker. Wojtczak and Viemeister (2005) showed nonsimultaneous (forward) masking in the envelope-frequency domain that has direct counterparts in audio-frequency psychophysics (e.g., Lüscher and Zwislocki, 1949). The ability to resolve tone level is also broadly similar across domains, with an approximately 1–2-dB increase in the standard level (SPL in audio frequency, $20 \log m$ in envelope frequency) required to reliably discriminate the levels of two supra-threshold sounds [e.g., Florentine *et al.*, 1987 (audio); Ewert and Dau, 2004 (envelope)]. These qualitative similarities suggest that the processes fundamental to perception in the two domains may be conceptualized in a single framework, despite the fact that the underlying mechanisms may be quite different.

 0001-4966/2007/121(4)/2168/14/$23.00

In this study, the hypothesis that two other robust audio-frequency phenomena would be observed in the envelope-frequency domain was tested. These phenomena are (1) a heightened sensitivity to increments with a continuous carrier (or a temporal fringe) relative to gated-carrier level discrimination performance and (2) lower thresholds in a tone-in-noise detection task with a temporally amplitude-modulated (AM) masker than in conditions with a random (unmodulated) masker. The second observation has been termed comodulation masking release (CMR) in the audio-frequency domain because it is most robust when several frequency channels are simultaneously and coherently modulated. Both audio-frequency observations can be at least partially attributed to AM-related cues (e.g., Gallun and Hafter, 2006; Schooneveldt and Moore, 1989). Therefore, upon transposition into the envelope-frequency domain, the analogous cues in such tasks would be related to the second-order envelope (Lorenzi *et al.*, 2001), or "venelope" (Ewert *et al.*, 2002).

The current understanding of venelope perception can be summarized as follows: Sinusoidal modulation of the depth of a first-order AM carrier is detectable under some conditions (e.g., Lorenzi *et al.*, 2001), but the perceptual salience of venelope components is generally found to be weaker than that of first-order envelope fluctuations (Ewert *et al.*, 2002). Venelope fluctuations can interact with envelope detection and vice versa (e.g., Moore *et al.*, 1999; Ewert *et al.*, 2002); this finding can be qualitatively accounted for with several physiologically realistic nonlinearities that effectively transform second-order envelope components into first-order envelope components (Shofner *et al.*, 1996). Alternatively, venelope cues could take the form of temporal variations in the output of envelope-frequency-tuned modulation filters (Ewert *et al.*, 2002; Füllgrabe *et al.*, 2005; Füllgrabe and Lorenzi, 2005), which could interfere at some later stage of processing with the output of the modulation filter tuned to the signal frequency.

In the audio-frequency domain, listeners are more sensitive to level differences presented as continuous-carrier level increments than presented as gated tones with different SPLs (e.g., Campbell and Lasky, 1967; Viemeister and Bacon, 1988; Bacon and Viemeister, 1994). An energy-based detection model cannot explain the difference in thresholds in the two conditions. Instead, the temporal structure of the standard, target, and interstimulus intervals must be taken into account. This finding can be considered from several perspectives. One possibility is that the memory requirements of the system are higher in gated-carrier level discrimination than for increment detection, where listeners could potentially perform the task without comparing across intervals even in a two-alternative forced-choice task (Harris, 1963). This explanation is less than satisfactory because, near threshold, there is certainly an element of comparison across intervals even in the continuous-carrier task: the listener must decide which interval sounded the most like it contained an increment or "bump." Also, the relatively short intervals between stimuli probably render memory noise negligible with respect to sensation noise in most two-or three-interval paradigms (Durlach and Braida, 1969).

Another related explanation holds that the improved sensitivity results because the system could be making decisions by detecting changes in the increment task (Macmillan, 1971; Hafter *et al.*, 1997). Onsets and offsets of gated stimuli result in excitation of the putative modulation filterbank (e.g., Dau *et al.*, 1997) in both standard and target intervals that depends on the shape and duration of the ramps applied to the carrier (Gallun and Hafter, 2006). Increment detection paradigms using a continuous pedestal, on the other hand, cause a change in the signal envelope only in the target interval. As a result, transient onset (and offset) responses are present in only one of the observation intervals. Also, gated paradigms require the listener to identify the interval containing the more intense sound, while continuous-carrier level discrimination can be performed without knowing the direction of the change in level (Hafter *et al.*, 1997).

Physiologically, absolute firing-rate changes in single auditory-nerve fiber (ANF) responses to increases in SPL do not depend on the temporal position of the increment with respect to the onset of the pedestal (e.g., Smith and Zwislocki, 1975; Smith and Brachman, 1982; Smith *et al.*, 1985). Instead, the *relative* increase in instantaneous rate increases as the delay between pedestal onset and increment is lengthened, because the response to the pedestal decreases with time. If it is assumed that a constant fractional increase in the response is required to reach threshold, these findings can also qualitatively account for the perceptual gated-continuous difference.

Fringe effects in level discrimination have provided a long-standing challenge for pure power-spectrum models of audio-frequency processing; tone-in-noise detection tasks that compare masking by modulated and unmodulated maskers have emerged more recently as challenges to such long-term energy-based models (e.g., Schooneveldt and Moore, 1989; Verhey *et al.*, 1999). This article focuses on a simple class of CMR experiments: those that use a single noise masker centered on the signal frequency. This class of CMR paradigms yields the most significant and robust release from masking when the masker is broadband and fully modulated (Verhey *et al.*, 2003). Several cues could potentially underlie a release from masking (i.e., lower thresholds with a modulated masker compared to unmodulated masker conditions). A "dip-listening" model suggests that the listeners are able to selectively attend to the periods of the stimulus with low masker amplitudes, thus improving the local signal-to-noise ratio (SNR; Buus, 1985). Another possible cue, which is mainly based on the processing in the peripheral channel tuned to the signal frequency (within-channel processing), is the overall smoothing of the masker fluctuations upon addition of the signal (Schooneveldt and Moore, 1989; Verhey *et al.*, 1999). Across-peripheral-channel comparisons of target-interval differences might also be used if the bandwidth of the masker is sufficiently broad (e.g., Hall *et al.*, 1984). All of these mechanisms have been used to understand CMR in the audio-frequency domain.

Based on these empirical audio-frequency observations, the current study presents envelope-frequency-domain versions of the experiments that led to them. Listeners' access to venelope cues should determine whether differential effects

will be observed in (1) continuous- and gated-carrier AM depth discrimination and (2) sinusoidal AM (SAM) detection in the presence of a noise AM masker with and without slow and regular fluctuations in overall modulation depth. The remainder of this article is divided into three main sections. Two lines of psychophysical experiments are described and discussed in the first two sections. The third part focuses on interpretation of the findings with the help of a physiologically motivated computational model.

II. EXPERIMENT I. LEVEL DISCRIMINATION AND INCREMENT DETECTION IN THE ENVELOPE-FREQUENCY DOMAIN

The goal of the first set of experiments was to determine whether continuous-carrier AM-depth-discrimination thresholds were lower than traditional gated-carrier thresholds. Two reasonable hypotheses lead to predictions of a difference in performance between the two paradigms. First, adaptation at the output of modulation-tuned channels could have the effect of masking across-interval depth differences, since neural responses often exhibit more variability at higher response amplitudes (e.g., Young and Barta, 1986; see Sec. IV of the current study). This would result in poorer performance in the gated conditions. Alternatively, because energy increment detection in the audio-frequency domain is at least partly associated with modulation detection (Wojtczak and Viemeister, 1999) and coded along the modulation dimension (Gallun and Hafter, 2006), a corresponding task in the envelope domain may provide another cue along an additional dimension, the hypothetical venelope dimension, which might lead to lower thresholds in the continuous-carrier condition relative to the gated case.

A. Methods

1. Listeners

The experiments were carried out at the Centre for Applied Hearing Research at the Technical University of Denmark (DTU). All of the listeners participated voluntarily and had pure-tone detection thresholds less than 20 dB HL at octave frequencies between 125 and 8000 Hz. Their ages ranged from 23 to 39 years. Three of the four subjects in the main experiments had significant experience in related psychoacoustic testing; two of the authors (PCN, TD) were part of this group. Four additional listeners were recruited to participate in the extension of experiment I (Sec. II B 2) because of the relatively large across-subject variability.

2. Apparatus and stimuli

Subjects listened diotically via Sennheiser HD 580 circumaural headphones in a double-walled, sound-attenuating booth. Stimulus generation and presentation were carried out in MATLAB using the AFC software package developed at the University of Oldenburg and at DTU. A 48-kHz sampling rate was used to digitally generate stimuli. The carrier was a 70-dB SPL, 5.5-kHz pure tone. Sinusoidal AM was applied for the entire 500-ms duration, and a 50-ms raised-cosine window was applied at the onset and offset of observation-interval stimuli. Inter-observation-interval durations (between possible target interval presentations) were also 500 ms in duration.

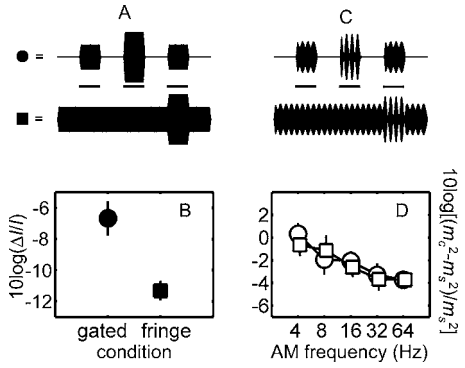


FIG. 1. Comparison of the gated-continuous difference in the audio (left) and envelope (right) frequency domains. Schematic illustrations of stimulus waveforms in the two experiments are shown in (a) and (c) (horizontal bars between the stimuli indicate the timing of the 500-ms observation intervals). (b) Audio-frequency level discrimination thresholds measured with gated (closed circles) and quasi-continuous pedestals (closed squares). (d) AM depth discrimination thresholds for a modulated standard (m_s=0.25) obtained with traditional gated intervals (open circles), and with quasi-continuous modulation presented before, between, and after the observation intervals (open squares). For both (b) and (d): f_c=5500 Hz; standard SPL =70 dB. Each symbol is the average threshold for four listeners; error bars indicate ±1 standard deviation of the individual mean thresholds.

For comparison to the envelope-domain results, thresholds for the audio-frequency versions of pure-tone gated level discrimination and quasi-continuous increment detection (i.e., with a temporal fringe) were also measured in the same subjects. Signal duration, inter-observation-interval, gating parameters, SPL, and carrier frequency were the same as in the envelope-frequency experiments. The corresponding example stimulus waveforms are shown in Fig. 1(a).

The modulating waveforms for the AM depth discrimination paradigm in the gated conditions were identical to those described in Ewert and Dau (2004). The observation-interval stimuli are described by the following equation:

$$s(t) = \sin(2\pi f_c t)[1 + m_s\sqrt{1 + m_{\text{inc}}}\sin(2\pi f_m t)],$$

where f_c is the carrier frequency (5500 Hz), m_s is the standard modulation depth, m_{inc} is the relative depth increment (zero in the standard intervals), and f_m is the modulation frequency. The comparison (target interval) depth can be related to the standard depth and depth increment simply as $m_c = m_s\sqrt{1+m_{\text{inc}}}$. Using a notation more in line with audio-frequency level discrimination literature, m_{inc} can also be thought of as the Weber fraction, i.e., $m_{\text{inc}} = (m_c^2 - m_s^2)/m_s^2$. Whereas the earlier study (Ewert and Dau, 2004) focused on the effects of standard-interval modulation depth for a fixed-frequency (16-Hz) sinusoidal AM, the current experiments used a fixed standard depth (m_s) of −12 dB (in $20\log m$; linear m=0.25) and varied two other parameters. Here, the influences of modulation frequency (f_m=4, 8, 16, 32, and 64 Hz) and gating choices were examined. The traditional AM depth-discrimination stimuli (e.g., Wakefield and Viemeister, 1990; Lee and Bacon, 1997; Ewert and Dau, 2004) are referred to as "gated" and the envelope-domain equivalent of increment

detection as "quasi-continuous" or "fringe" conditions.

The critical difference between the gated and fringe conditions was confined to the stimulus presented between observation intervals in the three-interval paradigm. In gated conditions, a silent interval separated three modulated tones (the two standard intervals contained tones with a modulation depth m_s of 0.25; the target interval was a tone with some AM depth higher than m_s). In contrast, the quasi-continuous conditions were comprised of a 500-ms modulated tone (m_s=0.25) that was present in the two inter-observation intervals and also before the first interval and after the third and final interval. Example stimulus waveforms for the gated (top) and fringe conditions (bottom) are shown in Fig. 1(c). Stimulus amplitudes in all three intervals were gated with 50-ms ramps, regardless of the gating mode (this was also the case for the audio-frequency level discrimination stimuli).

3. Procedure

Listeners were trained until four consecutive threshold estimates in each condition showed no evidence of learning. Two additional threshold estimates were obtained if the standard deviation of the four estimates was greater than 3 dB (this happened once in all of the experiments described here). Average data are presented as the mean and standard deviation of the subjects' final depth-discrimination threshold estimates.

A three-interval, three-alternative forced-choice paradigm with visual correct-answer feedback was used along with a two-down, one-up adaptive tracking procedure (Levitt, 1971). This combination of parameters yields convergence on the 70.7% point of the psychometric function and a threshold estimate that corresponds to a d' of unity. The listeners' task was to identify the observation interval containing the higher signal AM depth. Observation-interval timing was identified with visual cues presented synchronously with the standard and target interval stimuli on the computer monitor. The stimulus parameter that was varied in the tracking procedure was the fractional AM depth increment in dB ($10 \log m_{\text{inc}}$). The initial 4-dB signal-interval step size was halved after each of the first two track reversals occurring after consecutive correct responses. Six reversals were required after the final 1-dB step size was reached; threshold for a given track was taken as the mean signal level corresponding to the target-interval AM depth used at those six points. The order of stimulus presentation was randomized across parameters (gating mode and AM frequency) for each listener.

The audio-frequency level-discrimination experimental procedures were essentially identical to those used to measure AM depth-discrimination sensitivity. The tracking variable used was also similar. The Weber fraction in dB ($10 \log \Delta I/I$) was adjusted until the target interval was just noticeably different from the two standard observation intervals.

B. Results

1. Discrimination thresholds with gated and fringe presentation modes

The magnitude of the audio-frequency gated-continuous difference was measured first; mean level discrimination results are shown in Fig. 1(b). Enhanced sensitivity to increment (fringe) conditions has been demonstrated in previous studies; the average difference for the listeners in the current study was 4–5 dB. The magnitude of the effect in our listeners was in line with the average 4.6-dB difference found at SPLs above 35 dB by Viemeister and Bacon (1988), who used a continuous 1000-Hz carrier and 200-ms observation intervals. Absolute discrimination thresholds in the gated conditions in the current study ($10 \log \Delta I/I$=−6.5 dB) are slightly better than the thresholds for matching carrier frequencies and standard levels reported by Florentine *et al.* (1987); this may be attributable to differences in presentation mode (monaural in Florentine *et al.* versus diotic in the current study).

Average modulation-depth discrimination thresholds are shown in Fig. 1(d) for a range of modulation frequencies. The most relevant aspect of the data for the purposes of the current study is the similarity in performance for the gated (circles) and fringe conditions (squares), which is in contrast to the findings in the audio-frequency domain. Performance was broadly consistent across listeners, as suggested by the size of the standard deviation bars (<1.6 dB). Listener L4 was slightly more sensitive in the fringe conditions, while L2 exhibited lower thresholds in the gated conditions. Because these individual differences were similar in magnitude and stable across AM frequency for both listeners, they effectively cancelled out in the mean data.

Mean AM-depth Weber fractions [in $10 \log((m_c^2 - m_s^2)/m_s^2)$] dropped from approximately 0 dB at 4 Hz to −4 dB at 32 and 64 Hz. These values are equivalent to target-interval depths at threshold ranging from m_c=0.35 to 0.30 for a modulated standard (m_s=0.25) and are consistent with previous studies that have found decreases in threshold at higher AM frequencies with a fixed-duration stimulus (i.e., Lee and Bacon, 1997). Thresholds in the gated condition at 16 Hz (−2.1 dB) were within 1 dB of those reported by Ewert and Dau (2004), who used a 16-Hz signal with a standard depth of m_s=0.225, among others, imposed on a 65-dB SPL, 4-kHz carrier.

2. Gated and fringe AM detection thresholds and comparison with "static" level discrimination performance

Previous studies have reported an enhancement of SAM detection thresholds at low modulation rates ($\leqslant \sim 10$ Hz) when a temporal fringe was used instead of gating the carriers (Viemeister, 1979; Yost and Sheft, 1997). The current finding of identical discrimination thresholds in the first experiment (with m_s=0.25) appears incompatible with these earlier findings. In order to determine whether the listeners in the current study also exhibited a gated-continuous difference for AM detection, an extension of the first experiment was added: thresholds in several related envelope-processing

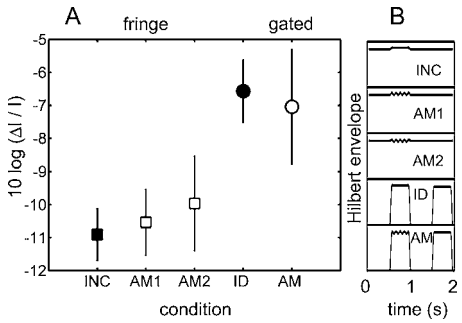


FIG. 2. (a) Mean audio-frequency level discrimination thresholds (solid symbols) and envelope-frequency detection thresholds (open symbols) under different gating conditions. Squares represent performance with a quasi-continuous carrier; circles correspond to thresholds with gated carriers. Conditions AM1 and AM2 are distinguishable based on the presence (AM1) or absence (AM2) of a dc component in the target interval. (b) Schematic illustrations of the stimulus envelopes used in each condition. Conditions INC and ID are re-plotted from Fig. 1(b). Error bars represent across-listener standard deviations (as in Fig. 1).

tasks were directly compared in an effort to better map out the differences between AM discrimination and detection, and to establish relationships between "dynamic" AM cues and "static" audio-frequency level discrimination. The question addressed is how gated and continuous level discrimination and AM detection thresholds are related. Based on the results of the first experiment, one hypothesis is that AM detection should *not* depend on the choice of gating parameters if the same mechanism underlies AM detection and AM depth discrimination.

To test this hypothesis, the dynamic AM stimuli were matched to the static level-increment stimuli in their carrier frequency (5.5 kHz), standard level (70 dB SPL), overall duration (500 ms), and onset/offset ramp duration (50 ms). Restricting the ramp duration parameter to be the same resulted in an AM stimulus that contained five cycles of 10-Hz SAM signal (50 ms on+50 ms off=100 ms period). Different combinations of gated versus fringe presentation modes and static versus dynamic fluctuations in the target interval resulted in the five test stimuli used in the extension of experiment I. Schematics of the envelope waveforms in each condition are shown in Fig. 2(b). For clarity, a two-interval task is represented (a 3AFC task was used in the actual experiment), with the target interval onset beginning at 0.5 s and the standard interval beginning at 1.5 s. Conditions with temporal fringes were identified as INC (static increment), AM1 (dynamic AM increment, with a dc offset), and AM2 (dynamic AM increment, no dc offset). Gated stimuli were labeled ID (static gated intensity discrimination) and AM (gated AM detection).

Mean data across eight listeners are shown in Fig. 2(a). Performance is defined in terms of a Weber fraction, where ΔI is determined by the difference between the maximum and minimum values of the envelope in the target interval for the AM conditions (open symbols), and by the difference in peak intensities across the standard and target intervals in the audio-frequency level-discrimination conditions (closed symbols).

One main result is that thresholds were similar for all three fringe conditions INC, AM1, and AM2 [Fig. 2(a)], and for both gated conditions ID and AM (t test p values for these comparisons were all greater than 0.13). This finding suggests that the system was not more sensitive when several dynamic temporal envelope fluctuations were presented than when a fixed energy increment with a single onset and offset was used. The similarity between thresholds in AM1 and AM2 suggests that the listeners were probably not using an overall level cue in condition AM1. The finding that the INC and AM1 and AM2 stimuli produce similar thresholds supports the hypothesis that increment detection is linked to modulation detection (and not primarily based on the detection of an energy change).

Another comparison to make in Fig. 2(a) is across the open symbols (conditions with SAM in the target interval). Listeners were more sensitive in a low-f_m AM-detection task with a temporal fringe (AM2) than with gated carriers (AM). This is in line with noise-carrier studies of Viemeister (1979) and Sheft and Yost (1990), and with the tone-carrier experiments of Yost and Sheft (1997), but seemingly at odds with the result from the depth-discrimination task, where gating parameters had no effect for discrimination of a suprathreshold AM depth. Converting the thresholds to $20 \log m$, the difference between thresholds in the AM2 condition (−32 dB) and the gated AM condition (−26.5 dB) amounts to about 5.5 dB.

C. Discussion

1. Adaptation and change detection

The similarity between gated and quasi-continuous AM depth discrimination thresholds can be interpreted in terms of the adaptation mechanisms that have been used to qualitatively explain the audio-frequency asymmetry in performance seen in gated-carrier level discrimination and continuous-carrier increment detection (see the Introduction). If an increased amount of adaptation in gated conditions underlies gated-continuous differences, then the current results suggest one of at least two conclusions in the envelope-frequency domain. Either there is little or no adaptation at the output of modulation-tuned channels or, if there is adaptation, then the response to an increment in AM depth must decrease with the same time course as the adaptation, so that the relative response increment remains constant as a function of time.

There is some peripheral physiology that initially appears consistent with a transformation supporting the latter interpretation. Smith *et al.* (1985) reported a decrease in the AM response modulation of ANFs as a function of time: the response modulation depth decreased with short-term adaptation (i.e., the effect lasted for approximately 10 ms). In contrast with the current study, Smith *et al.* (1985) used stimuli with high AM frequencies (150–600 Hz) and imposed the modulation on gated carriers with short (2.5 ms) rise-fall times. For the lower fluctuation rates (4–64 Hz) and slow (50 ms) ramp functions used here, it is unlikely that the small effect observed in peripheral physiology could have an impact on the observed similarity between gated and continuous AM depth-discrimination thresholds. This leads back to the alternative explanation, namely that there is negligible

perceptual adaptation to AM stimulation observed with the ramp and exposure durations used in the current study.

Perceptual coding of AM is usually assumed to be strongly influenced by central processing factors. This is because modulation-tuned channels are not found in the periphery, and the temporal responses of ANFs can robustly follow modulations to significantly higher rates than the several hundred Hertz (Joris and Yin, 1992) that human listeners can detect as a temporal (i.e., not spectrally resolved) cue (Kohlrausch *et al.*, 2000). The responses of cells in the inferior colliculus (IC) appear to be more tightly coupled to psychophysical measures than peripheral responses (Joris *et al.*, 2004), and temporal adaptation in responses of IC neurons to AM stimuli with relatively long onset and offset ramps is often negligible (Nelson and Carney, 2007).

An alternative way to conceptually account for the audio-frequency gated-continuous difference is to assume the existence of a modulation filter bank that processes the stimuli at the output of peripheral filters, generating an effective additional dimension. An increment in the SPL of a sound activates at least the low-frequency modulation channels, where the amount of activity depends on the exact stimulus characteristics and the transfer functions of the modulation filters. As recently shown by Gallun and Hafter (2006), increment detection thresholds can be quantitatively accounted for by assuming a modulation-frequency selective analysis. In contrast, in the gated-carrier level-discrimination conditions, the most effective cue is reflected in the dc component (or in the lowest available modulation filter) in such a model. The finding that a similar asymmetry between increment detection and level discrimination was not found in the AM domain may suggest that analogous circuitry, i.e., another "independent" (venelope) domain, does not exist, or has a negligible influence on perception.

2. Relation to previous work

Wojtczak and Viemeister (1999) compared intensity discrimination and low-f_m SAM detection with continuous-carrier pure tones across a wide range of standard SPLs and arrived at an empirical formulation of the relationship between the two measures: $10\log(\Delta I/I)=0.44(20\log m)+D(f_m)$, where $D(f_m)$ is a constant that depends only on modulation frequency. For a 4-Hz signal AM, Wojtczak and Viemeister determined this constant to be 1.7; for the 10-Hz signal AM used in the current study, $D(f_m)$ would probably take on a slightly lower value. With continuous 70-dB SPL carriers, a 10-Hz modulation rate, and assuming $D(f_m)$ to be 1.7, our data are reasonably consistent with the proposed empirical formula: $10\log(\Delta I/I)=-10.9$ dB (INC condition in Fig. 2 and $0.44(20\log m)+1.7=-12.5$ dB (AM2 condition in Fig. 2). Decreasing the value of $D(f_m)$ or inserting the modulation thresholds measured with the AM2 stimuli (SAM with a dc component) would make the equation's predictive ability worse. Wojtczak and Viemeister (1999) speculated that the empirical relationship might also hold for gated carriers, but did not test this hypothesis explicitly. The current data allow for such a test. In the gated intensity discrimination task (ID) here, $10\log(\Delta I/I)=-6.6$ dB, while in the gated AM detection condition (AM), $0.44(20\log m)+1.7=-10.0$ dB. The match to the proposed formula is worse in this condition, suggesting that it does not directly generalize to describe the relationship using gated stimuli.

III. EXPERIMENT II. TONE-IN-NOISE DETECTION WITH A MODULATED MASKER IN THE ENVELOPE-FREQUENCY DOMAIN

The goal of the second experiment was to determine the extent to which listeners could use slow and regular temporal fluctuations in the instantaneous depth of a (stochastic) masker modulation to aid in the detection of a (deterministic) sinusoidal signal modulation. The stimuli were designed to maximize the availability of potential release-from-masking cues in an envelope-domain transposition of a typical audio-frequency CMR paradigm.

A. Methods

Details of the listeners, apparatus, and procedure were the same as in the first set of experiments. This section addresses any remaining differences, which were mainly limited to stimulus parameters.

The carrier was again a 5.5-kHz pure tone with 50-ms raised-cosine windows applied to the onset and offset. The overall SPL of the standard and target were normalized to have the same rms level as a 70-dB SPL pure tone. Observation intervals were separated by a 500-ms silent interval. In the two standard intervals, the tonal carrier was modulated by a Gaussian noise, which had a 120-Hz bandwidth (BW) and was geometrically centered around 32 Hz, from 8 to 128 Hz. The average masker modulation depth was −13.2 dB rms ($m=0.22$; for a 120-Hz BW the noise had a "spectrum level" of −34 dB). This combination of masker depth and BW was chosen to (1) ensure significant masking of the 32-Hz signal AM (presented only in the target interval), (2) avoid overmodulation (no stimuli with modulation depths greater than one were presented to the listeners), and (3) provide a potential opportunity for across-modulation-channel processes to enhance detection performance [by using a BW greater than that of the putative modulation filters, which are typically described as having half-power Q-values between 0.5 and 2 (e.g., Lorenzi *et al.*, 2001; Wojtczak and Viemeister, 2005), or effective BWs between 16 and 64 Hz for a channel tuned to 32 Hz].

Masker waveforms in each interval were independent noise realizations, generated digitally by setting the Fourier coefficients outside the desired pass-band to zero. In the baseline conditions (analogous to unmodulated conditions in audio-frequency CMR experiments), no further manipulations were made of the masker waveform before the 32-Hz sinusoidal AM (always in sine phase) was added and the resulting envelope signal imposed on the carrier. A general equation for the final stimulus is

$$s(t) = c\{\sin(2\pi f_c t)[1 + m\sin(2\pi f_m t) + M(t)]\},$$

where c is a scalar that equalizes the overall audio-frequency power in each interval, f_c is the carrier frequency, m is the target modulation depth (zero in the standard interval), f_m is the signal modulation frequency (32 Hz), and $M(t)$ is the

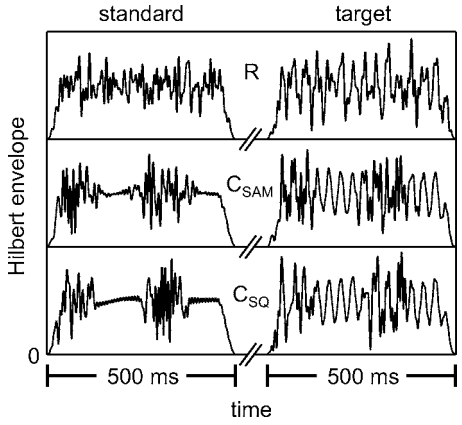


FIG. 3. Example temporal envelopes of the stimuli used to test for envelope-domain comodulation masking release. Standard-interval envelopes (left) are defined by the masker-alone waveform; target interval envelopes (right) are made up of an additive combination of masker and sinusoidal signal AM. R: Baseline (unmodulated, or random masker) condition. C_{SAM}: Sinusoidal venelope fluctuations. C_{SQ}: Square-wave venelope fluctuations. For all stimuli, f_c=5500 Hz; standard SPL=70 dB; masker BW =120 Hz, geometrically centered on the 32-Hz signal frequency; observation interval duration=500 ms; signal depth m=40%; venelope fluctuation rate=4 Hz.

masker waveform. In contrast to the baseline condition, the comparison, or "comodulated," masker waveforms were further processed before being combined with the signal SAM. Slow, coherent, and regular (sinusoidal or square wave) temporal fluctuations were imposed on the instantaneous masker modulation depth, $M(t)$. This resulted in a stimulus with a time-varying venelope. In all of the comparison conditions, the imposed venelope fluctuations were maximal in the sense that the nominal envelope depth of the masker varied between zero and the peak value. Examples of the time waveforms that were used to modulate the carrier are shown in Fig. 3; masker-alone (standard) waveforms are illustrated on the left, signal-plus-noise envelopes are shown on the right. Baseline unmodulated masker (R), sinusoidally modulated masker (C_{SAM}), and square-wave-modulated masker (C_{SQ}) conditions are shown.

Imposing slow fluctuations in the envelope can affect the resulting modulation spectra (i.e., sidebands are generated when the envelope is modulated, just as they are in the audio-frequency spectrum when a carrier is modulated). One way to avoid this complication is to filter the noise after it is modulated; the trade-off when using this strategy is that the temporal waveform is slightly changed, usually in the form of ringing caused by the band-limiting. To control for this issue, thresholds were measured when the masker envelope bandwidth was limited either before (condition C'_{SAM}) or after (condition C_{SAM}) imposing the slow venelope fluctuations in both the baseline and comparison conditions.

In the first part of the experiment, the venelope fluctuation rate was fixed at 4 Hz (two cycles were presented for each 500-ms signal), and the waveform used to modulate the (first-order) AM noise was varied, both in terms of its shape and its amplitude. In the extension of the experiment, the duration of the signals was extended to 2 s to allow for the use of even slower venelope fluctuation rates (from 0.5 to 4 Hz). An equal-energy (in terms of the envelope rms), square-wave venelope masker was used with the 2-s signals.

For comparison, thresholds were also measured for the same listeners in an audio-frequency CMR paradigm with parameters designed to (loosely) parallel those used in the envelope-frequency experiment. In both the audio- and envelope-domain experiments, detection thresholds of a mid-frequency sinusoidal signal embedded in a moderately intense and wideband (with regard to putative bandwidths of modulation or auditory filters) Gaussian masker were measured. Slow and regular fluctuations were imposed on the masker in both domains; release from masking was defined as improved thresholds in the conditions using modulated maskers over those using noises with flat temporal venelopes or envelopes.

Specific audio-frequency parameters were 2-kHz signal frequency, 800-Hz masker bandwidth (geometrically centered on the signal frequency), a masker spectrum level of 30 dB SPL (overall rms level=59 dB SPL), and a 32-Hz (first-order) sinusoidal AM imposed on the masker. Observation and interstimulus intervals were 500 ms. The tone level was adaptively varied initially in steps of 8 dB; the initial step size was halved after each of the first two track reversals occurring after consecutive correct responses until it reached 2 dB. Again, the mean of six reversals was taken as threshold for a given track. These parameters were chosen to maintain the same within-channel to across-channel energy ratios in the audio- and envelope-frequency CMR experiments. Assuming a typical 3-dB effective bandwidth of 200 Hz at 2 kHz in the audio-frequency task, the ratio of within-channel to across-channel energy was approximately 200 Hz: 800 Hz (or $\sim$1:4), which is similar to the ratio of 32 Hz: 120 Hz used in the envelope-domain task.

B. Results

1. No release from masking in the envelope-frequency domain for 500-ms *stimuli*

The magnitude of audio-frequency CMR with a wideband masker centered on the tone frequency and fully modulated by a deterministic waveform in the comodulated conditions is shown in the left panel of Fig. 4. Thresholds were about 10 dB lower in conditions with a comodulated masker (C) than in the random (flat masker) case (R). The magnitude of the effect is close to that observed in a previous study using similar stimulus parameters (Verhey *et al.*, 1999).

The new contribution of the current experiment was to translate the parameters that result in significant audio-frequency CMR into the envelope-frequency domain. Pure-tone carrier SAM detection thresholds were measured in the presence of several types of additive modulation maskers. A release from masking would take the form of lower thresholds in the conditions with slow and regular variations imposed on the instantaneous masker modulation depth when compared to performance in the conditions with a flat-envelope (Gaussian) masker modulation.

Within the right panel of Fig. 4, SAM detection thresholds are shown for the four masker conditions described

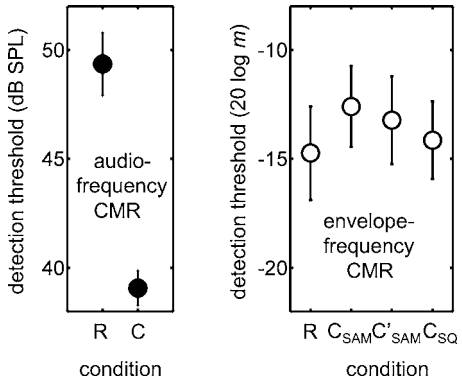


FIG. 4. Effects of imposing slow and regular fluctuations on the masker amplitude in the envelope- and audio-frequency domains. Conditions correspond to different temporal shapes imposed on the masker amplitude. Left panel: Audio-frequency thresholds. R: random flat masker envelope (unmodulated). C: 32-Hz SAM masker envelope, filtered after modulation and not equalized for overall energy increment caused by modulation. Right panel: Envelope-frequency thresholds. R: flat venelope masker (unmodulated). C_{SAM}: 4-Hz SAM venelope, noise filtered after modulation. C'_{SAM}: 4-Hz SAM venelope, noise only filtered prior to modulation. C_{SQ}: 4-Hz square-wave venelope, noise filtered after imposing the 4-Hz fluctuations. Conditions C_{SAM}, C'_{SAM}, and C_{SQ} were compensated for the small overall increase in masker energy caused by the modulation. Error bars indicate ±1 standard deviation of the individual mean thresholds

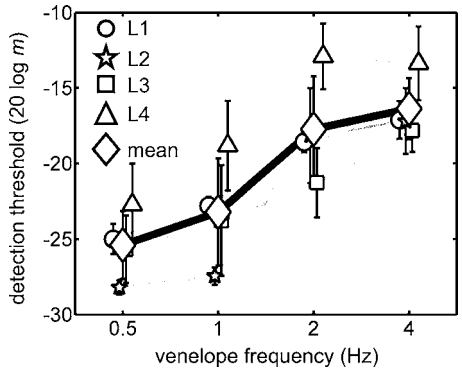


FIG. 5. Individual and mean 32-Hz SAM detection thresholds as a function of the frequency of the square-wave venelope fluctuations imposed on the first-order masker modulation. Stimulus parameters were the same as those in condition C_{SQ} of Fig. 4, except the overall duration was increased to 2 s. Error bars plotted with individual listener data represent across-track standard deviations; those with the mean data indicate across-listener standard deviations.

above. The average thresholds were all between −12 and −15 dB (20 log m), and none of the comodulated condition thresholds were significantly different from those measured in the random condition (t-test p values=0.18, 0.34, and 0.68). The results indicate that the listeners were unable to take advantage of the slow and regular venelope fluctuations imposed on the first-order masker.

2. Extending the time course of the slow masker fluctuations

In the square-wave venelope masker conditions above, the listeners were presented with two 125-ms segments of the unmasked pure SAM 32-Hz signal in the target interval (four complete cycles) between two 125-ms segments containing both the signal and masker modulation. This duration of pure-signal AM was insufficient to give rise to a release from masking. However, intuitively, one expects that there *must* be a release from masking if the periods of low masker energy are long enough. To further investigate the time course of the effect, it was necessary to increase the overall duration of each interval to 2 s to accommodate more than one cycle of the slow masker modulation. Square-wave venelope waveforms with rates of 0.5, 1, 2, and 4 Hz were imposed on the same modulation masker used with the 500-ms signals (120-Hz BW geometrically centered on the 32-Hz signal rate, with an average depth of −13.2 dB rms), and detection thresholds were determined again for a 32-Hz signal AM.

Detection thresholds for the 2-s stimuli are shown in Fig. 5. Individual thresholds are shown in addition to the mean results (diamonds) because of the relatively high intersubject variability. For all four of the listeners, performance improved with decreasing venelope fluctuation rates over the range of frequencies tested. The parameters of the stimuli used in the 4-Hz condition were identical to those used with the 500-ms stimuli (condition C_{SQ} in Fig. 4); mean performance improved by about 2.5 dB as a result of increasing the stimulus duration.

Thresholds asymptote near the expected pure-AM detection thresholds for listeners L1, L2, and L3; the remaining listener (L4) was less sensitive overall and continued to show improvement between the 1- and 0.5-Hz conditions. Overall, the listeners required at least 500-ms periods of unmasked SAM signal between masker bursts to reach performance near the 32-Hz SAM detection thresholds expected for a 70-dB SPL, 5.5-kHz pure tone carrier (between −25 and −30 dB; e.g., Kohlrausch *et al.*, 2000). The relatively high thresholds observed for the 2- and 4-Hz venelope frequencies suggest that the perceptually relevant decision variable is either integrated over a long interval, or that the internal representation of the signal AM is affected by preceding masker modulations for several hundred milliseconds. Anecdotally, listeners reported that the masker bursts were perceptually fused for venelope rates above 2 Hz and gradually became identifiable as temporally distinct and separable events at rates below 2 Hz.

C. Discussion

The findings from experiment II are consistent with those of experiment I in that venelope fluctuations did not appear to contribute to performance in envelope-frequency-domain versions of either task. In other words, there does not appear to be an additional independent coding dimension that the listeners have access to in the translated (modulation domain) experiments as there apparently is in the audio-frequency domain.

1. Relation to previous work

The results of our effort to measure CMR in the envelope-frequency domain are in qualitative agreement with the findings of several previous studies. Wojtczak and Viemeister (2005) also showed that a modulated envelope preceding a SAM signal imposed on the same carrier could affect detection performance for masker-probe delays of up

to 200 ms. Their AM forward-masking paradigm used a wideband noise carrier, a sinusoidal masker AM, and a short (50-ms) signal that was present only after the masker. The broadly similar time course of masking observed in the two studies suggests that a single mechanism could underlie both effects, and that it is independent of the statistical description of the carrier and masker. It is not obvious what this mechanism might be; in fact, if one assumes ringing or persistence as the underlying mechanism, neither effect would be expected based on the broadly tuned nature of the putative modulation filters, since the "trade-off" for implementing broad signal-processing filters is that the response recovers quickly from stimulation.

The information available to cochlear implant (CI) users is provided largely in the form of temporal envelope fluctuations imposed on the amplitude or duration of current pulses presented to stimulating electrodes. Nelson *et al.* (2003) and Nelson and Jin (2004) measured performance of CI users in a speech recognition task with a temporally modulated noise masker, and varied the "gate frequency" from 1 to 32 Hz. They found that CI users did not benefit from temporal gaps in the noise masker as long as 500 ms (performance was independent of the gate frequency). Conclusions from the CI experiments seem consistent with those from the current study: in conditions with severely impoverished spectral-frequency cues, listeners are unable to use relatively long temporal gaps in a noise masker to aid in the detection of a signal.

2. Interpreting time courses

The extended-time-course AM-detection experiment suggests a long integration time constant operating at some stage presumably central to the putative envelope-filtering process. Such a statement is consistent with "long-term" masked AM detection decision statistics that quantify responses based on an averaged representation of the processed stimulus envelope, such as envelope rms (e.g., Strickland and Viemeister, 1996; Ewert and Dau, 2000; Ewert *et al.*, 2002) or the average firing rate of a model IC cell (Nelson and Carney, 2006). However, it is worth pointing out that the current data set does not necessitate the assumption of such a time-averaged decision variable. It remains possible that a "local feature" decision variable (e.g., envelope max/min ratio or maximum local modulation depth) could be used, but that the listeners combine information from multiple looks (e.g., Sheft and Yost, 1990; Viemeister and Wakefield, 1991) of the details of the local features.

Cortical physiological studies have provided evidence for long-lasting modulation of responses to envelope fluctuations that might underlie the apparent perceptual sluggishness observed here. Using pure-tone forward masking paradigms in the primary auditory cortex (A1), several groups have shown that the response to a short probe signal could be affected by the presence of a masker that preceded the probe by as much as several hundred ms or longer (e.g., Calford and Semple, 1995; Brosch and Schreiner, 1997; Ulanovsky *et al.*, 2004). If the masker had a similar audio-frequency composition to that of the probe (as it did in the current study), the response to the probe was usually suppressed. In a recent study of the unanesthetized marmoset A1 that used stimuli more similar to those used in the current psychophysical experiments, Bartlett and Wang (2005) examined the contextual dependence of AM responses on previous stimulation. Their findings were in qualitative agreement with those of the AM forward masking studies, but the observed suppression (or facilitation) of a probe AM stimulus could last longer than 1 s in some neurons and depended on the modulation properties of the preceding stimulus. To date, the authors are not aware of physiological results at any level that address the effect of a nonsimultaneous masker modulation imposed on the same tone carrier as a deterministic signal modulation. In all of the studies mentioned above, the probe and masker were imposed at different times on separate carriers (i.e., the stimuli were gated). It would be interesting to know whether physiological time courses of adaptation to AM are different for gated and continuous carriers.

IV. MODELING

A. Methods

A physiologically motivated processing model developed to predict peripheral and central neural responses to pure SAM tones (Nelson and Carney, 2004) and psychophysical responses to masked SAM tones (Nelson and Carney, 2006) was used to simulate responses to the stimuli used in the current study. The peripheral model was a modification of previous AN models (Carney, 1993; Zhang *et al.*, 2001; Heinz *et al.*, 2001a), and the final model output can be compared to pure-tone onset response cells in the IC. Interactions between fast excitation and slow inhibition give rise to modulation tuning in the model IC cells. Since the two inputs are matched in audio-frequency CF, the model is referred to as the same-frequency inhibition and excitation (SFIE) model, as in Nelson and Carney (2004). In the current study, the relative strength of the inhibition with respect to the excitation of the model IC cells ($S_{INH,IC}$) was set equal to 1.0. This parameter was important for determining the threshold modulation depth required to elicit firing in the model cells: values of $S_{INH,IC} \leqslant 1$ resulted in lower depth thresholds than $S_{INH,IC}$ values greater than one (i.e., stronger inhibition re: excitation). The time constants of inhibition (τ_{inh}) and excitation (τ_{exc}) were chosen to yield a cell that was tuned to the signal f_m of interest (see Nelson and Carney, 2004).

B. Results and discussion

Simulation results are described and discussed with three specific psychophysical observations in mind: (1) audio-frequency level-discrimination thresholds depend on the choice of gating mode (experiment I), (2) AM detection thresholds depend on gating mode (experiment I extension) while AM depth-discrimination thresholds do not (experiment I), and (3) masked SAM detection thresholds do not readily improve when the masker is comodulated (experiment II). The first finding is examined most carefully with the model, and those results are used as justification for the assumptions made with the remaining sets of data. In general, the modeling work is meant to qualitatively test the ability of an existing, physiologically motivated envelope-

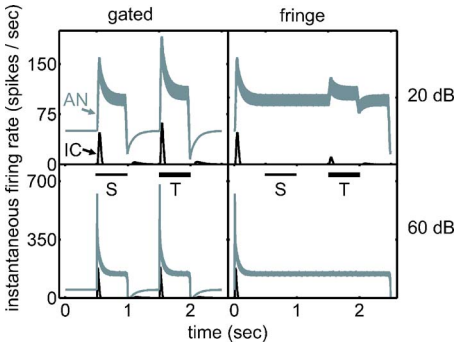


FIG. 6. Simulated responses to standard and target stimuli for the AN and IC levels of the SFIE model. Upper panels: 20 dB SPL standard level; lower panels: 60 dB SPL standard level. The target interval level was 3 dB higher than the standard. Left panels: gated stimuli; right panels: fringe presentation mode.

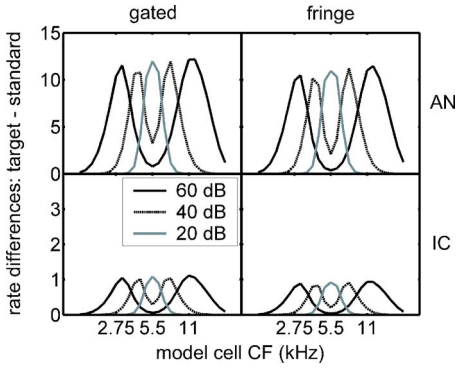


FIG. 7. Rate difference profiles for different gating modes (left and right panels) and levels of the model (upper and lower panels) in response to a 3-dB level increment. Each curve represents changes in the model rate responses for a 5500-Hz tone with a fixed standard SPL. Twenty-five model cells, log-spaced from 1375 to 22 000 Hz, were simulated for each standard level and gating mode.

processing model to account for broad and basic features of the data. This approach intentionally lies between speculating on potential mechanisms and explicitly predicting listeners' thresholds with a specific (i.e., fitted) model.

1. Audio-frequency domain level discrimination with gated and continuous carriers

By definition, modulation-tuned neurons are also envelope change detectors, and the properties that underlie AM responses in these neurons can also qualitatively explain the audio-frequency gated-continuous difference. To illustrate this point and to provide a concrete example of a component of a realistic neural circuit that predicts a heightened sensitivity to increments over changes in the level of gated tone bursts, the SFIE model was applied to the audio-frequency stimuli used in the current study.

Instantaneous firing rate (IFR) functions are shown in Fig. 6 for the responses of two stages of the model. The functions are comparable to physiological peri-stimulus time histograms (PSTHs) obtained for the AN (e.g., Harris and Dallos, 1979) and the IC (e.g., Langner and Schreiner, 1988) and were generated for eight illustrative conditions. Both AN and IC model responses are included in each of the panels of Fig. 6, which correspond to a specific combination of gating mode (gated or fringe) and standard SPL (20 or 60 dB). The timing of the standard (S) and target (T) stimulus presentation is marked by the two horizontal bars from 0.5–1 s and 1.5–2 s. A 3-dB level increment in the target interval was used. Other parameters of the stimuli were matched to those used in the psychophysical experiment. The IC model time constants ($\tau_{exc}=10$ ms; $\tau_{inh}=20$ ms) were chosen to yield a cell tuned to the effective 10-Hz modulation rate caused by the 50-ms onset and offset ramps.

First, consider the model outputs in response to a 20-dB standard-interval SPL tone (upper panels of Fig. 6). Here, the most critical differences between the modeled AN and IC responses are in the steady-state portion: ANFs respond with sustained firing to pure-tone stimulation, while the IC model only fires when the stimulus envelope elicits a change in the peripheral response. This is most clearly seen at envelope transitions with rising slopes, but offset adaptation in the peripheral model also results in a small response at the offset of the gated stimuli in the IC model. Note also that the IC model responds to both standard and target intervals when the intervals are gated (left panel), but only to the target interval in the fringe conditions (right panel). The AN model responses are always nonzero when a stimulus is present, although there is no change in the response from the baseline to the fringe-condition standard interval.

Because of rate saturation effects in the AN model, the pattern of model responses was quite different when a 60-dB standard tone SPL was used (Fig. 6, lower panels). Specifically, the target interval increment did not elicit a change in the fringe-condition responses of either the AN or IC model. As a result, a model consisting only of high-spontaneous rate (SR), on-CF ANFs at medium to high SPLs predicts unrealistically high level discrimination thresholds (Colburn *et al.*, 2003). There are two popular ways to account for psychophysical performance at high SPLs and high frequencies. One is to unevenly and heavily weight the contribution of low-SR ANFs (Winslow and Sachs, 1988; Delgutte, 1987; Viemeister, 1988). This approach is not completely satisfying, because such high-threshold, wide dynamic range ANFs make up only ~15% of the total population in cat (Liberman, 1978) and only exist at high CFs (>1500 Hz, Winter and Palmer, 1991). Another aspect of the response to high-level tones that may provide information for discrimination is in the spread of excitation across a population of neurons (Viemeister, 1972; Florentine *et al.*, 1987; Heinz *et al.*, 2001b; Colburn *et al.*, 2003). To address this issue, standard and target stimuli were presented to a group of model cells with different CFs.

Responses across the population were quantified in terms of their average rate over the entire 500 ms of the stimulus. Peripheral (AN) differences in the model's rate responses are shown in the upper panels of Fig. 7 and central (IC) rate differences are plotted in the lower panels. The parameter in each panel of Fig. 7 is the standard level. As in Fig. 6, gated and fringe conditions are illustrated in the left and right panels, respectively. For all four combinations of gating mode and model stage, the biggest differences in rate between the target and standard interval moved progressively away from the tone frequency (5500 Hz) as the standard SPL

was increased. This is consistent with previous studies using simulations of high-SR ANFs (e.g., Siebert, 1968; Heinz *et al.*, 2001b; Colburn *et al.*, 2003) and is caused entirely by saturation in the present model (because a linear basilar membrane model was used). A model version with level-dependent bandwidth and gain [i.e., a time-varying compressive nonlinearity (Heinz *et al.*, 2001a)] was also tested, and a similar pattern was obtained, suggesting that effects caused by saturation dominate those caused by compression when low-threshold, high-SR ANFs are used to estimate the population response.

Another feature of the differences in model rates was that the shapes of the profiles were similar for the responses of both stages of the model. This was a direct result of the simplified nature of the SFIE IC model neurons. Absolute values of the rate differences were significantly higher in the AN model (note the different scales for the upper and lower panels); this was not surprising given the sustained nature of responses to pure tones in the AN model and transient characteristics of the IC model responses. Finally, comparing across the gated and fringe conditions, the changes in model rates were not strongly dependent on the mode of gating for either the AN or IC model population.

The similarity of absolute rate differences for the gated and fringe conditions for both stages of the model does not provide a compelling explanation of the gated-continuous difference. It does, however, lead to a consideration of another feature of neural responses that must be known (or assumed) before predicting performance: the variability of rate estimates (e.g., Siebert, 1965). In actual ANFs, rate variability can be reasonably described as Poisson, with spike-count variance approximately equal to the mean count [at least at low rates, see Young and Barta (1986) and Winter and Palmer (1991) for a description of the reduced-variance deviation from Poisson at high rates]. The situation is less clear in more central processing stations, but, for simplicity, Poisson variance will be assumed in the responses of both stages of the model. These variance characterizations allow for a relatively simple formulation of the information provided by each frequency channel in the population rate response [following the approach of Siebert (1965); for details and derivations, also see Heinz *et al.* (2001b)]:

$$\delta'^2(icf) = \frac{[(\text{rate}_T - \text{rate}_S)/I_{\text{inc}}]^2}{\sigma^2_{\text{rate}}},$$

where $(\text{rate}_T - \text{rate}_S)$ is the rate difference term plotted in Fig. 7, I_{inc} is the size of the target-interval increment (a value of 1 results in the 3-dB amplitude increment in the examples shown), and σ^2_{rate} is the variance of the rate response. Our Poisson assumption allows for a simple estimate of σ^2_{rate}: it is assumed to be equal to the average rate (across both standard and target responses).

Information profiles, which incorporate both changes in rate and contributions of assumed neural variability, are shown in Fig. 8, in a format identical to that used to visualize the rate differences alone in Fig. 7. The limits of the ordinates are identical in all four panels. When the response variability is taken into account, the AN population model still predicts no advantage in the fringe condition relative to the gated presentation mode. In contrast, the envelope-change-detecting IC model clearly predicts a heightened sensitivity to the fringe condition, for all three tested standard levels. In addition, the overall summed population d' (related to the area under the information profile curve) is higher for the fringe-stimulus IC model rate responses than for the peripheral AN model responses. The gated-continuous difference in the IC model is strongly influenced by the lower average rate in the fringe condition, which translates into lower assumed variability and higher values of d'.

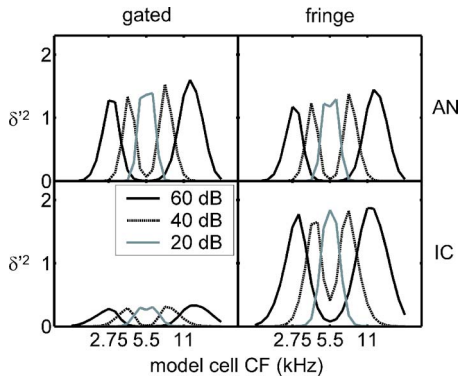


FIG. 8. Across-frequency information profiles, arranged in the same format as Fig. 7. This measure of sensitivity takes both neural variance and changes in average rates into account.

While the exact values of predicted d' and thresholds depend on details of the parameters of the model and the statistical description of the chosen internal noise assumption, overall trends and the difference between gated and fringe conditions for the SFIE model with equal amplitude inhibition and excitation do not. One of the key features of the IC model that underlies the current explanation of the audio-frequency gated-continuous difference is the fact that there is some response to both intervals in the gated condition, and only a response to the target interval in the fringe condition. The other critical assumption is an internal noise process that predicts response variability that increases with average rate. Such a change-detection mechanism could in theory be independent of peripheral adaptation, although there is an interaction between the two in the model, and some interplay probably exists in the real system as well.

In contrast to psychophysics, where at standard levels lower than about 20 dB above detection threshold, continuous and gated pedestals yield similar measures of performance (Carlyon and Moore, 1986; Viemeister and Bacon, 1988), the SFIE model predicts a fringe advantage for all supra-threshold standard SPLs. One way to potentially modify the model to account for this level dependence would be to add a second internal noise source with a fixed variance (in addition to the assumed Poisson noise) to the final model response.[1]

2. Envelope-frequency domain modulation detection and discrimination with gated and continuous carriers

AM detection thresholds (at least at low modulation rates) depend on whether the carrier is gated or quasi-continuous (Fig. 2 of the current study; see also Viemeister,

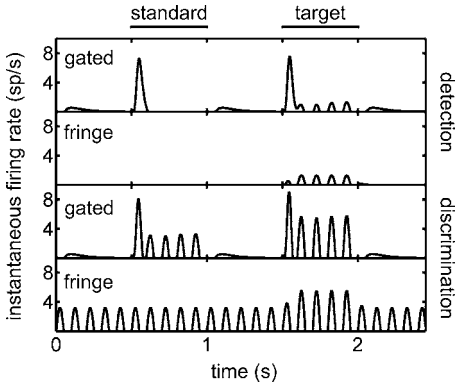


FIG. 9. SFIE model responses are qualitatively consistent with a fringe advantage in AM detection and no fringe advantage in AM depth discrimination. Model responses are shown for a 2.5-s window centered on the presentation of a standard followed by a target modulation. Simulated PSTHs are shown for an AM detection paradigm (top two panels) and an AM depth discrimination task (bottom two panels); gated and fringe conditions are included for both tasks. Stimulus parameters: f_c=5.5 kHz; SPL =70 dB; f_m=10 Hz; detection m=−20 dB; discrimination m_s=−12 dB, m_c =−7 dB. Key model parameters: τ_{exc}=10 ms; τ_{inh}=20 ms; $S_{INH,IC}$=1; AN CF=2000 Hz.

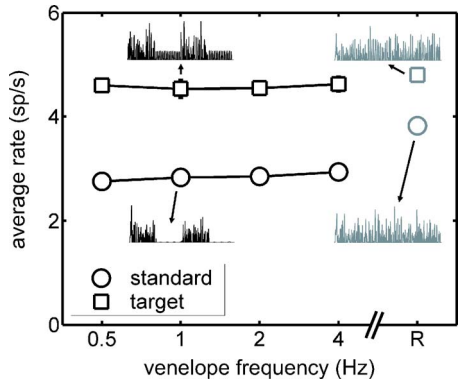


FIG. 10. Model IC cell average rates and example IFR functions in response to the stimuli used in the envelope-frequency CMR paradigm. Model parameters were the same as those used to generate the responses in Fig. 9, except the AN CF=2250 Hz, τ_{exc}=3 ms, and τ_{inh}=10 ms, which resulted in a cell rate-tuned to the 32-Hz signal AM frequency. The signal depth was fixed at −15 dB (20 log m) in the target interval, and a square-wave venelope was imposed on the masker in the comodulated conditions (as in the extension to experiment II). For comparison, rates and IFRs elicited by the random (R), or unmodulated, condition are also included in the plot. The duration of stimuli and IFRs was 2 s.

1979; Sheft and Yost, 1990; Yost and Sheft, 1997): thresholds are significantly higher when a gated carrier is used. In contrast, it was found that AM depth-discrimination thresholds (m_s=0.25) were statistically identical for gated and quasi-continuous carriers. Based on the results from the preceding section, analysis of the model in this section will be focused on IC model responses away from CF, where the biggest differences between standard and target were found.

IC model IFR functions are shown in Fig. 9 for 10-Hz SAM detection (top two panels) and 10-Hz SAM depth discrimination (bottom panels) paradigms. The observation intervals were 0.5 s; the standard interval started at 0.5 s and the target at 1.5 s. Labels in the upper left corner of each panel indicate the gating mode for each response. Simulated PSTHs for the AM-detection paradigm were similar to the IC model responses in Fig. 6 for audio-frequency level discrimination, in that the gated stimuli elicited a response in both the standard and target interval, while the fringe stimulus resulted in a model response only in the target interval. Again, if differences in both rate and variance are considered, model responses predict an enhanced sensitivity to the fringe condition compared to the gated condition (see preceding section).

In contrast, for AM depth discrimination (lower panels of Fig. 9) the IC model responded to both standard and target interval in the gated and fringe conditions. Since both rate differences and assumed rate variability are similar in these conditions, the IC model predicts little or no difference in thresholds between the gated and fringe presentation modes (as observed in the data of experiment I).

3. CMR experiment

Two questions were investigated concerning the ability of the model responses to qualitatively predict psychophysical trends observed with the stimuli used in the envelope-frequency-domain CMR experiment. First, does the model correctly predict the absence of a release from masking with the 4-Hz venelope fluctuation rate and 32-Hz signal SAM rate, as used in the base experiment? Second, is there an effect of venelope fluctuation rate over the range considered in the extension to the base CMR experiment?

To address the first question, a fixed-level signal SAM (−15 dB in 20 log m) was added to the masker in the target interval; model responses were quantified in terms of their average firing rates (across ten independent noise waveforms) in the random (R) unmodulated condition and in the 4-Hz square-wave comodulated masker condition for both standard and target intervals. If the model were consistent with the listeners' thresholds, the difference in the response to the standard and target intervals should be independent of the masker condition (random or comodulated at 4 Hz). Figure 10 shows that this was not the case; average rates in the random condition (gray symbols) were more similar in the standard (○) and target (□) intervals than the rates in the 4-Hz comodulated condition (right-most connected points). The disparity is caused by the reduced response magnitude in the standard intervals of the comodulated conditions; target-interval rates were largely independent of the venelope fluctuation patterns imposed on the masker.

The overall long-term envelope rms energy was identical in all of the noise-alone (standard) intervals shown in Fig. 10; the suppression in rate for the comodulated condition relative to that for the random condition was therefore caused by a nonlinear relationship between envelope rms and model IC cell average rate. The main factor contributing to this relationship was the change in the slope of the rate versus stimulus modulation depth function: at low m, the slope was shallower than at higher m. For example, when the modulation depth of a pure 32-Hz SAM signal was varied and presented to the cell simulated in Fig. 10, the slope of the function was ~0.1 sp/s/dB for $-30 \text{ dB} < 20 \log m < -25$ dB, and ~0.9 sp/s/dB for $-5 < 20 \log m < 0$ dB. This means that responses to small effective modulation depths (such as the "ripples" caused by postmodulation filtering of

the masker, or fluctuations in the masker away from the cell's best modulation frequency) were strongly attenuated in the model, which resulted in a reduced overall response to the modulated standard-interval stimulus.

Taken together, the rate responses of the model IC cell were not consistent with the listeners' inability to use the fluctuations in the masker to improve thresholds in the masked detection task. The schematic IFR functions included in Fig. 10 along with the rate quantifications show that the signal representation in the temporal responses of the model IC cells also suggest an increased salience of the signal in the comodulated conditions. The long effective time constants apparently necessary to explain the time course of release from masking observed in the extension of experiment II (on the order of hundreds of ms) are not included in model IC cells tuned to a 32-Hz signal frequency; as a result, the current model does not predict the psychophysical increase in thresholds with venelope fluctuation rate (connected symbols in Fig. 10.)

V. SUMMARY AND CONCLUSIONS

Two audio-frequency paradigms were translated into the envelope-frequency domain to assess the perceptual salience of venelope (second-order envelope) cues in continuous-carrier depth discrimination and SAM detection in the presence of a slowly varying noise masker. The experiments described here suggest a weak effect of temporal structure on performance in the two envelope-processing tasks. Several conclusions can be drawn from the empirical data and modeling results:

(i) Tone-carrier SAM-depth discrimination thresholds are not dependent on the gating mode of the carrier (i.e., gated or quasi-continuous), for a standard modulation depth of −12 dB (m_s=0.25) and modulation frequencies from 4 to 64 Hz. This contrasts with audio-frequency level discrimination results, which clearly indicate a heightened sensitivity to level increments when compared to gated-carrier level discrimination measurements.

(ii) SAM detection thresholds (or discrimination with a standard depth m_s=0) are approximately 5–6 dB lower when a quasi-continuous carrier is used than when the observation-interval stimuli are gated (for f_m=10 Hz).

(iii) Masked detection thresholds of a 32-Hz signal AM do not improve when the masker is slowly and coherently modulated with a 4-Hz venelope fluctuation rate. This is true for both sinusoidal and square-wave shaped comodulation. Audio-frequency tone detection thresholds, on the other hand, are strongly affected by the properties of the masker modulation.

(iv) To observe CMR in the envelope-frequency domain, the period of masker modulation must be lengthened until the masker bursts occur as perceptually distinct events (i.e., venelope fluctuation rates $\leqslant$1–2 Hz for a 32-Hz signal).

(v) A simple model developed to predict responses to SAM tones in the auditory midbrain can qualitatively account for several of the results, including the gated-continuous difference for pure-tone level discrimination and AM detection and the gated-continuous "similarity" for AM depth discrimination. The model does not, however, explain the listeners' inability to use relatively slow fluctuations in the instantaneous masker modulation depth to improve performance in the envelope-domain CMR task. Higher-order processing, possibly associated with auditory grouping and/or segregation mechanisms, may need to be considered to account for results in the CMR task.

ACKNOWLEDGMENTS

This research was supported by Grant Nos. NIH-NIDCD F31-7268 (PCN) and NIH-NIDCD R01-01641 (LHC, PCN) and by the Danish Research Council (TD, SDE). Discussions with Magdalena Wojtczak and Neal Viemeister were particularly helpful in preparing this paper.

[1]This adjustment would reduce the difference in rates at very low SPLs, when the standard level is near threshold. Because the effect of standard SPL was not a focus of the current behavioral experiments, such a modification was not included in the simulations. An addition of such a fixed-variance (stimulus-independent) internal noise would be conceptually similar to one of the sources of internal noise required to account for AM depth discrimination in Ewert and Dau (2004).

Bacon, S. P., and Viemeister, N. F. (**1994**). "Intensity discrimination and increment detection at 16 kHz," J. Acoust. Soc. Am. **95**, 2616–2621.

Bartlett, E. L., and Wang, X. (**2005**). "Long-lasting modulation by stimulus context in primate auditory cortex," J. Neurophysiol. **94**, 83–104.

Brosch, M., and Schreiner, C. E. (**1997**). "Time course of forward masking tuning curves in cat primary auditory cortex," J. Neurophysiol. **77**, 923–943.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958–1965.

Calford, M. B., and Semple, M. N. (**1995**). "Monaural inhibition in cat auditory cortex," J. Neurophysiol. **73**, 1876–1891.

Campbell, R. A., and Lasky, E. Z. (**1967**). "Masker level and sinusoidal-signal detection," J. Acoust. Soc. Am. **5**, 972–976.

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**, 461–464.

Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**, 401–417.

Colburn, H. S., Carney, L. H., and Heinz, M. G. (**2003**). "Quantifying the information in auditory-nerve responses for level discrimination," J. Assoc. Res. Otolaryngol. **4**, 294–311.

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

Delgutte, B. (**1987**). "Peripheral auditory processing of speech information: implications from a physiological study of intensity discrimination," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Dordrecht, Nijhoff), pp. 333–353.

Derleth, R. P., and Dau, T. (**2000**). "On the role of envelope fluctuation processing in spectral masking," J. Acoust. Soc. Am. **108**, 285–296.

Durlach, N. I., and Braida, L. D. (**1969**). "Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Ewert, S. D., and Dau, T. (**2004**). "External and internal limitations in amplitude-modulation processing," J. Acoust. Soc. Am. **116**, 478–490.

Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). "Spectro-temporal processing in the envelope-frequency domain," J. Acoust. Soc. Am. **112**, 2921–2931.

Florentine, M., Buus, S., and Mason, C. R. (**1987**). "Level discrimination as a function of level for tones from 0.25 to 16 kHz," J. Acoust. Soc. Am.

81, 1528–1541.
Füllgrabe, C., and Lorenzi, C. (**2005**). "Perception of the envelope-beat frequency of inharmonic complex temporal envelopes," J. Acoust. Soc. Am. **118**, 3757–3765.
Füllgrabe, C., Moore, B. C. J., Demany, L., Ewert, S. D., Sheft, S., and Lorenzi, C. (**2005**). "Modulation masking produced by second-order modulators," J. Acoust. Soc. Am. **117**, 2158–2168.
Gallun, F. J., and Hafter, E. R. (**2006**). "Amplitude modulation sensitivity as a mechanism for increment detection," J. Acoust. Soc. Am. **119**, 3919–3930.
Hafter, E. R., Bonnel, A. M., and Gallun, E. (**1997**). "A role for memory in divided attention between two independent stimuli," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 228–237.
Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.
Harris, J. D. (**1963**). "Loudness discrimination," J. Speech Hear. Disord. Monogr. Suppl. **11**, 1–59.
Harris, D. M., and Dallos, P. (**1979**). "Forward masking of auditory nerve fiber responses," J. Neurophysiol. **42**, 1083–1107.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001a**). "Auditory-nerve model for predicting performance limits of normal and impaired listeners," ARLO **2**, 91–96.
Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001b**). "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," Neural Comput. **13**, 2273–2316.
Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.
Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.
Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.
Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.
Langner, G., and Schreiner, C. E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol. **60**, 1799–1822.
Lee, J., and Bacon, S. P. (**1997**). "Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration," J. Acoust. Soc. Am. **101**, 3688–3693.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liberman, M. C. (**1978**). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. **63**, 442–455.
Lorenzi, C., Simpson, M. I. G., Millman, R. E., Griffiths, T. D., Woods, W. P., Rees, A., and Green, G. G. (**2001**). "Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers," J. Acoust. Soc. Am. **110**, 2470–2478.
Lüscher, E., and Zwislocki, J. (**1949**). "Adaptation of the ear to sound stimuli," J. Acoust. Soc. Am. **21**, 135–139.
Macmillan, N. A. (**1971**). "Detection and recognition of increments and decrements in auditory intensity," Percept. Psychophys. **10**, 233–238.
Moore, B. C. J., Alcantara, J. I., and Dau, T. (**1998**). "Masking patterns for sinusoidal and narrowband noise maskers," J. Acoust. Soc. Am. **104**, 1023–1038.
Moore, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). "Modulation masking produced by beating modulators," J. Acoust. Soc. Am. **106**, 938–945.
Nelson, P. B., and Jin, S. H. (**2004**). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **115**, 2286–2294.
Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Nelson, P. C., and Carney, L. H. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.
Nelson, P. C., and Carney, L. H. (**2006**). "Cues for masked amplitude-modulation detection," J. Acoust. Soc. Am. **120**, 978–990.
Nelson, P. C., and Carney, L. H., (**2007**). "Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus," J. Neurophysiol. **97**, 522-539.
Schooneveldt, G. P., and Moore, B. C. J. (**1989**). "Comodulation masking release (CMR) as a function of masker bandwidth, modulator bandwidth, and signal duration," J. Acoust. Soc. Am. **85**, 273–281.
Sheft, S., and Yost, W. A. (**1990**). "Temporal integration in amplitude-modulation detection," J. Acoust. Soc. Am. **88**, 796–805.
Shofner, W. P., Sheft, S., and Guzman, S. J. (**1996**). "Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes," J. Acoust. Soc. Am. **99**, 3592–3605.
Siebert, W. M. (**1965**). "Some implications of the stochastic behavior of primary auditory neurons," Kybernetik **2**, 206–215.
Siebert, W. M. (**1968**). "Stimulus transformations in the peripheral auditory system," in *Recognizing Patterns*, edited by P. A. Kolers (MIT, Cambridge, MA), pp. 104–133.
Smith, R. L., and Brachman, M. L. (**1982**). "Adaptation in auditory-nerve fibers: A revised model," Biol. Cybern. **44**, 107–120.
Smith, R. L., and Zwislocki, J. J. (**1975**). "Short-term adaptation and incremental responses in single auditory-nerve fibers," Biol. Cybern. **17**, 169–182.
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.
Strickland, E. A., and Viemeister, N. F. (**1996**). "Cues for discrimination of envelopes," J. Acoust. Soc. Am. **99**, 3638–3646.
Ulanovsky, N., Las, L., Farkas, D., and Nelken, I. (**2004**). "Multiple time scales of adaptation in auditory cortex neurons," J. Neurosci. **24**, 10440–10453.
Verhey, J. L., Dau, T., and Kollmeier, B. (**1999**). "Within-channel cues in comodulation masking release (CMR): Experiments and model predictions using a modulation-filterbank model," J. Acoust. Soc. Am. **106**, 2733–2745.
Verhey, J. L., Pressnitzer, D., and Winter, I. M. (**2003**). "The psychophysics and physiology of comodulation masking release," Exp. Brain Res. **153**, 405–417.
Viemeister, N. F. (**1972**). "Intensity discrimination of pulsed sinusoids: The effects of filtered noise," J. Acoust. Soc. Am. **51**, 1265–1269.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Viemeister, N. F. (**1988**). "Intensity coding and the dynamic range problem," Hear. Res. **34**, 267–274.
Viemeister, N. F., and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84**, 172–178.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Wakefield, G. H., and Viemeister, N. F. (**1990**). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," J. Acoust. Soc. Am. **88**, 1367–1373.
Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.
Winslow, R. L., and Sachs, M. B. (**1988**). "Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle," Hear. Res. **35**, 165–189.
Winter, I. M., and Palmer, A. R. (**1991**). "Intensity coding in low-frequency auditory-nerve fibers of the guinea pig," J. Acoust. Soc. Am. **90**, 1958–1967.
Wojtczak, M., and Viemeister, N. F. (**1999**). "Intensity discrimination and detection of amplitude modulation," J. Acoust. Soc. Am. **106**, 1917–1924.
Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.
Yost, W. A., and Sheft, S. (**1997**). "Temporal modulation transfer functions for tonal stimuli: Gated versus continuous conditions," Aud. Neurosci. **3**, 401–414.
Young, E. D., and Barta, P. E. (**1986**). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," J. Acoust. Soc. Am. **79**, 426–442.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," J. Acoust. Soc. Am. **109**, 648–670.

Lateralization discrimination of interaural time delays in four-pulse sequences in electric and acoustic hearing[a)]

Bernhard Laback[b)] and Piotr Majdak
Acoustics Research Institute, Austrian Academy of Sciences, Reichsratsstrasse 17, A-1010 Vienna, Austria

Wolf-Dieter Baumgartner
ENT-Department, Vienna University Hospital, Währinger Gürtel 18-20, A-1097 Vienna, Austria

(Received 9 May 2006; revised 18 January 2007; accepted 18 January 2007)

This study examined the sensitivity of four cochlear implant (CI) listeners to interaural time difference (ITD) in different portions of four-pulse sequences in lateralization discrimination. ITD was present either in all the pulses (referred to as condition Wave), the two middle pulses (Ongoing), the first pulse (Onset), the last pulse (Offset), or both the first and last pulse (Gating). All ITD conditions were tested at different pulse rates (100, 200, 400, and 800 pulses/s pps). Also, five normal hearing (NH) subjects were tested, listening to an acoustic simulation of CI stimulation. All CI and NH listeners were sensitive in condition Gating at all pulse rates for which they showed sensitivity in condition Wave. The sensitivity in condition Onset increased with the pulse rate for three CI listeners as well as for all NH listeners. The performance in condition Ongoing varied over the subjects. One CI listener showed sensitivity up to 800 pps, two up to 400 pps, and one at 100 pps only. The group of NH listeners showed sensitivity up to 200 pps. The result that CI listeners detect ITD from the middle pulses of short trains indicates the relevance of fine timing of stimulation pulses in lateralization and therefore in CI stimulation strategies. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2642280]

PACS number(s): 43.66.Pn, 43.66.Ts, 43.66.Mk [AJO] Pages: 2182–2191

I. INTRODUCTION

The growing number of bilateral cochlear implantations has raised interest in studies investigating the sensitivity of bilateral cochlear implant (CI) listeners to binaural cues. In particular, the sensitivity to interaural time difference (ITD) in electric hearing has been a subject of interest (see below). ITD is an important cue for the localization of sound sources in the left/right dimension (e.g., Macpherson and Middlebrooks, 2002), for binaural unmasking of speech in noise (e.g., Bronkhorst and Plomp, 1988), and for the perceptual segregation of competing speech sounds (Drennan *et al.*, 2003).

Several psychophysical studies have investigated ITD perception in CI listeners (van Hoesel *et al.*, 1997, 2002, 2003; Lawson *et al.*, 1998; Lawson *et al.*, 2001; Long *et al.*, 2003; Wolford *et al.*, 2003; Laback *et al.*, 2004; Senn *et al.*, 2005; Majdak *et al.*, 2006). They showed that CI listeners are sensitive to ITD, although there is a large interindividual variability in performance. These studies (except for the last three cited) used, besides other stimuli, unmodulated, rectangularly gated pulse trains as stimuli. The results obtained with this type of stimulus do not reveal to what extent listeners exploit ITD information in the ongoing signal as opposed to the information in the gating portions (onset and offset) of the stimulus. Motivated by those studies, the present study addressed the question if CI listeners are sensitive to ITD information presented either in the ongoing or the gating portions of a pulse train, using a lateralization discrimination task. For that purpose, a specific stimulus was chosen that allowed the strict separation and independent control of ITD in the ongoing and in the gating portions. The stimulus consists of a sequence of four pulses with constant amplitude, in which the first and last pulse represent the gating portions, and the two pulses in the middle represent the ongoing portion. In different experimental conditions, ITD information was present either in the ongoing signal, onset, offset, both onset and offset, or in the entire pulse train (to be called waveform ITD). It should be mentioned that while this stimulus has advantages, as will be described later, its short duration may complicate the generalization of the outcomes to longer durations.

For an unmodulated electrical pulse train, ongoing ITD is present solely in the fine timing of the individual pulses, which can be referred to as the "fine structure,"[1] a term used in the psychoacoustic literature to define the rapidly varying carrier frequency of an acoustic waveform. Thus, in this case, the lateralization discrimination performance for ongoing ITD is a measure of fine structure ITD sensitivity.

It is known from the normal hearing literature that the relative importance of ongoing and gating ITD depends on the rate of the stimulus. Ongoing ITD is the major lateralization cue as long as the frequency of the signal component conveying the ITD, the carrier or the envelope, does not exceed a certain limit. In the case of carrier ITD, the frequency limit is around 1500 Hz (Klumpp and Eady, 1956;

[a)]Parts of this work were presented at the 28th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology in February, 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: bernhard.laback@oeaw.ac.at

 0001-4966/2007/121(4)/2182/10/$23.00

TABLE I. Etiology of the four CI listeners completing the experiments (CI1, CI3, CI8, and CI12). Also included are the data of the four CI listeners who showed too poor sensitivity in the baseline condition (four pulses; 100 pps; waveform ITD of 600 μs) and thus were not included for participation in the experiments. The rightmost columns show the lateralization discrimination scores for the baseline condition and the 300 ms version of the baseline condition.

Subject	Participating in the experiments	Etiology	Age (yr)	Age at implantation (yr)		Duration of desfness		Binaural electrical stimulation experience	Performance for baseline condition (In % correct)	Performance for 300 ms version of baseline condition (In % correct)
				L	R	L	R			
CI1	Yes	Meningitis	20	14	14	5.5 mo	1.5 mo	6 yr	80.0	98.3
CI3	Yes	Meningitis	21	21	21	2 mo	2 mo	1 mo	96.0	99.0
CI8	Yes	Otosclerosis	41	41	39	3 yr	12 yr	2 mo	73.0	77.0
CI12	Yes	Sudden hearing loss	40	35	34	8 yr	3 yr	5 yr	95.0	99.0
CI2	No	Skull trauma	58	54	48	21 yr	25 yr	4 yr	58.3	70.0
CI6	No	Progressive	42	41	39	8 yr	8 yr	1 yr	60.0	65.0
CI5	No	Otosclerosis	44	35	42	2 yr	9 yr	2 yr	59.4	73.7
CI9	No	Progressive	58	50	51	5 yr	5 yr	7 yr	60.5	75.9

Zwislocki and Feldmann, 1956; Boerger, 1965). In the case of envelope ITD imposed on a high frequency carrier, the frequency limit appears to be lower, depending on the temporal characteristics of the stimulus (Henning, 1974; Hafter *et al.* 1983; Bernstein and Trahiotis, 1994, 2002). ITD in the gating portions, in particular the onset, is more influential at higher signal frequencies (Saberi and Perrott, 1995), particularly in case of ambiguous ongoing ITD cues (Freyman *et al.*, 1997).

To examine these dependencies in electrical hearing, the current study examined lateralization discrimination for each of the ITD conditions described above as a function of pulse rate. The measurements obtained from the CI listeners were complemented by measurements on normal hearing (NH) subjects who listened to an acoustic simulation of electric stimulation. Previous studies have shown that some aspects of temporal pitch perception of NH subjects listening to such an acoustic simulation correspond to the perception of CI listeners (McKay and Carlyon, 1999; Carlyon *et al.*, 2002). Since both temporal pitch perception and ITD perception are based on the temporal properties of the stimulus, the simulation technique used in the cited studies may also mimic some aspects of ITD perception in electric hearing. However, it has to be kept in mind that NH subjects listening to the simulation can discriminate rate pitch up to much higher pulse rates than CI listeners.

Previous studies on NH listeners demonstrated that ITD information in the temporal fine structure is important for the lateralization of sound sources (Wightman and Kistler, 1992; Smith *et al.*, 2002) and for speech perception in noise (Nie *et al.*, 2005; Zeng *et al.*, 2005). If CI listeners were found to be sensitive to ITD in the ongoing signal (and thus the fine structure), then encoding of fine structure cues in future CI stimulation strategies would appear to be a promising approach for improving the ability to lateralize sound sources and to comprehend speech in noise. A better understanding of the particular contribution of ITD information in the ongoing fine structure and in the gating portions as a function of pulse rate could help to improve stimulation strategies for cochlear implants, aiming to maximize the transfer of ITD information.

II. METHOD

A. Subjects and implant system

Four postlingually deafened CI listeners, implanted bilaterally at the Vienna University Hospital (CI1, CI3, and CI8) and at the University Clinic Würzburg (CI12), were tested. They were selected from a total of eight CI listeners invited for participation in the study. These four listeners fulfilled the selection criterion, as defined by the ability to reproducibly perform left/right discrimination on the basis of 600 μs waveform ITD in a sequence of four pulses at a pulse rate of 100 pps. The remaining four listeners showed very low discrimination scores for this baseline condition, even after a full day of training. Table I shows, for all eight listeners invited, the percent correct scores achieved for the baseline condition in a final test at the end of the training. Also included is the performance for a 300 ms version of the same stimulus, to allow comparison with other studies using this stimulus duration. These scores are based on at least 180 item repetitions. Table I also contains data on the patients' etiology. The data of the patients not participating in the experiments are included in the table to make them available for future analysis.

All of the implantees had been supplied with the C40+ system by MED-EL Corp. It generates nonsimultaneous biphasic current pulses (cathodic phase first) on up to 12 electrodes (2.4 mm spacing). It provides stimulation in monopolar configuration with an extracochlear ground electrode. The electrodes are numbered in ascending order from apical to basal position in the cochlea.

All four listeners participating in the tests had normal hearing before the onset of deafness. The duration of binaural deafness, i.e., the time period between the beginning of deafness at the first ear and the activation of the second CI, was two months (CI3), 5.5 months (CI1), eight years (CI12), and 12 years (CI8). Subject CI3 was supplied with CIs in one operation at both ears. Subjects CI1, CI12, and CI8 were successively supplied with CIs at the two ears with a temporal gap of four months, one year, and two years, respectively. The binaural electric stimulation experience ranged from one month to six years.[2]

TABLE II. Electric stimulation parameters. The electrodes are numbered from apex to base in ascending order. With respect to the parameter "right electrode higher" in percent, and value ≤24 or ≥76 indicates significant pitch discriminability ($p<0.01$). The stimulation levels are specified in μA.

Subject	Test electrodes L/R	Right electrode higher (in %)	Current levels (in μA) used in exp. I and II L/R
CI1	4/1	50.0	261/248
CI3	4/3	42.3	265/283
CI8	7/5	45.0	376/358
CI12	2/2	53.8	547/601

Five normal hearing subjects, aged 25–35 years old, participated in this study. None of them had any indication of present or past hearing disorder. Two of the subjects were the first two authors of this study (NH2, NH4). All except one NH subject (NH6) had previous experience with psychoacoustic experiments.

B. Apparatus for electric and acoustic stimulation

A personal computer system was used to control electric and acoustic stimulation. Each implant was controlled by a Research Interface Box (RIB), manufactured at the University of Innsbruck, Austria. The two RIBs were synchronized, providing an interaural accuracy of stimulation timing of 2.5 μs. Both RIBs were connected to the personal computer system via serial interfaces. Prior to the experiment, the stimuli were verified using a pair of dummy implants (Detektorbox, MED-EL), connected to a two-channel storage oscilloscope (softDSP, SDS 200).

The stimuli for acoustic stimulation were output via a 24 bit stereo analog-to-digital, digital-to-analog (A/D-D/A) converter (ADDA 2402, Digital Audio Denmark) using a sampling rate of 96 kHz per channel. The converter received the data from the computer via a digital audio interface (DIGI 96/8 PRO, RME). The analog signals were sent through a headphone amplifier (HB6, TDT) and an attenuator (PA4, TDT) and presented to the subjects via a circumaural headphone (K501, AKG). Calibration of the headphone signals was performed using a sound level meter (2260, Brüel & Bjær) connected to an artificial ear (4153, Bruel & Kjær). The headphone signals were further inspected by digital signal analysis software after digitizing them through the A/D-D/A converter.

C. Stimuli

The electric stimuli were trains of four biphasic current pulses with constant amplitude and a phase duration of 26.7 μs. The current level of the stimuli was adjusted to a comfortable loudness (see Table II). The pulse rate varied between 100 and 800 pps.

The acoustic stimuli used to simulate electric stimulation at a single electrode were similar to those used by McKay and Carlyon (1999) and Carlyon *et al.* (2002). Monophasic pulses with a duration of 10 μs were passed through a bandpass filter with −3 dB cutoff frequencies at 3900 and 5400 Hz. The filter was designed as a digital eighth-order Butterworth filter, with slopes of 48 dB/octave. The bandwidth of the filter was broad enough to preserve the modulation in the stimuli. The stimuli were gated with the halves of a Hamming window (duration: 0.6 ms) to avoid truncation of the impulse response, which could cause detection of transient cues. The presentation level was set so that a continuously presented pulse train with a pulse rate of 100 pps yielded a rms value of 78 dB sound pressure level (SPL).

The center frequency of the filter (4590 Hz) was a compromise between two effects. First, the auditory filter bandwidth increases with center frequency; thus, the smearing effect of the auditory filters on the temporal envelope of the stimulus decreases with increasing center frequency. This favors the choice of a high frequency. Second, the sensitivity to ITD in amplitude modulated tones decreases for carrier frequencies exceeding 4–6 kHz (Bernstein and Trahiotis, 2002; Henning, 1974). This favors the choice of a low frequency.

The method of simulating electrical stimulation in acoustical hearing by using bandpass filtered pulse trains implies that each pulse corresponds to the impulse response of the bandpass filter used. Thus, each acoustic "pulse" is a complex waveform, having a fine structure and an envelope. Although this differs from electrical pulses which have no fine structure of their own, we consider the bandpass filtered pulse trains as an appropriate simulation of electrical stimulation. This method assumes that information in the envelope of bandpass filtered acoustic pulse trains is analogous to information in the "fine structure" of electric pulse trains, even though the fine structure is not effectively represented in the neural response to high acoustic frequencies.

In acoustical stimulation, interaurally uncorrelated pink noise signals (50–10,050 Hz) were presented continuously at both ears to mask signals outside the desired frequency band. The spectrum level at 4.6 kHz was 15.2 dB SPL. The noises were generated in real time and mixed with the pulse trains.

Using a continuous background noise in acoustic stimulation may result in an overall decrease of ITD sensitivity. However, ITD sensitivity measurements with the five NH listeners using 300 ms pulse trains[3] revealed mean 80% just noticeable differences (JNDs) of as low as 40 μs (SD: 5 μs), which is quite close to the minimum detectable ITD observed in normal hearing. Thus, the noise can have had only a very small effect on ITD sensitivity.

III. PRETESTS

Each CI listener completed a series of pretests in order to locate an interaural pair of electrodes eliciting the same pitch percept. These pretests were performed using electric pulse trains with a pulse rate of 100 pps and a duration of 300 ms.

Interaural electrodes with similar pitches seem to be more likely to show ITD sensitivity, although the effect of increasing the interaural place difference can be small (van Hoesel, 2004; Wolford *et al.*, 2003; Long *et al.*, 2003). The procedure to find a pitch-matched electrode pair involved the following steps: (a) determination of electric dynamic range and comfortable level for electrodes 1–8 on each ear; (b) estimation of monaural pitch sensation across the electrode arrays to determine pitch-matched interaural electrode pair candidates; (c) interaural loudness balancing for each interaurally pitch-matched pair candidate; and, (d) measurement of pitch discriminability for each interaurally pitch-matched pair candidate and final selection of one pitch-matched pair. More details on the methods can be found in Majdak *et al.* (2006).

For all CI listeners at least two pitch-matched electrode pairs ($p<0.01$) were identified. Table II indicates, for each subject, the electrode pair members finally selected for presenting stimuli in the ITD studies and the corresponding percentage of trials in which the right electrode was judged to be higher in pitch.

The pretests also served to determine comfortable and interaurally loudness-balanced levels of the four-pulse stimuli to be used in the experiments.

IV. EXPERIMENT I: LATERALIZATION DISCRIMINATION

This experiment studied the effects of ITD in different signal portions in lateralization discrimination as a function of pulse rate.

A. Procedure

ITD sensitivity was measured using a "lateralization discrimination" task requiring left/right judgments of a target relative to a comparison stimulus. The first interval contained a comparison stimulus with zero ITD, evoking a centralized auditory image. The second interval contained the target, which differed from the comparison stimulus in that pulses at one ear were delayed relative to the other ear. The subjects were requested to indicate whether the second stimulus was perceived to the left or to the right of the first stimulus by pressing the appropriate button on a response pad. The stimulus intervals were separated by a silent period of 300 ms. Visual indication of the stimulus intervals was provided on a computer screen. Visual feedback about the correctness of the response was provided after each trial. The method of constant stimuli was applied to determine the JND in ITD with respect to lateralization discrimination. Percent correct scores were collected at four to six ITD values to estimate the psychometric function. JNDs were estimated from a maximum-likelihood cumulative Gaussian fit to the percent correct data.[4] The largest ITD presented was 800 μs and depended on the sensitivity of each subject for the individual conditions. In case of ongoing ITD (the two middle pulses) at the pulse rate of 800 pps, however, the largest ITD value was restricted to 500 μs, since ongoing ITD approaching half of the interpulse interval can introduce ambiguous cues (Majdak *et al.*, 2006). Each stimulus was presented at least 60 times. In cases where the psychometric function did not exceed 66% correct (which occurred for some conditions in case of listener CI8), at least 60 further item repetitions were presented in order to reduce the randomness (noise) in the data. A completely randomized design was applied in which all levels of the independent variables and their repetitions were pooled in one "item list" which was then randomized. The subjects took a break after a block of 30 min. Depending on the constitution and motivation of the subject, 6–10 blocks were completed in one testing day. Before the start of the experiment, the subjects received training using the same procedure as in the main experiment. The training was conducted in three stages. The first stage used a 300 ms version of the baseline condition, the second stage used the baseline condition, and the third stage used a list containing all stimulus conditions of the main experiment. For each stage, the training was continued until the subjects showed stable performance. The four CI listeners fulfilling the criterion for participation required about 2 h of training. The other four CI listeners who showed poor performance were trained for one day (6–10 blocks) before it was decided to exclude them from this study.

B. Stimulus conditions

Unmodulated pulse trains, consisting of four pulses, were presented at an interaurally pitch-matched electrode pair, selected in the pretests. The rationale for using a constant number of four pulses at each pulse rate was twofold. First, the pulse amplitude could be held constant across pulse rates, thus avoiding any confounding effects of loudness. It was verified for the CI listeners and the NH listeners by an informal loudness estimation task that these stimuli elicit the same loudness at different pulse rates. Second, the number of information units (in terms of pulses) containing ITD remained constant across pulse rates, thus avoiding confounding effects of the number of pulses containing ITD information.

Figure 1 illustrates schematically the different types of ITD tested, by showing the amplitude versus time representations of the respective pulse trains at the two ears. The stimulus shown on the top of the figure represents the comparison condition having zero ITD. The stimulus shown beneath, referred to as waveform delay (Wave), contains ITD in each of the interaural pulse pairs. In the condition containing delay in the ongoing signal (Ongoing) the two pulse pairs in the middle of the train contain ITD, whereas the first and last pulse pairs have a zero ITD. In the gating delay condition (Gating) the first and last pulse pairs contain ITD, whereas the two pairs in the middle have zero ITD. The conditions of onset delay (Onset) and offset delay (Offset) contain ITD in the first (onset) pulse pair only and in the last (offset) pulse pair only, respectively. The ITD is always di-

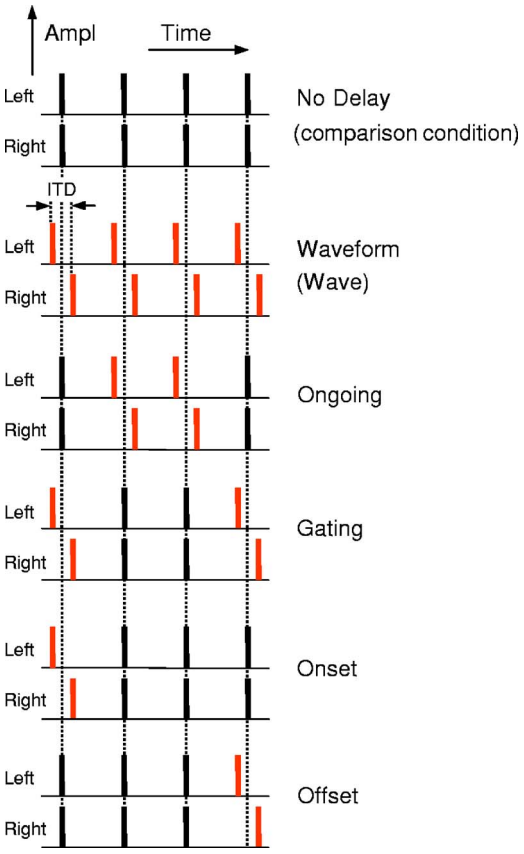


FIG. 1. (Color online) Schematic illustration of the different ITD types: The amplitude vs time representations of the pulse trains at the two ears are shown. Note that electric pulses were actually biphasic and acoustic pulses were monophasic (as shown). In Experiment II, only one channel (left or right) was presented.

vided between the two ears, i.e., the leading pulse already starts at half the ITD before the reference position at one ear and the lagging pulse starts at half the ITD after the reference position at the opposite ear. This was done to reduce the temporal irregularity, which is a potential monaural discrimination cue.

All ITD types were tested at pulse rates of 100, 200, 400, and 800 pps. Hence, the duration of the stimuli ranged from 40 ms (at 100 pps) to 5 ms (at 800 pps). CI listener CI8 was not tested with the pulse rate of 200 pps.

C. Results

All subjects reported hearing fused images for all of the conditions tested. Inspection of the distribution of the left/right judgments for each listener revealed sufficient symmetry. Therefore, an adjustment of the percent correct scores to remove response bias was not required. The CI listeners show large inter-individual differences in the overall lateralization discrimination performance. In order to allow the determination of JNDs for all subjects, a threshold criterion of 65% was used. For the conditions revealing sensitivity at the defined threshold criterion, the psychometric functions are monotonic.

1. CI listeners

Figure 2 shows the JNDs in μs for each of the four CI listeners derived from the lateralization discrimination scores, as a function of pulse rate, for the various ITD types. Error bars indicate the 95% confidence intervals.[5] The significance of the difference between two JNDs was evaluated using a test based on Monte Carlo simulations of the fits to the underlying psychometric functions.[4] Note that some overlap of 95% confidence intervals for two conditions does not preclude a significant difference between the mean values. JNDs which could not be determined at the specified threshold criterion are marked as ND.

The JNDs for the reference condition Wave increase with the pulse rate in case of listeners CI1 (significant difference between JNDs at 100 and 800 pps: $p=0.003$) and CI8 (JND at 100 pps: of 398 μs; JND at 800 pps: undeterminable). In case of CI3 and CI12, the JNDs are approximately constant across pulse rates.

All CI listeners are sensitive to gating ITD at all pulse rates for which sensitivity to waveform ITD was observed. The JNDs are constant across different pulse rates. The apparent decrease of JNDs with increasing rate for listener CI3 is not statistically significant.

For all CI listeners and at the lowest pulse rate (100 pps), the JNDs for the conditions Ongoing and Gating are larger than those for the condition Wave. These differences are significant for all listeners (largest p value: 0.043), except for CI8. The finding that omission of ITD in either the ongoing or the gating signal portion causes degradation in performance relative to condition Wave implies that both ongoing and gating ITD contribute to lateralization discrimination.

With increasing pulse rate, the CI listeners differ with respect to their sensitivity to ongoing ITD. Listener CI3 shows sensitivity up to 800 pps.[6] Listeners CI8 and CI12 show sensitivity up to 400 pps, and listener CI1 shows sensitivity at 100 pps only.

Three CI listeners (CI1, CI3, and CI12) reveal sensitivity to onset ITD. They also show increasing sensitivity with increasing pulse rate. The significance of this effect for each of the three listeners CI1, CI3, and CI12 is revealed by the significant differences between the JNDs at 100 and 800 pps ($p=0.0001$, 0.04, and 0.034, respectively).

The JNDs for condition Offset are determinable at the pulse rates of 100 and 200 pps only (CI3), at the pulse rate of 100 pps only (CI1 and CI12), and at neither pulse rate (CI8).

2. NH listeners

The five NH listeners showed homogeneous effects, and therefore Fig. 3 shows their mean JNDs. The error bars indicate ±1 standard deviation of the mean values across the listeners. The left panel of Fig. 3 plots the JNDs determined using the 65% threshold criterion. For comparison, the right panel shows the JNDs determined for the same data but using the 80% threshold criterion. Note the different scaling of the ordinates in the two plots. Comparison of JNDs obtained for the two threshold criteria reveals similar effects of the

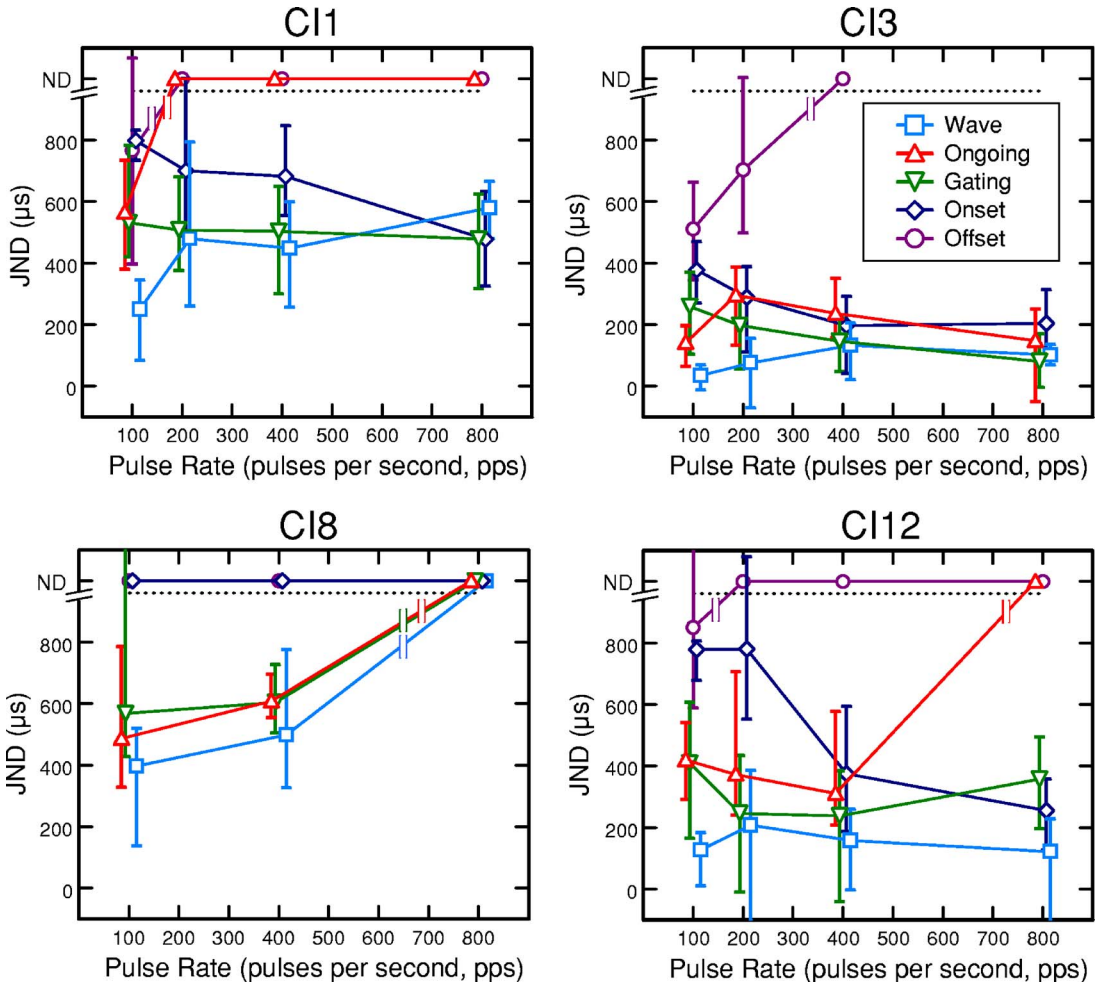


FIG. 2. (Color online) JNDs as a function of pulse rate obtained from Experiment I for each of the CI listeners. The parameter is the type of ITD, as depicted in Fig. 1. Error bars indicate the bootstrap 95% confidence intervals. Cases where a JND was not determinable at the specified percent correct point are plotted at the top of the figure, marked with ND. For better visual separation of the data points, different conditions are horizontally shifted relative to each other by a small amount. For some conditions the error bars are smaller than the symbols.

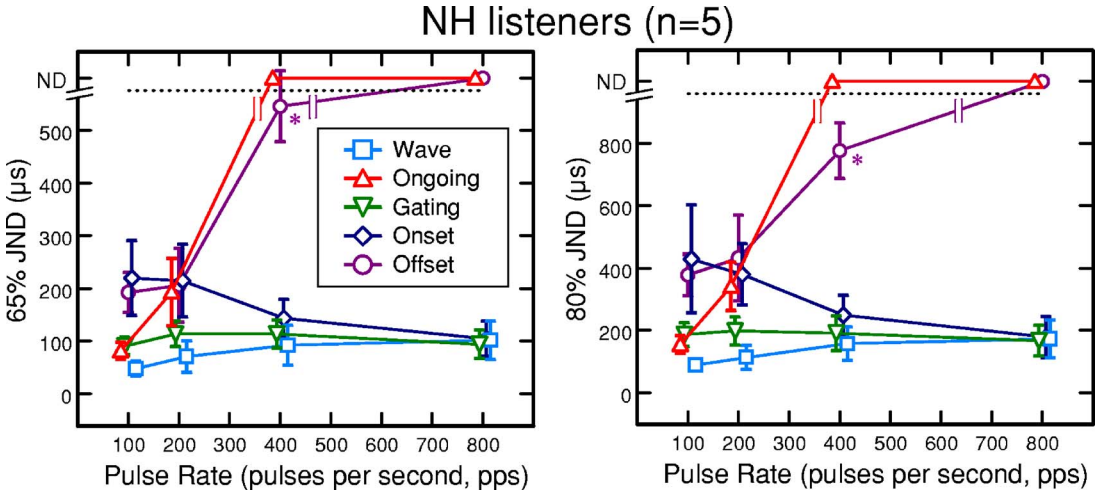


FIG. 3. (Color online) The mean JNDs of five NH listeners are shown. The left panel plots the JNDs determined using the 65% threshold criterion. For comparison, the right panel shows the JNDs using the 80% criterion. Error bars indicate ±1 standard deviation of the mean. Note the different scaling of the ordinates in the two panels. The JND for condition Offset at 400 pps (marked with an asterisk) is based on two listeners for which a JND could be determined. All other conventions are as in Fig. 2.

stimulus conditions. This suggests that the choice of the low criterion of 65% had no major effect on the outcomes.

The significance of differences in JNDs between different test conditions was tested by two-tailed *t* tests. The performance in condition Wave decreases with the pulse rate. This is reflected by the significantly higher JNDs at 800 pps than at 100 pps ($p=0.008$).

The listeners show sensitivity in condition Gating at all pulse rates. The performance is constant across different pulse rates (JNDs at 100 pps vs 800 pps: $p=0.64$).

At the lowest pulse rate (100 pps), the mean JNDs for the conditions Ongoing (82 μs) and Gating (91 μs) are higher than those for condition Wave (48 μs). The differences Ongoing vs Wave and Gating vs Wave are significant ($p=0.01$ and 0.001, respectively). In addition, the mean JNDs for conditions Onset (220 μs) and Offset (193 μs) are significantly higher than those for condition Gating (91 μs) ($p=0.02$ and 0.01, respectively). Taken together, the results at 100 pps indicate perceptual contribution of each interaural pulse pair.

With increasing pulse rate, the sensitivity to ongoing ITD decreases monotonically. The highest rate showing determinable JNDs is 200 pps in four listeners and 400 pps in one listener. The JND determinable for that one listener at 400 pps amounts to 486 μs.

The sensitivity to onset ITD increases monotonically with the pulse rate. The JNDs at 100 and 800 pps differ significantly from each other ($p=0.007$). At 800 pps, the JND in condition Onset (104 μs) is similar to the conditions Gating (94 μs) and Wave (102 μs) ($p=0.22$ and 0.84, respectively). In contrast, the sensitivity to offset ITD decreases with the pulse rate. At 400 pps, the JNDs for condition Offset are high for two of the listeners (mean value: 547 μs) and undeterminable for the other three listeners. Figure 3 shows the mean JND of the two listeners for whom a JND could be determined. At 800 pps, they are undeterminable for all listeners. Comparing the JNDs for conditions Onset and Offset with those for condition Gating reveals complementary contributions of onset and offset ITD as a function of pulse rate.

D. Discussion

1. CI listeners

Three main effects can be observed from the data collected in Experiment I with the four CI listeners, despite considerable inter-individual differences. First, gating ITD contributes to lateralization discrimination at all pulse rates for which sensitivity to waveform ITD was observed. Second, the three CI listeners revealing sensitivity to onset ITD show increasing contribution of onset ITD with increasing pulse rate. Third, and most importantly, all subjects show sensitivity to ITD in the two pulses in the middle of the train, the condition referred to as Ongoing: one listener up to 800 pps, two listeners up to 400 pps, and one listener at 100 pps only. Sensitivity in condition Ongoing implies that ITD in the temporal fine structure of the short pulse trains contributes to lateralization discrimination. These stimuli present "pure" fine structure ITD, containing no other ITD cues in the onset, offset, or ongoing envelope. Majdak *et al.* (2006) studied fine structure ITD sensitivity in four CI listeners using amplitude modulated pulse trains with a 300 ms duration. In that study, fine structure ITD was created by delaying the pulses at one ear and subsequently applying a trapezoidal envelope with zero ITD. This could potentially involve confounding effects of conflicting ITD cues at the onset. Namely, the first audible pulse could be that on the lagging side, since the first pulse on the leading side was at the absolute threshold. However, the finding of high fine-structure-ITD sensitivity up to 800 pps in that study indicates that conflicting ITD cues during the onset played a minor role.

The conditions Ongoing and Wave at 800 pps may have introduced ambiguous cues for ITD values approaching half of the interpulse interval. This would be reflected by a decline of the psychometric function in that ITD range. In fact, these psychometric functions showed no such decline at the largest ITDs tested, which were 500 and 600 μs in conditions Ongoing and Wave, respectively. Thus, it is unlikely that ambiguity had a strong influence. This is in contrast to the results of Majdak *et al.* (2006) who observed a pronounced decline of the psychometric functions for ITDs between approximately one quarter and one half of the interpulse period, using 300 ms pulse trains with ITD in the fine structure only. The absence of nonmonotonicity of the psychometric functions in the current study could be due to the short duration of the stimuli, assuming that for short stimuli the onset is more dominant in resolving the ambiguity than for longer stimuli.

For the listeners CI1, CI3, and CI12, the sensitivity to gating ITD is qualitatively consistent at the different pulse rates with the corresponding contributions of onset and offset ITD. This leads to the assumption that in the condition Gating the contributions of onset and offset ITD are combined. For CI8, however, gating ITD was found to contribute despite undeterminable JNDs in conditions Onset and Offset. This may indicate that either onset or offset ITD alone are too weak to be evaluated.

2. NH listeners

For the NH listeners, the decrease in sensitivity in condition Ongoing with increasing pulse rate is qualitatively consistent with studies from the NH literature. Hafter *et al.* (1983) and Saberi (1996) investigated the effect of pulse rate on the perceptual contribution of the ongoing signal using bandpass filtered clicks. These studies observed decreasing contribution of ITD in the clicks following the onset for rates exceeding 200 clicks/s. The maximum pulse rate where all NH listeners of the current study were sensitive to ITD in the two pulses in the middle of the train (condition Ongoing) was 200 pps. This is qualitatively comparable with the results of a study by Bernstein and Trahiotis (2002). They found sensitivity to ongoing ITD for modulation frequencies up to 256 Hz, using sinusoidally amplitude-modulated and "transposed" 4 kHz tones.

The high sensitivity to onset ITD at 400 pps and the complete dominance of onset ITD observed at 800 pps in the current study is qualitatively consistent with the results of

Saberi and Perrott (1995) and Freyman *et al.* (1997). They showed that click trains with rates equal to or greater than 500 pps are lateralized toward the ear favored by ITD in the onset click even if the remaining clicks in the stimulus have an ITD favoring the opposite ear.

The performance in condition Gating was found to be constant across different pulse rates. This seems to be an effect of the complementary contributions of onset and offset ITD as a function of pulse rate, as has also been observed for three of the CI listeners.

V. EXPERIMENT II: MONAURAL DETECTION

This experiment was intended to verify that the lateralization judgments obtained in Experiment I were based on binaural information rather than on monaural cues such as periodicity pitch or timbre. Such cues could theoretically have been exploited by the listeners in case of all ITD types besides Wave. Condition Ongoing, for example, involves a change in the interpulse interval from the first to the second and from the second to the third (and last) interpulse interval. The experiment tested if the subjects exceed chance performance in detecting monaural versions of the stimuli used in Experiment I. If the subjects do not exceed chance performance then the performance in Experiment I was not based on monaural cues. In case the subjects exceed chance performance, the data obtained in Experiment I could have been based on monaural cues.

A. Method

A three-interval, two-alternative forced-choice procedure (oddity task) was used in which the listener had to judge which stimulus was different from the other two. Visual feedback about the correctness of the response was provided after each trial. The "odd" stimulus was a monaural version of the target stimulus used for lateralization discrimination (having an irregular interpulse interval). The comparison stimulus was a monaural version of the respective reference stimulus (having a regular interpulse interval). The magnitude of the deviation from the regular interpulse interval, chosen for each condition, corresponded to half the value of the ITD tested in Experiment I which was just above the lateralization JND. All combinations of pulse rates and ITD types (except for Wave) of Experiment I were tested. A total of 36 stimulus presentations were applied for each condition. The stimuli were presented in completely randomized order in a single test session. A training phase with 20 presentations of each stimulus condition was completed before collecting data. All other aspects of the method and stimuli were identical as in Experiment I.

B. Results

The highest performance achieved by each of the listeners for all of the conditions was: CH: 47.2%, CI3: 39%, CI8: 44.4%, NH2: 47.2%, NH3: 47.2%, NH4:36%, NH5: 42%, NH6: 47.2%. In all cases the performance fell within the range of chance rating ($p > 0.05$). Thus, it was concluded that the lateralization discrimination thresholds obtained in Experiment I were entirely based on binaural cues.

VI. GENERAL DISCUSSION AND CONCLUSIONS

This study investigated the contribution of ITD in various portions of short unmodulated sequences of four pulses in lateralization discrimination as a function of pulse rate. All four cochlear implant listeners were sensitive, at least at low rates, to ITD in the signal portion referred to as ongoing signal, consisting of the two pulses in the middle of the train. Thus, they were sensitive to ITD in the temporal fine structure of the pulse sequences. However, the listeners differed greatly with respect to the highest pulse rate showing sensitivity (100 pps in one listener, 400 pps in two listeners, and 800 pps in the fourth listener). Furthermore, all listeners were found to be sensitive to ITD in the gating pulses at all pulse rates for which they showed sensitivity to waveform ITD. For the three listeners revealing sensitivity to onset ITD, its contribution was shown to increase with the pulse rate.

Because only four CI listeners participated in the experiments, no general conclusions can be drawn for the population of bilateral CI listeners. The data should rather be considered as case studies. As already mentioned, four of the eight CI listeners invited for participation in the tests showed low and unstable sensitivity for a baseline condition and were therefore not included for further participation in the experiments.

The stimuli were designed to avoid the influence of confounding parameters such as differences in the number of pulses and the amplitude in the comparison across different pulse rates. By using the same number of four pulses and a constant amplitude, the same information units were presented at each pulse rate, which facilitates the comparison across pulse rates.

It is also important to keep in mind that the four-pulse sequences used in this study are quite short, in particular at the higher pulse rates. It is possible that for the higher rates the neurons were in refractory state immediately after the onset pulse and thus did not respond to the two pulses in the middle presenting "ongoing ITD." In everyday situations, sustained timing cues are likely to continue for more than just two electrical pulses, and thus the weighting of different types of ITD cues may differ.

To check the outcomes for longer stimuli with a constant duration across pulse rates, additional data have been collected. The stimuli had the same duration of 300 ms at all pulse rates (100, 400, and 800 pps). The stimulus amplitude at 100 pps was the same as in Experiment I and for higher pulse rates it was adjusted to elicit equal loudness. The conditions Wave, Ongoing, and Gating were tested. The methodology was the same as in Experiment I. Since this experiment was done with just one CI listener (CI3, who was available for further testing), the interpretation of the results has to be considered as preliminary. The results confirmed two main findings from Experiment I. First, the listener was able to lateralize upon ongoing ITD. Second, gating ITD contributed to lateralization discrimination at all three pulse rates tested. Furthermore, up to 400 pps, the sensitivity to ongoing ITD was higher for the 300 ms stimuli than for the four-pulse stimuli. This improvement is most likely due to

temporal integration of ITD information. At 800 pps, however, no sensitivity to ongoing ITD was observed, which is in contrast to the results for the four-pulse stimuli. This may be related to the lower amplitude compared to the four-pulse stimuli. Lowering of the amplitude was necessary to obtain equal loudness increasing the stimulus duration. In summary, these results reveal an interaction between the effects of the parameter pulse rate and the parameters pulse number and amplitude. To use a constant number of information units containing ITD at each pulse rate in Experiment I circumvented these interactions.

Individual differences in the upper rate limit for sensitivity in condition Ongoing, as observed in Experiment I, suggest that some unknown factors besides those controlled in the experiments (interaural pitch and loudness matching) can limit the perception of ITD in the two pulses following the onset in electric hearing. One potential factor could be the decay of internal excitation caused by pulsatile electric stimulation. Different studies have shown that the recovery from forward masking, which may be related to the decay of excitation, can vary considerably between CI listeners (Chatterjee, 1999; Nelson and Donaldson, 2002).

One of the CI listeners showed sensitivity in condition Ongoing up to 800 pps, whereas four out of five NH listeners showed an upper rate limit of 200 pps. This difference may be related to the specific properties of electric and acoustic stimulation. First, a limiting factor in acoustic hearing may be the critical band filtering on the basilar membrane, which is bypassed in electric hearing. Ringing of the auditory filters in acoustic stimulation effectively reduces the modulation depth. This could make it difficult to extract ITD information from the pulses following the onset at higher pulse rates.[7] Second, phase locking is known to be stronger in electric hearing than in acoustic hearing due to bypassing the synaptic mechanism at the hair cell (Abbas, 1993).

Experiment II verified that monaural cues had no influence on the lateralization discrimination scores obtained in Experiment I. It was concluded that the performance in Experiment I was entirely based on binaural cues.

The finding that temporal fine structure cues can be exploited in electric hearing for pulse rates as high as 800 pps is supported by a recent study on monaural rate discrimination by Chen and Zeng (2004), who reported the ability of three CI listeners to detect sinusoidal frequency modulation at rates of the standard stimulus up to 1000 pps. This finding differs from previous studies, which have shown that CI subjects typically cannot detect a pitch difference based on rate above 300–500 pps of the standard stimulus (e.g., Zeng, 2002). A recent study performed by the authors of the present study (Majdak *et al.*, 2006) investigated the effects of ITD in the fine structure, using amplitude-modulated pulse trains. As in the current study, large differences between individual CI listeners were observed with respect to the highest pulse rate showing effects of fine structure ITD. Van Hoesel and Tyler (2003) and van Hoesel (2004) presented, for the first time, performance measures of CI listeners tested with a new stimulation strategy designed to encode fine structure ITD cues. No clear advantage in sound source localization could be observed for the new strategy compared to conventional strategies, which discard fine structure information. More work is needed to determine the potentials of stimulation strategies encoding ITD information in the fine structure, considering the complexity of the parameters and effects involved. For example, channel interactions due to current spread may disrupt low frequency ITD cues in the fine structure. The practical conclusion from the data collected in this study is that bilateral CI listeners may benefit from encoding fine structure ITD information in future CI stimulation strategies with respect to the localization of sound sources in the left/right dimension.

VII. SUMMARY

Four bilateral cochlear implant listeners were tested on their ability in left/right discrimination on the basis of ITD in different portions of four-pulse sequences, as a function of pulse rate. ITD information was presented in the two middle pulses, in the gating portions (onset and offset pulses), or in the entire train. Furthermore, five normal hearing subjects were tested, listening to simulations of electrical stimulation.

(1) One of the CI listeners showed sensitivity to ITD in the two middle pulses up to 800 pps, two CI listeners up to 400 pps, and one CI listener up to 100 pps. Four NH listeners showed sensitivity up to 200 pps, one up to 400 pps.
(2) For all CI and NH listeners, gating ITD contributed at all pulse rates. The sensitivity to onset ITD increased with the pulse rate for three CI listeners as well as for all NH listeners.
(3) A monaural detection experiment verified that the listeners did not make use of monaural cues when performing the lateralization discrimination task.

ACKNOWLEDGMENTS

We are indebted to our test persons, in particular the CI listeners, for their patience while performing the lengthy tests. We thank MED-EL Corporation for providing the equipment for direct electric stimulation. We are grateful to Steve Greenberg and Christopher Long for fruitful discussions and to Matthew Goupell and Brian Gygi for helpful comments on an earlier version of this paper. This study was supported by the Austrian Academy of Sciences.

[1]Referring to the "fine structure" of an electrical pulse train without envelope modulation is uncommon, but appears appropriate in this context.
[2]The short binaural experience of CI listeners C13 and CI8 may be considered as a potential shortcoming. We had the opportunity to perform repeated tests with these two listeners two years and one year, respectively, after the main tests and observed no change in ITD sensitivity (see also Majdak *et al.*, 2006). It should be considered that clinical CI systems which use constant pulse rates and thus discard fine binaural timing cues provide no stable fine structure ITD cues in everyday listening. This supports our opinion that the short binaural CI experience of the listeners CI3 and CI8 did not negatively influence their ITD sensitivity.
[3]Applying waveform ITD in unmodulated pulse trains with 100 pps and using the same methods as in experiment I.
[4]Using psignifit version 2.5.41 (see http://bootstrap-software.org/psignifit/), a software package for fitting psychometric functions to psychophysical data, described in Wichmann and Hill (2001a) and Wichmann and Hill (2001b).

[5]Confidence intervals were found by the BC_a bootstrap method implemented by psignifit, based on 1999 simulations (see Wichmann and Hill, 2001b).

[6]Since the observation of sensitivity up to 800 pps was unexpected, the measurements for condition Ongoing at 100, 400, and 800 pps were repeated at another testing day revealing exactly the same results.

[7]The JNDs obtained for the NH listeners might depend on the frequency region of the bandpass filter applied on the stimulus and therefore be somewhat arbitrary. Assuming that ringing of the auditory filters is the limiting factor, lower JNDs would be expected at higher frequency regions where the impulse responses of the auditory filters are shorter. On the other hand, Bernstein and Trahiotis (2002) found that the upper rate limit in the sensitivity to ongoing ITD does not increase with the center frequency for center frequencies from 4 to 10 kHz, thus not supporting the assumption that cochlear filtering is the limiting factor. However, the stimuli used by Bernstein and Trahiotis (2002) had a constant bandwidth in Hz, thus the bandwidth in ERB decreased with the center frequency. It may be that using a constant bandwidth in the ERB scale would favor the performance at higher center frequencies relative to that at the lower frequencies because of the larger number of stimulated neurons and/or because of the larger modulation depth at the output of the cochlear filter due to the larger bandwidth. A study attempting to clarify these issues with normal hearing listeners is currently under way.

Abbas, P. J. (**1993**). "Electrophysiology," in *Cochlear Implants: Audiological Foundations*, edited by R. S. Tyler (Singular, Publishing Group Inc., San Diego).

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.

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.

Boerger, G. (**1965**). "The localization of Gaussian tones," unpublished dissertation, Technical University of Berlin.

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Chatterjee, M. (**1999**). "Temporal mechanisms underlying recovery from forward masking in multielectrode-implant listeners," J. Acoust. Soc. Am. **105**, 1853–1863.

Chen, H., and Zeng, F. G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

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.

Freyman, R. L., Zurek, P. M., Balakrishnan, U., and Chiang, Y. C. (**1997**). "Onset dominance in lateralization," J. Acoust. Soc. Am. **101**, 1649–1659.

Hafter, E. R., Dye, R. H. Jr., and Wenzel, E. (**1983**). "Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number," J. Acoust. Soc. Am. **73**, 1708–1713.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Laback, B., Pok, S. M., Baumgartner, W. D., Deutsch, W. A., and Schmid, K. (**2004**). "Sensitivity to interaural level and envelope time differences of two bilateral cochlear implant listeners using clinical sound processors," Ear Hear. **25**, 5, 488–500.

Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, J. C. Jr., McElveen, J. T. Jr., and Roush, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Am. J. Otol. **19**, 758–761.

Lawson, D. T., Wolford, R., Brill, S., Schatzer, R., and Wilson, B. S. (**2001**). Twelfth quarterly progress report: Speech processors for auditory prostheses, Center of Auditory Prosthesis Research, Research Triangle Institute, NIH project No. N01-DC-8-2105. Bethesda. http://www.nidcd.nih.gov/staticresources/funding/programs/npp/pdf/archived/N01-DC-8-2105QPR12.pdf (last viewed 11/10/06).

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565–1574.

Majdak, P., Laback, B., and Baumgartner, W. D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–3622.

McKay, C. M., and Carlyon, R. P. (**1999**). "Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains," J. Acoust. Soc. Am. **105**, 347–357.

Nelson, D. A., and Donaldson, G. S. (**2002**). "Psychophysical recovery from pulse-train forward masking in electric hearing," J. Acoust. Soc. Am. **112**, 2932–2947.

Nie, K., Stickney, G., and Zeng, F. G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.

Saberi, K., and Perrott, D. R. (**1995**). "Lateralization of click trains with opposing onset and ongoing interaural delays," Acustica **81**, 272–275.

Saberi, K. (**1996**). "Observer weighting of interaural delays in filtered impulses," Percept. Psychophys. **58**, 1037–1046.

Senn, P., Kompis, M., Vischer, M., and Haeusler, R. (**2005**). "Minimum audible angle, just noticeable interaural differences and speech intelligibility with bilateral cochlear implants using clinical speech processors," Audiol. Neuro-Otol. **10**, 342–352.

Smith, Z. M., Oxenham, A. O., and Delgutte, B. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

van Hoesel, R. J., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implant subjects," J. Acoust. Soc. Am. **102**, 495–507.

van Hoesel, R. J., Ramsden, R., and Odriscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–149.

van Hoesel, R. J., and Tyler, R. S. (**2003**). "Speech perception, localization, and lateralization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.

van Hoesel, R. J. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.

Wichmann, F. A., and Hill, N. J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Wichmann, F. A., and Hill, N. J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wolford, R., Lawson, D. T., Schatzer, R., Xiaoan, S., and Wilson, B. S. (**2003**). Fourth quarterly progress report: Speech processors for auditory prostheses, Center of Auditory Prosthesis Research, Research Triangle Institute, NIH project No. N01-DC-2-1002. Bethesda. http://www.nidcd.nih.gov/staticresources/funding/programs/npp/pdf/N01-DC-2-1002QPR04.pdf (last viewed 11/10/06).

Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.

Zwislocki, J., and Feldman, R. S. (**1956**). "Just noticeable differences in dichotic phase," J. Acoust. Soc. Am. **28**, 860–864.

Sensitivity to binaural timing in bilateral cochlear implant users

Richard J. M. van Hoesel[a)]
CRC for Cochlear Implant and Hearing Aid Innovation, 384-388 Albert Street, East Melbourne, 3002 Victoria, Australia

(Received 15 May 2006; revised 10 January 2007; accepted 12 January 2007)

Various measures of binaural timing sensitivity were made in three bilateral cochlear implant users, who had demonstrated moderate-to-good interaural time delay (ITD) sensitivity at 100 pulses-per-second (pps). Overall, ITD thresholds increased at higher pulse rates, lower levels, and shorter durations, although intersubject differences were evident. Monaural rate-discrimination thresholds, using the same stimulation parameters, showed more substantial elevation than ITDs with increased rate. ITD sensitivity with 6000 pps stimuli, amplitude-modulated at 100 Hz, was similar to that with unmodulated pulse trains at 100 pps, but at 200 and 300 Hz performance was poorer than with unmodulated signals. Measures of sensitivity to binaural beats with unmodulated pulse-trains showed that all three subjects could use time-varying ITD cues at 100 pps, but not 300 pps, even though static ITD sensitivity was relatively unaffected over that range. The difference between static and dynamic ITD thresholds is discussed in terms of relative contributions from initial and later arriving cues, which was further examined in an experiment using two-pulse stimuli as a function of interpulse separation. In agreement with the binaural-beat data, findings from that experiment showed poor discrimination of ITDs on the second pulse when the interval between pulses was reduced to a few milliseconds. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2537300]

PACS number(s): 43.66.Pn, 43.66.Ts, 43.66.Mk, 43.66.Ba [AJO] Pages: 2192–2206

I. INTRODUCTION

Results from several studies with adult bilateral cochlear implant (CI) users have demonstrated substantial improvements in speech-intelligibility in noise and sound-source direction identification, compared to unilateral device use (van Hoesel and Clark, 1999; Gantz *et al.*, 2002; Müller *et al.*, 2002; Tyler *et al.*, 2002; van Hoesel *et al.*, 2002; van Hoesel and Tyler, 2003; Litovsky *et al.*, 2004; Seeber *et al.*, 2004; van Hoesel, 2004). However, most of these benefits appear to be derived from the use of level information at each ear, with little indication that interaural timing cues contribute. Although this might be expected from the consideration that clinical sound-processing algorithms generally discard interaural fine-timing cues, there is evidence that perceptual limitations also arise as a direct consequence of electrical stimulation. The extent of those limitations is under investigation in the work presented here.

The largest speech-intelligibility gain with bilateral CI users is seen when speech and noise are spatially separated, and is derived predominantly from the listener's ability to attend to the ear with the higher signal-to-noise ratio (SNR). Binaural speech unmasking, resulting from comparison of the differences between the signals at both ears, has not been clearly demonstrated. Although some subjects, in some studies, show a bilateral advantage when compared to the ear with the higher SNR (i.e., squelch), that is not necessarily the result of binaural unmasking. In addition to confounding listening experience with one versus two ears (van Hoesel and Tyler, 2003), asymmetries in performance between the ears might also result in a bilateral advantage over the ear with the higher SNR, if that ear is the poorer one, and remains so despite its physical SNR advantage over the contralateral ear.

The ability to identify sound-source direction has also been shown to be much improved for many adult CI users when using both ears. For pink noise bursts arriving from loudspeakers closest to the direction directly in front of the listener, four of the five subjects tested by van Hoesel and Tyler (2003) showed rms errors of 5° less.[1] For more laterally placed loudspeakers, errors often increased several-fold, which the authors attributed to smaller or more ambiguous changes in interaural level differences (ILDs) at those locations. The nature of the available cues in localization with CI was further examined in a study by van Hoesel (2004). In that study both free-field localization abilities with pink noise and low-rate audio click trains, as well as ILD and interaural time delay (ITD) sensitivity for the same signals connected to the direct audio input connector of the sound processors, were measured in two subjects using three different sound-processing strategies. Two of the strategies were clinical ones that discard fine-timing cues, and present only envelope information via modulation of a fixed high-rate pulse train. The third was a research strategy referred to by the author as PDT (peak derived timing), which was designed to preserve fine-timing cues in the electrical stimulus. Results showed that, irrespective of the strategy used, both ILDs and ITDs could be discriminated for low-rate click trains, but only ILDs could be discriminated with the pink noise bursts. It was postulated that the inability to process ITDs with pink noise was due to the rate of fluctuations, of

[a)]Electronic mail: rvanh@bionicear.org

 0001-4966/2007/121(4)/2192/15/$23.00

either fine-timing cues with PDT or envelope-modulation cues with the clinical strategies, being too fast for subjects to hear. It was also considered possible that ITD sensitivity with all three strategies was compromised by unwanted interaction due to current spread for stimulation on nearby electrodes (e.g., McKay and McDermott 1996; Cohen *et al.*, 2001). Such interaction could disrupt ITD cues by causing nerves near a particular electrode to fire at times corresponding to stimulation on other electrodes, and consequently also increases the effective stimulation rate for those nerves. Although interaction effects clearly need to be considered in practical applications, the experiments described in this paper are confined to stimuli comprising activation of a single electrode in each ear.

Before the present study only a modest number of experiments had been published describing psychophysical ITD data with electrical pulses under direct experimental control (van Hoesel *et al.*, 1993; van Hoesel and Clark, 1997; Lawson *et al.*, 1998; van Hoesel *et al.*, 2002; van Hoesel and Tyler, 2003; Long *et al.*, 2003; van Hoesel, 2004). Only static-ITD sensitivity had been investigated, and most of the studies described results for pulse trains at 100 pulses-per-second (pps) or less. The data from van Hoesel and Tyler (2003) included measurements up to 800 pps using constant loudness pulse trains in five subjects. At 50 pps, best performance corresponded to thresholds on the order of 100 μs. At 200 pps, performance was more variable and lowest thresholds were closer to 200 μs. At 800 pps, none of the subjects could perform the task with the maximal 400 μs ITDs tested. It was considered possible that those results are indicative of a common peripheral mechanism that might also limit monaural CI rate-pitch to rates below a few hundred hertz (e.g., Shannon, 1983; Tong and Clark, 1985; McKay *et al.*, 1994; Zeng, 2002). However, a confounding factor in the study by van Hoesel and Tyler (2003) was the use of constant-loudness stimuli, which at higher rates required lower stimulation levels that may have affected ITD sensitivity. In the present paper, experiment 1 extends the previous work with static-ITD signals and addresses the above-discussed questions. It includes measure of static ITD sensitivity for unmodulated pulse trains as a function of rate, level and duration, of monaural rate discrimination, and of ITD sensitivity with high-rate amplitude modulated signals as a function of modulation rate.

Everyday listening is usually not constrained by a fixed head position or stationary sound sources, so that ITDs will not be static. In experiment 2, sensitivity to time-varying ITDs was assessed using binaural-beat stimuli. Constant-rate electrical pulse trains in one ear were paired with slightly higher pulse rates in the contralateral ear (van Hoesel and Clark, 1997). This produced signals with ITDs that varied with time as a function of the rate difference between the ears, and were to be distinguished from a diotic reference condition. In normal-hearing (NH) listeners, the use of sinusoids of slightly different frequencies results in phase differences at the ears that cycle at the beat frequency. At beat frequencies up to about 2 Hz, the sound image remains well fused and its lateral position moves slowly in agreement with the phase relations at the ears, whereas at higher beat frequencies, loudness, roughness, and unilateral pitch are affected (e.g., Durlach and Colburn, 1978). NH listeners are sensitive to binaural beats with sinusoids below 1500 Hz (e.g., Licklider *et al.*, 1950), as well as high-frequency signals using differences in envelope rates at the two ears (McFadden and Pasanen, 1975; Bernstein and Trahiotis, 1996). In the study by van Hoesel and Clark (1997) it was shown that electrical pulse trains at slightly different rates at the two ears can also introduce binaural cues that allow substantially improved ability to detect dichotic rate changes compared to monaural ones, even in listeners with poorer than average ITD sensitivity, as long as the stimulation rate was 100 pps or lower. It was hypothesized that CI users with better static-ITD sensitivity, such as the three subjects in the present study, might also hear beats at higher pulse rates.

Another situation in which ITDs in subsequent parts of the signal can vary is in reverberant environments, for which later arriving reflections can provide misleading cues regarding sound-source location. In NH listeners a considerable number of experiments have demonstrated a "precedence effect," which describes reduced sensitivity to later cues in the signal compared to the onset (see, e.g., Litovsky *et al.*, 1999, for a review). In one such experiment, two-click stimuli are presented over headphones to determine sensitivity to ITDs applied only to the first, or the second, but not both clicks. Experiment 3 describes the results from a similar experiment with the three CI users using electrical two-pulse stimuli applied to the same electrodes as in experiments 1 and 2. Subjects were asked to discriminate between left, and right-ear delays at four interpulse intervals between 1 and 8 ms. The comparison of results from all three experiments, in the same subjects, was expected to provide an opportunity to better understand the potential of electrical stimulation to convey binaural timing cues.

II. SUBJECTS AND STIMULI

Three bilaterally implanted research volunteers, ME1, ME3, and ME4, participated in these experiments at the CRC Hear in Melbourne, Australia. Although the number of subjects was limited by availability of bilateral implant users and the considerable amount of testing required, all three subjects were moderate-to-good performers with regard to ITD sensitivity at 100 pps, when compared to data from previously published studies (van Hoesel and Clark, 1997; Lawson *et al.*, 1998; van Hoesel and Tyler, 2003; van Hoesel, 2004). The exclusion of subjects with poorer than average low-rate ITD sensitivity was considered appropriate in a study that aimed to examine the effects of electrical stimulation per se, rather than the additional factors that may confound data from such subjects. All three subjects received several training sessions allowing familiarization with tasks and signal cues prior to formal data collection. Subject ME1 was particularly well trained, having participated in similar experiments for almost 4 years. The data from this subject, for the present study, were collected over about 9 months, during which he was tested for three consecutive days every 6–8 weeks. Subject ME3 had also participated in bilateral CI experiments every week, or every second week, for nearly

TABLE I. Subject and electrode details. The stimulation levels specified in the last two columns refer to those corresponding to 80% of the dynamic range at 400 pps (see experiment-1 text for details).

Subject (Age, years)	First implant date (ear)	Second implant date (ear)	Hearing loss	Sound processor (strategy)	Left ear electrode (80%DR400)	Right ear Electrode (80%DR400)
ME1(67)	05-00 (L)	12-00 (R)	Menier's (adult onset)	Spear (PDT)	11 (915 μA)	11 (844 μA)
ME3 (50)	06-02 (L)	02-04 (R)	Genetic (adult onset)	Esprit (ACE)	11 (499 μA)	13 (530 μA)
ME4 (39)	08-01 (R)	09-04 (L)	Genetic (late childhood onset)	Esprit (ACE)	11 (763 μA)	11 (749 μA)

1 year prior to data collection. Her data, for the present study, were collected over a 6-month period of approximately weekly sessions. ME4 had been using bilateral devices for a shorter period of 6–7 months, and was only available on a weekly basis over a more limited period of 4 months. All three subjects were successful users of sequentially implanted bilateral Nucleus-24 systems. Subject details and electrodes used in the experiments are summarized in Table I. The third-last column in Table I indicates the sound-processor and coding strategies used most of the time whilst participating in the studies. Subject ME1 was a regular user of the PDT research strategy referred to in Sec. I, whereas the other two subjects used clinical behind the ear devices with commercial envelope-based strategies. This is provided for reference only. All the tests were done using direct activation of the electrode currents, using a custom laboratory system that is able to control binaural pulse timing within a few μs. The last two columns in Table I indicate the electrode numbers used in the experiments. Electrode 22 is the most apical, and electrode 1 the most basal band on each array. Spacing between adjacent electrodes is 0.75 mm. Electrodes near the middle of the array were selected for all three subjects. This was done after conducting pilot tests that confirmed that using very apical electrodes resulted in no better, and at times worse, ITD sensitivity, which is in agreement with the finding that none of the CI studies to date has shown systematic effects of absolute place of stimulation on ITD sensitivity. Place matching between ears was ensured using digitized modified Stenver's view x rays (Xu*et al.*, 2000), as well as subjective pitch comparisons (van Hoesel, 2004). Stimuli always comprised monopolar biphasic current pulses, using extra-cochlear reference electrodes. Current pulses were 25 μs per phase in duration, with a gap of 8 μs between cathodic and anodic phases.

III. EXPERIMENT 1: STATIC ITD SENSITIVITY

A. Methods

Experiment 1 describes ITD sensitivity with identical ITDs applied to each pulse in the signal. Thresholds were determined using a two-interval, two-alternative forced-choice (2-AFC) procedure, in which stimuli with equal magnitude ITDs leading in opposite ears were presented in random order. This procedure avoids the ambiguity inherent in methods that assess sensitivity to cues in a target signal that follows a preceding reference stimulus (Hartmann and Rakerd, 1989). Subjects were required to respond which of the two intervals produced a sound lateralized further to the right. As the total cue size available was twice the ITD used in either interval, results presented throughout the paper report the doubled cue. In each test block, 40 repeats of 4 fixed cues were included, resulting in 160 stimuli, which were presented in unconstrained randomized order. The four cue sizes included were logarithmically spaced between 100 μs ($\pm$50 μs) and 800 μs ($\pm$400 μs). However, if ITDs in that range did not result in performance both above and below 75% correct, additional blocks with four different cue sizes were tested in the same way until that criterion was fulfilled. Although interval bias was generally found to be small, bias-corrected percent-correct scores, $P(C)_{\max}$, were calculated and used to fit cumulative Gaussian functions using weighted linear regression. ITD thresholds were determined for $P(C)_{\max}=76\%$, corresponding to $d'=1$, and standard errors of the means (SEM) were estimated using a bootstrap procedure (Foster and Bischof, 1991). The same procedure was used to calculate the thresholds and SEM values for the other experiments in the paper.

In experiment 1A, pulse rates were varied from 100 to 600 pps, with stimulation level held fixed at approximately 80% of the dynamic range (DR) for a 300-ms pulse-train at 400 pps, henceforth denoted as "80%DR400." For subjects ME1 and ME3, additional measurements were also made at 60% of the DR at 400 pps (60%DR400). The measured 80%DR400 levels are shown for each electrode in Table I. Subjects reported that the loudness of the signals at that level was comparable to that experienced with their sound processors when listening to everyday speech. For both fixed levels, balance between the two ears was fine-tuned at 100 pps, if needed, by adjusting levels slightly until subjects reported that alternating presentation with 800 μs delays applied to left and right ears resulted in equidistant lateral shifts toward each ear. Note that if loudness varied differently as a function of rate in each ear, lateralization shifts at rates other than 100 pps may not have been symmetrical about the midline. However, none of the subjects commented that this was the case, and the impact of any such asymmetries will likely have been further reduced by the use of the 2-AFC procedure. Preliminary measurements with ME4 indicated that this subject's ITD thresholds were

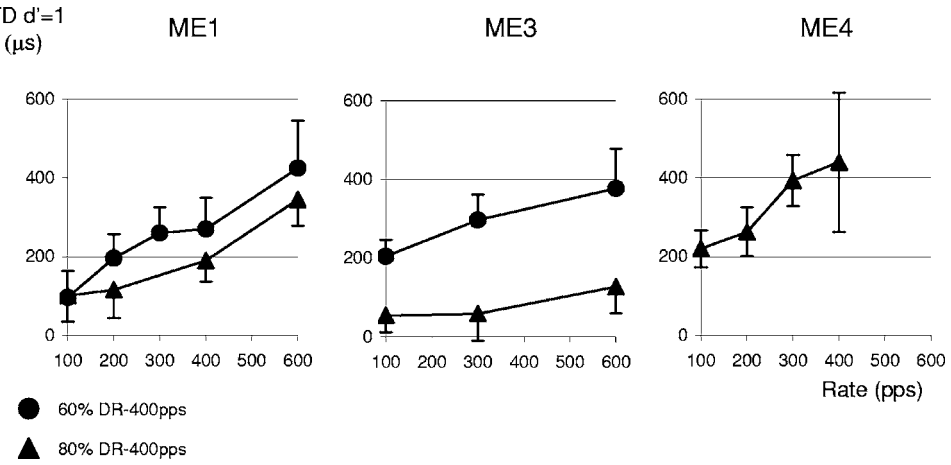


FIG. 1. ITD thresholds $[P(c)_{\max}=76\%,\ d'=1]$ with unmodulated pulse-trains for three bilateral CI users as a function of rate at the fixed stimulation level corresponding to 80% DR400 (triangles, see the text), and for ME1 and ME3 also at 60%DR400 (circles). Error bars indicate standard errors of the means (SEM), estimated using a bootstrap procedure.

somewhat higher than for the other two subjects, and that at 600 pps, ITDs in excess of half the rate interval would probably be needed. As this introduces an ambiguous ongoing ITD cue that potentially conflicts with the onset information, and because of time constraints with this subject, it was decided to limit the highest rate with her to 400 pps in the formal data collection.

In experiment 1B, monaural rate-pitch discrimination thresholds were measured using a three-interval 2-AFC procedure, that included an initial reference interval and two subsequent intervals containing either reference or probe signals in random order. The subject's task in this case was to determine whether the second or third interval contained the signal with a different rate pitch to that contained in the first interval. Subjects were specifically instructed to attend to pitch, and ignore any other cues. This "odd-man-out" paradigm was used to facilitate comparison with the results from experiment 2, in which a similar procedure was used to allow subjects to attend to either pitch or lateralization cues when detecting dichotic rate changes. The majority of monaural rate-discrimination measurements were made without level-roving, which is sometimes used to eliminate loudness cues that can result from large pulse-rate variations. The potential contribution from loudness cues was limited by restricting allowable rate changes to 30% higher than the reference rate, beyond which thresholds were not determined. As subject ME1 was available for the largest number of test sessions, the potential contribution of loudness cues was further assessed with this subject by including additional measurements with level-roved stimuli at the highest rate tested, for which rate-difference thresholds were largest and therefore loudness cues were expected to be maximal.

In experiment 1C, ITD sensitivity was measured as a function of stimulus duration at both 100 and 400 pps. ITD thresholds were determined for a single pulse, as well as durations of 50, 100, and 300 ms, using the same 2-AFC procedure as in experiment 1A. Stimulation levels were held fixed at 80%DR400 so that the single pulse results might also serve as an estimate of the contribution from the first pulse in the signals used in experiment 1A.

In experiment 1D, 6000 pps amplitude-modulated (AM) pulse trains were used to measure sensitivity to ITDs applied to the entire electrical stimulus, as a function of the modulation frequency over the range 100–400 Hz. Delaying the entire wave form, rather than just the modulator, avoids the need to resample the envelope at the carrier pulse times in one ear, and therefore eliminates the possibility that interaural level cues might be introduced that conflict with the ITD[2] (van Hoesel and Tyler, 2003). Sinusoidal AM was applied to the "current levels" (CL) used to control stimulation currents in these implants. Each CL increment corresponds to approximately 0.17 dB increase in stimulation current. The peak-to-peak depth of the modulator was set at 40 CL (approximately 6.8 dB), which for all subjects spanned at least 35%, and usually considerably more, of the electrical dynamic range at 6000 pps. The starting phase of the modulator was fixed at 0° so that envelope-ILDs at the onset of the delayed signal always reinforced the lateralization cue available from the applied ITDs.[3] All signals were 300 ms in duration, and were presented at a fixed peak-stimulation level referred to as "80%DRmod300," which corresponded to 80% of the dynamic range for a 6000 pps signal with 6.8 dB AM at 300 Hz. The same 2-AFC paradigm was used as in experiment 1A, and level balance between the two ears was again fine-tuned by ensuring equidistant lateral shifts resulting from 800 μs ITDs applied to each ear with 100 Hz AM signals. Final peak stimulation levels with the AM signals were, on average, about 2 dB lower than those used with the unmodulated 80%DR400 pulse-trains in experiment 1A.

B. Results and discussion

1. Effect of rate at constant stimulation levels (experiment 1A)

Figure 1 shows ITD thresholds $[P(C)_{\max}=76\%,\ d'=1]$ for each subject as a function of rate, at a fixed stimulation level corresponding to 80%DR400, and for subjects ME1 and ME3 also at 60%DR400. Thresholds at 100 pps and 80%DR400 were about 55, 100, and 220 μs for ME3, ME1, and ME4, respectively. These values correspond well with the range of thresholds between 90 and 200 μs described in the literature for good performers tested at 50 or 100 pps (Lawson *et al.*, 1998; van Hoesel and Tyler, 2003; van Hoesel, 2004), as well as a few reports indicating thresholds of 50 μs or better in a small number of subjects (Lawson *et al.*,

2001; Litovsky *et al.*, 2005). The thresholds found are considerably better than those reported for poor performing subjects, which can be in excess of a millisecond (e.g., van Hoesel and Clark, 1997). Recent work by Litovsky *et al.* (2005) indicates that subjects who experience hearing loss early in life are more likely to display large ITD thresholds. Although it is possible that electrical ITD sensitivity in the CI users tested here was also affected by deafness, only ME4 experienced any hearing loss during (late) childhood.

The data from this experiment were subjected to an analysis of variance (ANOVA) employing a general linear model (GLM) with rate and level as fixed factors, and subject as a random factor. The effects of rate ($F[4,11]=6.9$; $p=0.005$) and level ($F[1,11]=12.5$; $p=0.005$) were both found to be significant. A further ANOVA, conducted with just the two subjects who completed the tests at both stimulus levels, similarly showed significant effects of both rate ($F[1,8]=10.95$, $p=0.011$) and level ($F[4,8]=4.73$, $p=0.03$) on ITD thresholds. One mechanism that may have contributed to elevated thresholds at higher rates is neural refractory behavior associated with electrical stimulation, which has been demonstrated even at rates as low as a few hundred pps (e.g., Wilson *et al.*, 1997), and is possibly exacerbated by the effects of deafness (e.g., Shepherd *et al.*, 2004). In addition, mechanisms that limit ITD sensitivity in NH listeners with high-frequency (HF) signals at higher envelope-fluctuation rates may have played a role, as will be discussed further when comparing the data with normal-hearing listeners' performance. Recently, Smith (2006) reported ITD thresholds for electrical stimulation that were estimated from single unit recordings in the IC of anaesthetized cats. Despite interspecies differences, thresholds based on those data were around 150 μs for rates of 40 and 80 pps, which is comparable to the mean subjects' threshold of about 125 μs at 100 pps (and 80%DR400) in Fig. 1. At 160 and 320 pps, the single-unit derived thresholds increased to about 300 and 500 μs, respectively, whereas the behavioral data in Fig. 1 show minimal increases over a similar range, which suggests that pooling across the population may become more effective as rates increase.

The improved ITD sensitivity at higher levels may be the result of a larger portion of the neural population providing input to binaural ITD comparison circuits. Higher levels may also result in decreased variance in response latency (e.g., Javel and Shephard, 2000), resulting in better synchronized input to those circuits. The single unit data from Smith (2006) showed that some units responded to ITDs over a very limited dynamic range. Even if a considerable number of neurons showed such nonmonotonicity, the effect might be linearized via pooling across a large number of neurons with best responses over different level ranges, and therefore be absent in the behavioral data. The two subjects tested at both levels showed different amounts of threshold elevation at the lower compared to the higher level, despite the use of equivalent level adjustments in terms of the dynamic range. This too might be explicable in terms of numbers of neurons firing. Subject ME3 required lower overall stimulation currents than ME1, and if this is due to better neural survival, that subject may experience more nonlinear growth of loudness with increasing stimulus level (Cohen *et al.*, 2006). This may have resulted in larger loudness differences between stimulation levels at 60 and 80%DR400, corresponding to larger differences in the number of neurons firing, and consequently ITD sensitivity at the two levels. The significant effect of level found in the present experiment suggests that the elevated thresholds at higher rates in the data from van Hoesel and Tyler (2003) may indeed have partly resulted from the level reductions used to maintain constant loudness in that study.

Individual slopes of the threshold-rate functions in Fig. 1 varied considerably for these three subjects. At 80%DR400, subjects ME1 and ME4 showed larger threshold increases with rate than did ME3. For ME4, thresholds at 300 pps were about twice as large as those at 100 pps, whereas for ME3 thresholds remained unaffected over that range. ME4 also showed the largest absolute thresholds, which may be due to earlier loss of hearing, and reduced testing and bilateral CI listening experience compared to the other two subjects. Casual observation and preliminary measurements of ITD sensitivity with these three subjects showed substantially improved ITD sensitivity with time over the first 6 months after bilateral implantation, as well as with increased test sessions. The least effect of rate was seen in ME3. At the higher level, her thresholds at 100 and 600 pps do not differ significantly according to the bootstrap error estimates. It was considered whether that result might be attributable to good sensitivity to the "onset" ITD, defined in this paper as the ITD applied to the first pulse (unless noted otherwise), with minimal "ongoing" contributions from later pulses. That hypothesis was tested by making an additional measurement at 1000 pps (not shown in Fig. 1). At that rate, the subject's threshold increased five- or sixfold compared to the value at 100 pps. If the onset cue was the main reason for good sensitivity at 600 pps, it seems curious that it would not also be available at 1000 pps. Although smearing of timing cues across several pulses at higher rates is one potential mechanism that could lead to such a result, the data to be discussed in experiment 3 do not support this for an interpulse interval of 1 ms. It seems more likely that this subject was indeed able to attend to ongoing ITD cues at 600 pps, but not 1000 pps. Further evidence that some CI users can hear ITD cues at rates as high as 600 pps has been reported recently by Laback *et al.* (2005), Jones *et al.* (2006) and Majdak *et al.* (2006). Laback *et al.* measured ITD sensitivity in three bilateral CI users using four-pulse sequences at rates in the range 100–800 pps, and found ITD sensitivity that generally deteriorated for whole waveform delays as rates increased, but also showed sensitivity to ongoing delays at 800 pps in at least one subject, even when the first pulse was diotic. Jones *et al.* tested ITD sensitivity on several electrodes in each of seven subjects, and in agreement with Fig. 1, on average found increasing thresholds with rate but substantial variation across subjects and across electrodes within subjects. In at least one subject, good ongoing ITD sensitivity was also indicated at 600 pps when using pulse trains with 50 ms rise times to reduce the effect of the onset cue. Majdak *et al.* (2006) applied ITDs to electrical carrier pulses in 320-ms-long pulse trains, comprising four cycles of 100%

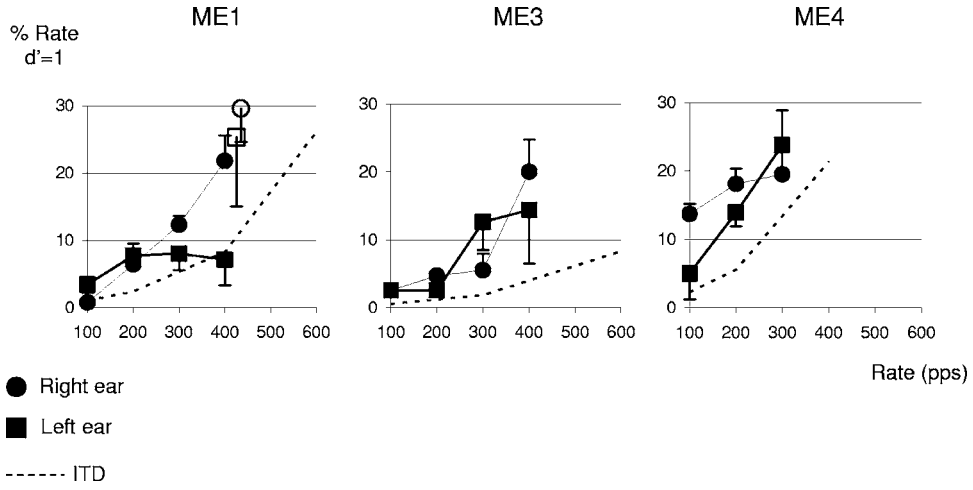


FIG. 2. Monaural rate discrimination thresholds (solid lines, $d'=1$) for the same electrical stimulation parameters and subjects as in Fig. 1. Circles and squares show rate-pitch DLs for left and right ears, respectively. Open symbols for subject ME1 show rate DLs at 300 pps when level roving was employed. Dashed lines show the ITD thresholds from Fig. 1, expressed as the change in rate that would result if the threshold ITD were subtracted from the stimulation-rate interval at each reference rate.

trapezoidal modulation at 12.5 Hz whilst holding the envelope ITD fixed independently of the carrier ITDs. They too found best (carrier) ITD sensitivity at low pulse rates, and showed evidence that ITDs could be heard at 400 pps, and possibly also at 800 pps, in two of four subjects.

It is interesting to compare the effect of electrical pulse rate on ITD thresholds with both high- and low-frequency ITD sensitivity in normal-hearing listeners. Normal-hearing listeners' ITD thresholds for 100-Hz pure tones with slow rise times are quite similar to the 100 pps thresholds in Fig. 1 (e.g., Klumpp and Eady, 1956). Although comparison of actual threshold values is not particularly informative because stronger onset cues were likely available with the electrical pulse trains, there is a clear contrast between increasing thresholds in the CI data as rates increased, and *decreasing* thresholds observed in NH listeners as pure-tone frequencies increase from 100 Hz to about 1 kHz. The auditory nerve response to simple electrical pulse trains is substantially different than that resulting from low-frequency sinusoids traveling along the basilar membrane in NH listeners, and this may be the primary reason for reduced ITD sensitivity at higher rates with simple electrical pulse trains. In addition to the neural refractory effects already mentioned, an important difference may be that with electrical stimulation the place and rate information are generally not matched as they are for pure tones in normal hearing. Such mismatch has been argued to result in reduced pitch sensitivity in NH listeners (Oxenham *et al.*, 2004), although Carlyon and Deeks (2002) showed that stimuli with mismatched rate and place information in NH listeners results in better rate discrimination than is usually seen in implant users, which implies other factors must also contribute to the electrical rate-pitch result. Whereas in normal hearing specific phase relations exist at different locations along the basilar membrane, the neural response to electrical pulses is synchronous over a fairly broad region along the cochlea. If matching of place and rate information, and preservation of appropriate phase relations are critical to monaural temporal coding, as proposed in some models for pitch perception (e.g., Loeb *et al.*, 1983; Loeb, 2005; Carney *et al.*, 2002), perhaps it is the case that they are also needed to optimally convey ITDs. Although the data from Oxenham *et al.* (2004) do not support that conjecture below about 150 Hz where ITD discrimination is similar for transposed and pure tones, at higher frequencies ITD (and pitch) discrimination is clearly poorer with transposed tones.

ITD sensitivity in NH listeners does decrease with increasing *envelope* rate for HF signals. Hafter and Dye (1983) tested four subjects with 4 kHz bandpass-filtered sequences of 32 clicks, and found ITD thresholds between 15 and 100 μs at a repetition rate of 100 Hz, and approximately double those values at 500 Hz. Although those listeners may have had access to low-frequency cues that were incompletely removed by the filtering employed, as rates increased there is a clear decrease in ITD sensitivity not unlike that seen in the CI users' data. When onset cues are removed, and care is taken to avoid low frequency cues in NH listeners, the effect of rate can be much stronger. ITD sensitivity has been reported to be very poor with HF signals with slow rise times and AM rates above about 150 Hz (Bernstein and Trahiotis, 2002), even when transposed tones (van de Par and Kohlrausch, 1997) are used to optimize envelope-ITD cues. The 150 Hz limitation is also seen in monaural AM detection tasks in NH listeners (Kohlrausch *et al.*, 2000; Ewert and Dau, 2000), but is presently not well understood, so that it is difficult to determine whether electrical pulse trains are likely to be affected by the same mechanism. Mounting evidence that some CI users can hear ongoing ITD cues up to 600 Hz or higher suggests that may not be the case. However, as with the CI data, NH listeners show considerable intersubject variation, and one of the subjects tested by Bernstein and Trahiotis (2002) was also able to use ongoing ITD cues at an AM rate of 512 Hz.

2. Monaural rate discrimination (experiment 1B)

The solid lines in Fig. 2 indicate the monaural rate-difference thresholds ($d'=1$) for each subject, expressed as a percentage of the reference rate, and using the same electrodes and 80% DR400 stimulation levels as used in experiment 1A. Circles and squares show thresholds for right and left ears, respectively. The dashed lines show the ITD thresholds from Fig. 1 expressed as an "equivalent" percentage change in stimulation rate, calculated by subtracting the threshold ITD at each rate from the interpulse interval at that rate.[4] This measure was chosen to better allow comparison with the monaural rate thresholds, which were measured as

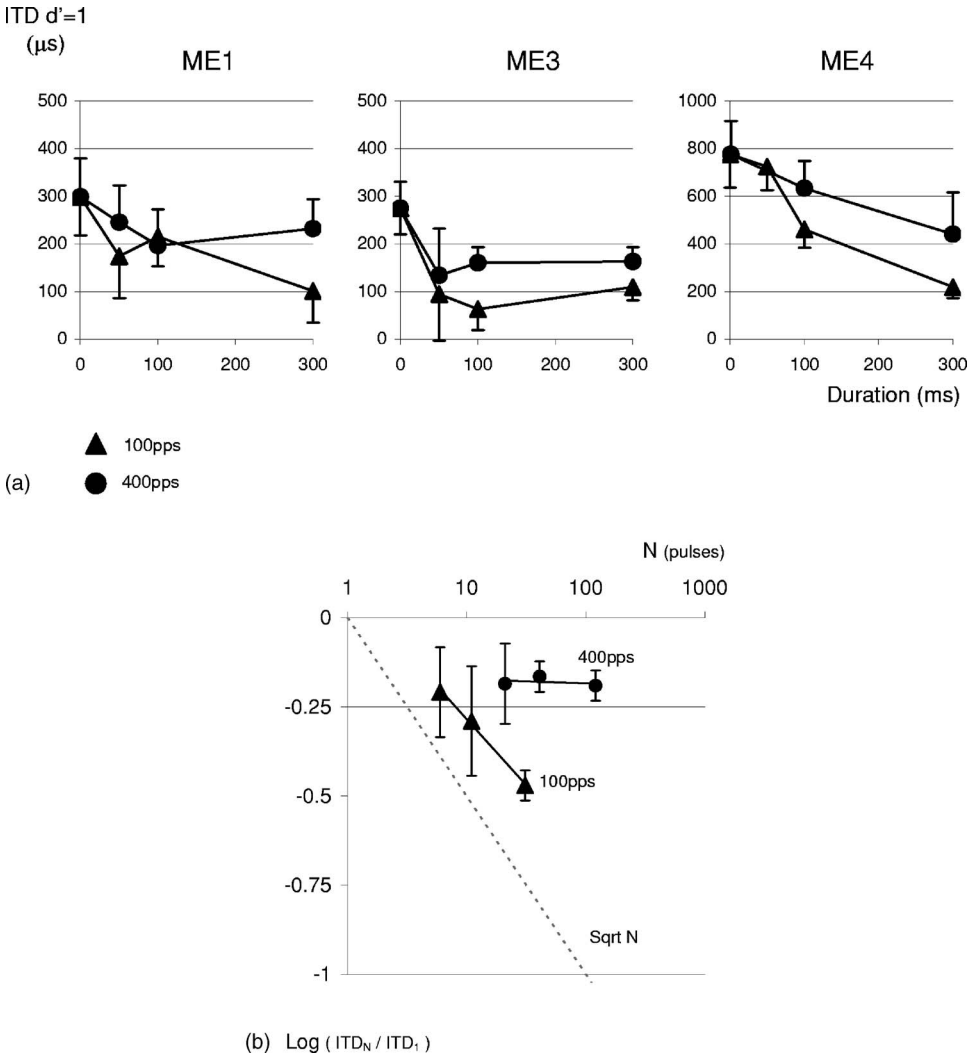


FIG. 3. (a) ITD sensitivity as a function of pulse-train duration, from a single pulse (shown as "0 ms") to 300 ms, at stimulation rates of 100 pps (triangles) and 400 pps (circles). (b) Data from (a) averaged across subjects after normalization with respect to each subject's single-pulse ITD threshold. The logarithm of the subjects' average improvement in ITD threshold for a duration of N pulses, compared to the single-pulse ITD threshold, is shown as a function of number of pulses (N) plotted logarithmically. The dashed line indicates ideal-observer behavior with perfect integration of the ITD information from each pulse, resulting in a slope of −0.5.

the smallest increase in rate, and therefore decrease in interpulse interval, that could be heard. Monaural rate-discrimination thresholds were not determined for ME4 at 400 pps because rate variations of 30% were not audible. The rate-discrimination data were subjected to an ANOVA using a GLM, treating each ear for each subject as a random factor. The effect of reference rate was found to be highly significant ($F[3,13]=10.16$, $p=0.001$).[5] When comparing results at 400 and 100 pps, the monaural rate data for ME3 in both ears, and ME1 in the right ear, show considerably larger threshold elevation than is seen in those subjects' ITD data. When modest intensity roving of up to ±0.34 dB was employed to eliminate potential loudness cues at 400 pps with ME1, rate-pitch thresholds at 400 pps increased further (open symbols). Because the rate-detection thresholds were small at 100 pps, loudness cues were likely negligible at that rate. Consequently, ME1's roved-level thresholds at 400 pps imply that in the absence of loudness cues, monaural functions might be even steeper than those plotted in Fig. 2.

To compare the effect of pulse rate on ITD and rate-pitch discrimination tasks, the data from each subject, and in each task (rate or ITD discrimination), were normalized with respect to performance at 100 pps, and subsequently used to fit regression lines for the effect of rate in each task. An accumulated ANOVA showed that the effect of rate in the ITD and rate-pitch tasks differed significantly ($F[1,18]=8.8$, $p=0.008$), and indicates that ITD cues were available at rates for which monaural rate-pitch cues were not. This is likely to be at least in part attributable to the availability of an onset ITD cue from the first pulse in each ear, which would result in a slower decline of the total cue as ongoing ITDs became more difficult to discern at higher rates. In contrast, even the first interpulse interval depends on information beyond the first pulse in the monaural rate discrimination task. A common peripheral limitation could therefore in principal be responsible for performance in both tasks, despite the larger range of rates over which ITDs were available. In support of this idea it is interesting to note that the ordering of both monaural and binaural performance is the same across subjects. However, the data from ME3, showing sensitivity to ongoing ITD cues at 600 pps (and as shall be shown in experiment 1C, at 400 pps for all three subjects), combined with very poor rate discrimination at 400 pps, implies that monaural rate discrimination may be subject to additional limitations that do not affect ITD perception.

3. Effect of duration (experiment 1C)

Figure 3(a) shows ITD sensitivity as a function of stimulus duration from a single pulse (0 ms) to 300 ms, at both

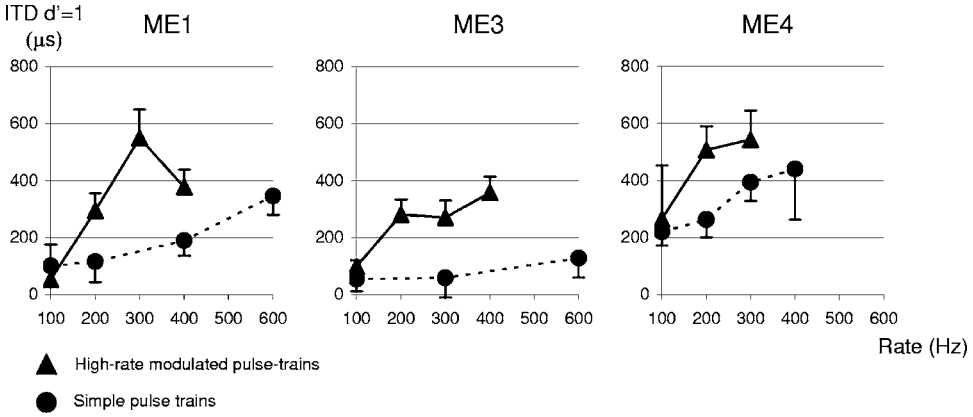


FIG. 4. Whole-wave-form ITD thresholds ($d'=1$) with 6000 pps AM pulse trains (triangles, solid lines) as a function of modulation rate. The unmodulated ITD thresholds from experiment 1A (circles, dashed lines) are shown for comparison at matched rates.

100 and 400 pps. The data were normalized with respect to each subject's single-pulse ITD threshold prior to applying an ANOVA at each rate. The effect of duration was found to be significant at both 100 pps ($F[3,8]=6.01$, $p=0.019$), and 400 pps ($F[3,7]=5.36$, $p=0.031$), which means ongoing ITDs contributed significantly to overall ITD thresholds up to at least 400 pps for these three subjects. Comparison of subject ME3's single-pulse ITD threshold, with her 80% DR400 ITD-threshold for a 300 ms long pulse train at 600 pps from experiment 1A, shows that the latter was about half the single-pulse threshold, which lends further support to the claim that this subject could make use of ongoing ITDs up to 600 pps. To quantitatively asses the reduction of ITD threshold when adding more pulses in the stimulus, the normalized ITD data (with respect to single-pulse thresholds) were averaged across subjects, and plotted as a function of number of pulses in the stimulus on log-log axes for both pulse rates in Fig. 3(b). Error bars indicate the standard error of the mean across the three subjects. An ideal observer with perfect integration of the ITD information from each pulse would show performance that improves as the square root of the number of pulses (e.g., Houtgast and Plomp, 1968), which corresponds to a slope of −0.5 on this plot, as indicated by the dashed line. The solid lines in Fig. 3(b) show lines of best fit over the range of durations spanning 50–300 ms at each rate. For the 100 pps pulse train the slope of that fit is close to −0.5, but at 400 pps it is substantially less steep. Despite the normalization of the data with regard to the single pulse thresholds, there was large intersubject variation with regard to the effect of rate. As a result, an accumulated ANOVA associated with a simple regression analysis, grouped by rate, indicated that the effect of duration did not differ significantly at 100 and 400 pps when data from all three subjects were included ($F[1,13]=1.67$, $p=0.22$). However, when the data from subject ME3, who showed little effect of rate on ITD sensitivity in experiment 1A, were excluded, the difference in slopes showed information from the later pulses was not used as effectively at 400 pps, as at 100 pps ($F[1,7]=5.38$, $p=0.05$) for those two subjects, Although ME3 showed little effect of duration over the range 50–300 ms at either rate in Fig. 3(a), absolute thresholds over that range were significantly lower at 100 than 400 pps (one-way ANOVA, $F[1,4]=15.1$, $p=0.018$). Assuming that the onset cue was equally available at both rates, that result indicates that for this subject too ITD cues from pulses beyond the first were more effective at 100 than 400 pps, and conversely that the *relative* contribution of the onset to the total cue increased at the higher rate.

The decreased effectiveness of later pulses at the higher rate is in agreement with data from NH listeners attending high-frequency click trains when rates exceed about 100 Hz (Hafter and Dye, 1983; Saberi, 1996; Stecker and Hafter, 2002). Hafter *et al.* (1988) discuss this result in terms of "binaural adaptation," which is monaurally mediated, perhaps at the cochlear nucleus, but is not the result of auditory nerve adaptation. Although adaptation at the auditory nerve may differ with electrical stimulation, compared to normal hearing, there is evidence that such adaptation is probably less in the electrical case (Loquet *et al.*, 2004), so that the data in Fig. 3 suggest binaural adaptation in CI users resembles that seen in NH listeners. Hafter and Dye (1983) modeled shallower slopes at higher rates, similar to those in Fig. 3(b), by assuming successively decreased contributions from each pulse throughout the stimulus duration. However, more recent "observer weighting" studies that directly assess relative weighting of ITDs applied at each click (Saberi, 1996; Stecker and Hafter, 2002) show no evidence of a gradual decline, and instead indicate an immediate reduction of weights beyond the first click. The effectiveness of information beyond the first click with electrical stimulation is further investigated in experiment 3.

4. Effect of modulation rate with high-rate carriers (experiment 1D)

Figure 4 shows whole-wave-form ITD thresholds for 6000 pps AM signals, as a function of modulation rate between 100 and 400 Hz (triangles). Performance at 400 Hz AM was not measured for subject ME4 due to time limitations. The threshold-rate curves from Fig. 1, for low-rate unmodulated signals on the same electrodes, and at pulse rates equal to the modulation rates used here, are included for comparison (circles). AM and unmodulated signal ITD thresholds were included in a GLM ANOVA, with subject as a random factor. Results showed significant effects of rate ($F[3,10]=12.86$, $p<0.001$), and signal type ($F[1,10]=34.4$, $p<0.001$), as well as near significant interaction between those factors ($F[3,10]=3.55$, $p=0.056$). Further

analysis using a 5% least significant difference of means (LSD) criterion revealed no difference between the thresholds for the two signal types at 100 Hz, but at higher rates, thresholds were higher with AM signals than with the unmodulated pulse trains. Thresholds for the AM signals were significantly lower at 100 Hz (5% LSD), than higher AM rates, but did not differ significantly over the range 200–400 Hz.

The reduced ITD sensitivity for AM rates in the range 200–400 Hz, compared to unmodulated signals at those rates, is of both practical and theoretical interest. It implies that ITDs for frequencies in that range might be better coded using a single pulse per cycle, as in the PDT strategy, than by using a high-rate modulated envelope as in present clinical strategies, at least for these subjects and in the absence of electrode interaction. Several factors may have contributed to the difference in performance with the two signal types. One possibility, in accordance with the level effect seen in experiment 1A, is that the ITD thresholds were larger with the AM signals as a result of the lower stimulation currents needed to produce acceptable loudness levels at 6000 pps. The AM signals may also have included weaker onset cues than the unmodulated signals, both because of the more gradual rise time associated with the log-sinusoidal modulation, as well as the lower level and additional 3.4 dB reduction in stimulation level on the first pulse (for a starting phase of 0° and a modulation depth of 6.8 dB) compared to the largest pulses in the signal. The difference in performance with AM and unmodulated signals may have been reduced at 100 Hz due to good sensitivity to ongoing cues at that low rate, thereby reducing the reliance on the onset cue. Ongoing cues at higher AM rates may have been disrupted by poor coding of the envelope due to strong refractory effects associated with the 6000 pps carrier, whereas at 100 Hz modulation neurons might have been able to recover during the low-level portion of the signal envelope. This might also explain nonmonotonic behavior with rate suggested by ME1's data. To test the possibility that increased periods with low stimulation levels might improve ITD sensitivity, a further pilot test was conducted with subject ME1, comparing performance with an extremely deep modulation of 25 dB with that using 6.8 dB, whilst maintaining the same peak stimulation level. However, results were very similar at both modulation depths, perhaps because the amplitude of the onset was also reduced for deeper AM. An interesting extension of the present experiment might be to relax the requirement that the ITD and (interpolated) onset-envelope ILD be congruent,[3] and adjust the starting phase of the sinusoidal modulator such that the first pulse in the signal occurs at a maximum and remains unaffected by the modulation.

IV. EXPERIMENT 2: TIME VARYING ITD SENSITIVITY (BINAURAL BEATS)

A. Methods

Sensitivity to time-varying ITDs was measured using a binaural-beat task, in which a pulse train with diotic onsets (0-ITD) followed by increasing ITDs was to be distinguished from one that remained diotic. Increasing ITDs were generated by using slightly different pulse rates in the two ears. Stimuli comprised 300 ms pulse trains, at nominal rates of 100, 200, and 300 pps, applied to the same electrodes and at the 80%DR400 stimulation level described in experiment 1A. Since different cues were potentially available depending on the (interaural) rate difference, subjects were asked to identify whether the second or third interval, in a three-interval 2-AFC task, contained the signal that differed from the diotic reference in the first interval. The use of a fixed duration signal means that a fixed beat frequency, or equivalently interaural rate difference, results in an ITD at the end of the stimulus that is a fixed fraction of the interpulse interval, and is therefore larger in absolute terms at lower stimulation rates. For example, a binaural-beat frequency of 1 Hz and signal duration of 300 ms, results in a final ITD of 3 ms at 100 pps, but only 1 ms at 300 pps. Note also that, for a duration of 300 ms, binaural-beat rates above 1.67 Hz result in final ITDs that correspond to more than half the interpulse interval.[6]

Within each test block, the reference rate was held fixed at 100, 200, or 300 pps. Thirty repetitions of four different beat frequencies, resulting in 120 stimuli in total, were presented in random order in each block. An initial pilot test was conducted with each subject at each reference rate to determine an appropriate range of beat frequencies to include in the formal evaluation. In the pilot test, subjects indicated whether any differences in lateral position, image spread, or pitch could be heard when listening to continuous alternating presentation of 300 ms pulse trains that were diotic in one interval and comprised a 1 Hz beat in the other. If a difference was discernible at 1 Hz, the test block was constructed using four beat frequencies spaced at one-octave intervals from 1 Hz down to 0.125 Hz. If not, the rate was further increased in one ear until the dichotic stimulus could reliably be distinguished from the diotic condition, and the formal test block was similarly constructed with four rate differences decreasing in one-octave intervals from that rate. If performance from the test block did not bracket 75% correct, repeat tests with additional blocks containing smaller or larger rate differences were conducted until that criterion was satisfied. Thresholds corresponding to $d' = 1$ were estimated in the same way as described in experiment 1A.

B. Results and discussion

The solid lines in Fig. 5 show the dichotic rate thresholds for each subject. To facilitate comparison with the monaural rate discrimination data from experiment 1B, beat-detection thresholds are shown as a percentage change of the rate in one ear relative to the reference rate at 100, 200, or 300 pps. Error bars around each threshold indicate bootstrap estimates of one standard error of the mean. Circles and triangles are for rate changes in left and right ears, respectively, whilst keeping the contralateral signal fixed at the reference rate. The dashed lines show the monaural rate discrimination thresholds from experiment 1B. Compared to the subjects tested by van Hoesel and Clark (1997), all three subjects in the present study showed lower dichotic-rate thresholds (Fig. 5), as well as static ITD thresholds at 100 pps [Fig. 1(a)]. At

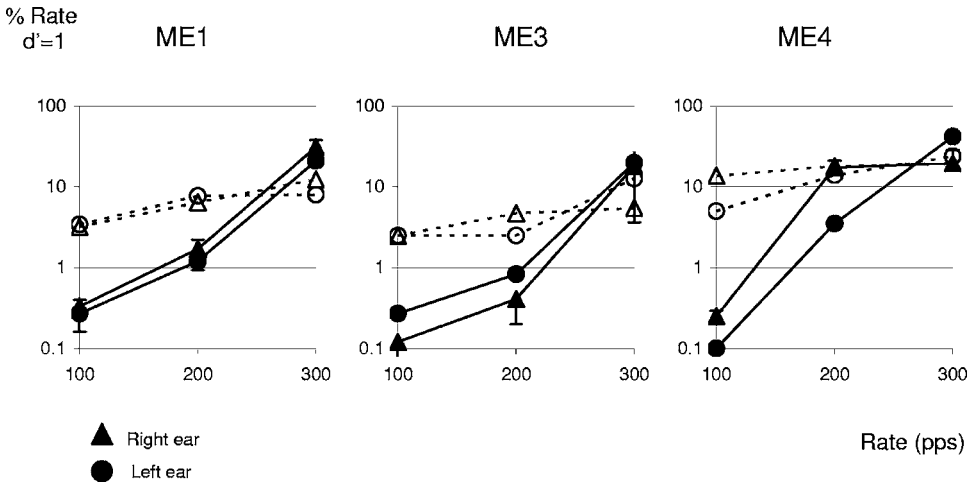


FIG. 5. Dichotic rate difference thresholds ($d'=1$) for the binaural-beat detection task in experiment 2 (closed symbols, solid lines). Thresholds are shown as a percentage change relative to the reference rates of 100, 200, and 300 pps. Circles and triangles are for rate changes in left and right ears, respectively, with the contralateral signal held fixed at the reference rate. Monaural rate discrimination thresholds, from experiment 1B, are also shown for comparison (open symbols, dashed lines).

100 pps, subjects spontaneously reported lateralization shifts for the dichotic stimuli and thresholds were close to 0.1%, which is at least an order of magnitude better than their monaural results. At 200 pps, performance was closer to 1% for ME1 and ME3, which was still considerably better than their monaural results at that rate, but approached monaural performance for ME4. At 300 pps, all three subjects showed dichotic thresholds that were no better, and sometimes worse than their monaural thresholds, and none reported hearing lateralization shifts. Median dichotic thresholds were 0.2%, 1.3%, and 18%, corresponding to beat-frequencies of 0.2, 2.6, and 48 Hz, at 100, 200, and 300 pps, respectively. A GLM ANOVA was applied to the log-transforms of the %-rate thresholds (transformed to improve uniformity of the variance), and showed highly significant effects of rate ($F[2,28]=77.7$, $p<0.001$) and listening mode (monaural or dichotic) ($F[1,28]=40.8$, $p<0.001$), as well as their interaction ($F[2,28]=30.8$, $p<0.001$). 5% LSDs indicated significantly lower thresholds for dichotic compared to the monaural listening at both 100 and 200 pps, but *higher* thresholds indicating an adverse interaction between ears at 300 pps. The results show that the time-varying ITDs in the binaural-beat signal were audible at 100 and 200 pps, but not at 300 pps. Although NH listeners show good binaural-beat sensitivity with sinusoids up to 1500 Hz, envelope-beat sensitivity in HF complexes also diminishes beyond 100 or 200 Hz (McFadden and Pasanen, 1975; Bernstein and Trahiotis, 1996; Bernstein and Trahiotis, 2002).

Comparison of the dichotic-signal thresholds in Fig. 5, with the static ITD results in Fig. 1, shows that binaural beats were available over a more limited range of pulse rates than static ITDs. Because performance at 200 and 300 pps, compared to 100 pps, was ratio-metrically higher in the binaural-beat task than with static ITDs, linear scaling of the maximal (or average) ITD in the 300 ms stimuli in the binaural-beat task does not explain the data. To illustrate this, at 100 pps, the median beat-detection threshold of 0.2 Hz corresponds to a maximal ITD of 600 μs (or an average ITD of half that value) at the end of the 300 ms pulse train, which is around four times larger than the subjects' mean static-ITD thresholds at 100 pps in experiment 1A. At 200 pps, static-ITD sensitivity was very similar to that at 100 pps, so that if the cue used in the beat-detection task was linearly related to the maximal available ITD, at threshold we would expect to see the same final ITD of 600 μs, which would require a beat frequency of 0.4 Hz. Instead, the median beat-detection threshold at 200 pps was considerably higher at 2.6 Hz, corresponding to a maximum ITD of 3.9 ms[7] at the end of the 300 ms burst, which is more than an order of magnitude larger than the mean static-ITD sensitivity at that rate. Finally, at 300 pps, binaural cues were not heard at all and performance was in fact worse than monaural rate detection. The fact that faster beat-frequencies are needed at higher stimulation rates to produce the same range of ITDs over a fixed duration, raises the possibility that binaural sluggishness (e.g., Grantham and Wightman, 1978) was responsible for poorer performance as the stimulation rate increased. However, that conjecture is not supported by the observation that 1 Hz binaural beats were clearly audible at 100 pps, but often not at 200 pps (even though the maximal ITD produced by a 1 Hz beat at 200 pps was 1.67 ms, which is considerably larger than the 600 μs cue that allowed beat detection at 100 pps).

One important difference between the binaural-beat and the static-ITD signals is the availability of useful information from the first pulse in the latter, but not the former. Consequently, as rates increased, and ongoing ITDs became more difficult to discern (in accordance with the results from experiments 1A and 1C), the total available cue will have decreased more slowly in the static-ITD discrimination task. However, it is worth noting that, even at 100 pps, for which ongoing cues appear to contribute strongly, the averaged ITD cue over the 300 ms signal duration at beat-detection threshold was several times higher than the static-ITD thresholds, when both measures were averaged across subjects. Perhaps subjects were able to benefit more from "multiple looks" at the same cue size with each pulse pair in the static-ITD signal, whereas in the beat-detection task the time varying ITDs may have provided a more ambiguous cue that caused ongoing information to be weighted less strongly (e.g., Freyman *et al.*, 1997). At higher rates such ambiguity might be more pronounced because comparable ITDs correspond to a larger

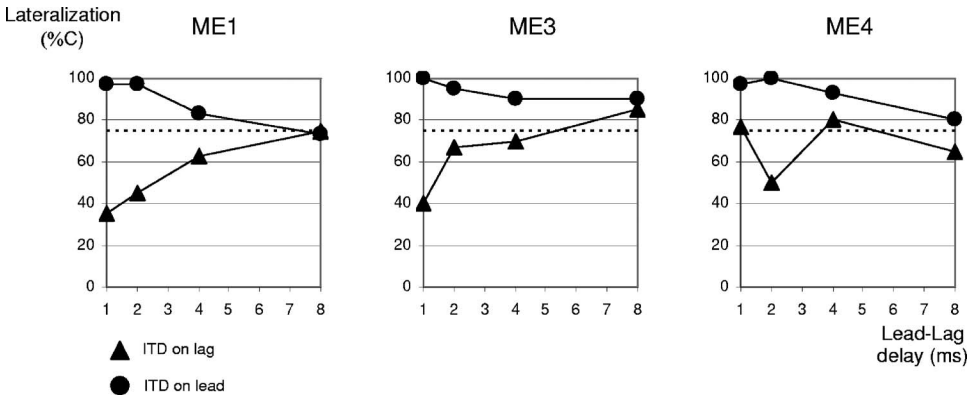


FIG. 6. Percent correct lateralization for ITDs applied to either the lead (circles), or lag (triangles) pulse in the "two-click" stimuli employed in experiment 3, as a function of interclick interval over the range 1–8 ms. Dashed lines indicate the 76%-correct criterion used to determine thresholds in experiments 1 and 2.

fraction of the interpulse interval. Sensitivity to conflicting cues from successive pulses is further considered in the next experiment.

V. EXPERIMENT 3: PRECEDENCE (LAG DISCRIMINATION SUPPRESSION)

A. Methods

A lateralization task was conducted using electrical binaural "click pairs" that consisted of two bilateral pulses, a lead and a lag, separated by an interclick interval (ICI) in the range 1–8 ms. Two experimental conditions were tested. In the first, ITDs were applied to the lead and the lag was diotic. In the second, ITDs were applied to the lag and the lead was diotic. In each experimental trial, two click pairs were presented in random order in a two-interval 2-AFC task. One interval comprised a click pair with a delay applied to the (lead or lag) left-ear signal, and the other to the right. Listeners were required to respond whether the first, or second, click pair was lateralized further to the right. The fixed ITD used in each interval was 400 μs for subjects ME1 and ME3, resulting in a total cue of 800 μs between intervals. For ME4, who showed somewhat higher ITD thresholds, the fixed ITD was increased to 600 μs. For each experimental condition, with ITDs applied to the lead or lag, a single test block included 30 repeat trials with each of four ICIs of 1, 2, 4, and 8 ms, presented in random order (120 trials in total) and without feedback. The same electrodes were used as in experiments 1 and 2, but stimulation levels were increased by 1–3 dB, compared to the 80% DR400 levels in experiment 1, so that these brief stimuli were similar in loudness to that experienced when listening to everyday speech. Stimulation levels were balanced across the ears by ensuring that ITDs applied to left and right components of a single binaural pulse resulted in equidistant lateral shifts toward each ear.

B. Results and discussion

Figure 6 shows percent correct lateralization for the click pairs, as a function of lead-lag separation over the range 1–8 ms. Circles show results for ITDs applied to the lead, and triangles those to the lag. The dashed lines indicate the 76% correct performance criterion used to determine ITD thresholds in experiments 1 and 2. The data were subjected to a GLM ANOVA, specifying ICI, and whether the ITD was applied to the lead or lag, as fixed factors, and subject as a random factor. ITDs applied to the lead were significantly better discriminated than those applied to the lag ($F[1,14]=47.19$, $p<0.001$). The difference was largest at small ICIs, as confirmed by significant interaction between ICI and which click received the ITD ($F[3,14]=5.8$, $p=0.009$). 5% LSDs showed that for ITDs applied to the lag, discrimination was significantly worse at ICIs of 1 ms compared to either 4 or 8 ms, as well as at 2 ms compared to 8 ms. The lag-discrimination data for subject ME4 show unusually good performance at an ICI of 1 ms compared to the other subjects. However, given the same subject's poor discrimination performance at 2 ms, combined with the consideration that this subject showed the most adverse effect of increasing rate on ITD sensitivity in experiment 1, it seems likely that this is due to noise in the data with the limited number of repeats in the present experiment. Unfortunately the subject was not available for retesting. The group trends are in good agreement with a preliminary report from Agrawal *et al.* (2006), who have tested a larger group of CI users with similar electrical signals to those used here. Poorest lag ITD discrimination is also seen as ICIs are reduced to a few ms using click trains in NH listeners (e.g., Zurek, 1980), and is often considered as demonstrative of precedence (Litovsky *et al.*, 1999). Although various aspects of precedence are generally thought to be centrally mediated, Hartung and Trahiotis (2001) have proposed that some of that behavior with successive transients in NH listeners can be explained through the combined action of auditory filtering, hair-cell responses, and binaural cross correlation. In particular, they demonstrate in their model that large interaural cues, which do not necessarily correspond to those imparted to the stimulus, can arise at the outputs of various auditory filters in response to click pairs with short ICIs despite filter ringing times that exceed those ICIs. When further subjected to the compressive effects of the hair cell response and across-frequency correlation, the resulting functions show good agreement with behavioral data from NH listeners that indicate much stronger effects of ITDs applied to the lead than lag at ICIs of a few milliseconds. However, the data in Fig. 6 show comparable lag discrimination suppression with electrical stimulation, despite the absence of both auditory filters and hair cell responses. It may be the case that peripheral effects specific to electrical stimulation coincidentally lead to results that are similar to those seen in NH listeners. Alternatively, a presumably more central, common mechanism unrelated to

NH mechanical transduction properties may be responsible for both CI and NH results. Although lag discrimination suppression is more frequently described in the context of precedence, the reduced contribution of ITDs on pulses beyond the first at shorter ICIs is also in good agreement with the binaural-beat data from experiment 2. Both tasks required an ability to hear ITD cues on pulses following a diotic onset. The binaural-beat data showed that beats could only be heard at 100 pps, or sometimes 200 pps, for which ICIs are 10 and 5 ms, respectively. Correspondingly, the results from the present experiment show ITDs on the lag could be best perceived at the larger ICIs of 4–8 ms. At shorter ICIs lag-ITD discrimination was poor and binaural beats in experiment 2 were inaudible.

When ITDs were applied to the lead instead of the lag, performance was less affected by the ICI. At intervals of 1 ms performance was better than at 8 ms, but the difference was not quite statistically significant (difference of means 17%, LSD=18%). A similar effect can be seen with NH listeners in the data from Tollin and Henning (1998). It may be the case that, at the largest ICIs, the task was made more difficult due to fusion breakdown. Although none of the CI users reported difficulties lateralizing due to presence of multiple sounds, when ME1 was asked to subjectively describe the effect of ICI in a separate test, he commented that the sound became more "echoic" at 8 ms than at shorter intervals. The possibility that lateralization was made more difficult by decreased fusion at large ICIs does not affect the main finding that discrimination of ITDs on the *lag* was poorest at small ICIs.

The results from this experiment suggest that in a reverberant situation, CI users who are able to hear onset ITDs should be able to focus on that cue despite conflicting cues from later arriving reflections. From a sound-processing point of view, the finding that information on the lag had little effect at ICIs as short as a millisecond, implies that binaural strategies for CI processors might benefit from using filters with time-averaging windows no longer than that to prevent smearing of temporal cues from the onset and later portions of the signal. Although ringing times of auditory filters in NH listeners can be longer than this, as shown by Hartung and Trahiotis (2001) the interaction of stimulus ITD and ICI with filter responses at multiple frequencies can largely alleviate the adverse effects of those long ringing times. It is unlikely that a similar mechanism can be exploited in present CI processors because of the relatively limited number of fixed filters, as well as substantial interaction between sites of stimulation due to electrical current spread. The finding that at an ICI of 1 ms lag-ITDs were very difficult to hear, but lead ITDs were readily discriminated, implies that perceptual integration of the electrical onset-ITD cue had an associated time constant no longer than 1 ms, otherwise ITD information would have been averaged across the two pulses and ITDs would have been equally effective when applied to either pulse. Such a short integration time for the onset cue supports the use of the ITD threshold for the single-pulse stimulus in experiment 1C as an indication of the information available from the first pulse in longer duration stimuli, and therefore the conclusion that ongoing ITDs were available up to at least 400 pps for all three subjects because ITD thresholds decreased at longer durations. It also indicates that in experiment 1A, the reason ME3 could hear ITDs at 600 pps, but not 1000 pps, was probably not because of smearing of the onset cue at 1000 pps, and that ongoing ITD sensitivity was indeed available up to 600 pps (but not 1000 pps).

VI. CONCLUSIONS

The data from these experiments indicate that, at 100 pps, ITDs on each pulse in the signal contributed substantially to overall performance, whereas at higher rates, the cues from pulses beyond the first (onset) became more difficult to discern. When the first pulse was diotic and ITD cues on subsequent pulses differed, none of the three subjects was able to hear ITDs at rates as low as 300 pps in the binaural-beat detection task, and lag ITD discrimination was very poor in the precedence experiment when the two pulses were separated by only a few milliseconds. In contrast, for static-ITD signals comprising identical cues on all pulses, including the first, ITD sensitivity was much less affected over that range of rates. Although this was presumably at least in part because the onset cue remained available in the static-ITD signal as rates increased, comparison between static and time-varying ITD sensitivity at 100 pps, for which ongoing cues were well perceived, indicates that other factors such as cue ambiguity may also play a role. Results from experiment 1C show that contributions from ongoing static ITD cues were available up to at least 400 pps for all three subjects, because static-ITD thresholds decreased as stimulus duration increased at both 100 and 400 pps. In experiment 1A, subject ME3 showed evidence of ongoing ITD sensitivity up to 600 pps, because thresholds remained relatively unchanged up to that rate, but increased substantially at 1000 pps. The finding that sensitivity to static-ITD cues was preserved at rates for which monaural rate-pitch discrimination in experiment 1B was very poor was probably also affected by the availability of the onset cue in the ITD task, so that result alone does not rule out a common peripheral effect as the primary limitation in both tasks. However, the evidence that ongoing ITD cues were available up to at least 400 pps for all three subjects, and 600 pps for ME3, suggests that monaural rate discrimination may be subject to further constraints that do not affect ITD sensitivity.

Static ITD sensitivity for these three subjects, with low-rate pulse trains at 100 pps, or with high-rate AM pulse trains with 100 Hz modulation, was comparable to thresholds at 50–100 pps in better performing bilateral CI subjects in previous reports. Although these thresholds are somewhat higher than those reported by Hafter and Dye (1983) for NH listeners with filtered click trains, that may be partly due to the incomplete removal of low frequency cues in the NH study. The effect of rate on ITD sensitivity in the present study confirms the report by van Hoesel and Tyler (2003), although the effect of level found in experiment 1 demonstrates that the reduced levels used at higher rates in that earlier study may also have played a role. Elevated ITD thresholds for pulse rates above a few hundred pps have

generally been confirmed in recent studies by Laback *et al.* (2005), Jones *et al.* (2006), and Majdak *et al.* (2006). However, as in the present work, some subjects in those studies also showed sensitivity to ITDs at rates as high as 600 or 800 pps. This was the case despite the use of slow rise times to reduce onset cues in the studies by Jones *et al.* and Majdak *et al.*, or diotic onsets in the study by Laback *et al.*

Sensitivity to ongoing ITDs with electrical stimuli therefore is available well beyond the 150 Hz limitation sometimes described for NH listeners with HF AM signals with slow rise times. However, some NH subjects also appear to hear ongoing ITDs beyond that limit (e.g., Bernstein and Trahiotis, 2002), so it is unclear whether electrical stimulation differs in this regard from acoustic stimulation with HF signals. In contrast, it is clear that the *improvement* in ITD thresholds seen with increasing pure tone frequencies up to about 1000 Hz in NH listeners is not available with electrical pulse trains, which demonstrates that those electrical stimuli are unable to convey important timing relations that are available to NH listeners with low-frequency sinusoids. This result is unlikely to stem just from insufficiently apical stimulation with the electrical stimuli, as systematic variation of ITD sensitivity with bilateral place of stimulation has not been observed in any CI studies to date, and the most apical electrodes in present electrode arrays readily reach to places associated with 1000 Hz or lower in NH listeners. Instead, it seems more likely due to differences such as the synchronous response to electrical stimulation across a relatively broad region of the cochlea, associated refractory behavior potentially compounded by the effects of deafness, and distorted place-rate relations in comparison to the specific phase-locked responses to low frequency sinusoids traveling along the basilar membrane in normal hearing.

The reduced contribution from ITDs beyond the onset at higher rates, shown in experiment 1C, as well as the limited range of rates over which binaural beats were heard in experiment 2, are in good agreement with binaural adaptation effects at rates above 100 or 200 Hz in NH listeners (Hafter and Dye, 1983; Saberi, 1996; Stecker and Hafter, 2002). Similarly, in experiment 3, the finding that ITD cues on the lag were most difficult to discriminate when ICIs were reduced to a few milliseconds is comparable to similar measures in NH listeners (e.g., Zurek, 1980), and indicates that NH auditory filter and hair-cell mechanisms are not prerequisite to evoke lag discrimination suppression at those ICIs. Although auditory filter and hair cell responses can alleviate the adverse effects of long auditory-filter ringing times in NH listeners (Hartung and Trahiotis, 2001), the use of similar processes in CI sound processors is likely precluded by comparatively coarse signal representations and broad regions of activation along the cochlea. Consequently, shorter time constants than those associated with auditory filters in NH subjects may be required in CI strategies. According to the present studies, ringing times no longer than a millisecond may be beneficial.

When electrical pulse-trains at 6000 pps were amplitude modulated at 100 Hz in experiment 1D, thresholds were comparable to those found using 100 pps unmodulated pulse trains. However, at modulation rates between 200 and 400 Hz, performance was worse than with unmodulated pulse trains at those rates. Recent data from Jones *et al.* (2006) similarly show better performance with unmodulated pulse trains than AM signals at 300 Hz, but not at 100 Hz. Factors contributing to elevated thresholds with the modulated signals may include the lower level and weaker onset cues with the AM signals, as well as poor envelope representation as a result of elevated refractory effects associated with the 6000 pps carrier. The better performance with unmodulated pulse-trains compared to AM signals at 200 and 300 Hz suggests that these subjects would obtain more salient ITD cues for frequencies in this range by using a fine-timing based strategy, such as PDT, than envelope based strategies. However, interaction between nearby electrodes in multichannel stimulation may need to be alleviated before practical benefits ensue.

ACKNOWLEDGMENTS

The author is particularly indebted to the three research volunteers who generously donated their considerable time and effort in completing this work. The author is also grateful for the insightful comments made by two anonymous reviewers and Andrew Oxenham, as well as helpful discussions with Chris Stecker at The University of Washington, Andrew Vandali and Laurie Cohen at CRC-Hear in Melbourne, Ruth Litovsky, Gary Jones, and Smita Agrawal at the University of Wisconsin, and Zach Smith at MIT. Organizational and technical support was gratefully received from Bob Cowan and Mark Harrison, respectively. This work was funded by the CRC for Cochlear Implant and Hearing Aid Innovation (CRC-Hear) in Melbourne, Australia.

[1]Although Seeber *et al.* (2004) argue that the large interloudspeaker spacing used by van Hoesel and Tyler (2003) may have resulted in an underestimating rms errors near 0°, data from van Hoesel (2004) with much reduced loudspeaker spacing shows similar performance.

[2]Consider the case where the first pulse is placed at the minimum of the modulation cycle in both ears. Any time delay applied to the modulator or carrier alone in one ear can only increase that pulse's stimulation level, irrespective of whether that level change is in agreement with, or counter to the ITD cue between the ears. Simular arguments apply to all the pulses in the stimulus, and particularly with harmonically related modulator and pulse rates this can lead to confounding level cues.

[3]In contrast, a starting phase of 90° would result in a conflict between the interpolated instantaneous envelope ILD, and ITD cues at the onset of the signal. When considering the ILD of paired, time-delayed pulses, the onset ILD remains zero irrespective of the modulator starting phase.

[4]By way of example, for an ITD threshold of 100 μs at 100 pps, the "equivalent rate interval" changes from 10 ms without the ITD, to 9.9 ms when the ITD threshold is subtracted, so that the rate changes by about 1%.

[5]An additional analysis with subject, rather than ear, as the random factor resulted in even higher levels of significance with regard to the effect of rate.

[6]Because only rates higher than the reference were used to generate beats, and the duration of 300 ms is an integer multiple of all three reference-rate intervals, as the beat frequency increased from 0 Hz, the stimulus on the high-rate side finished earlier than the fixed-rate side until the beat was 3.33 Hz, for which an extra pulse was added at exactly 300 ms, etc.

[7]Note that for a 2.6 Hz beat the ongoing ITD cue at the end of the 300 ms pulse train exceeds half the 200 pps rate-interval, presenting potentially confounding ongoing lateralization cues. However, since the task involved comparison with a diotic condition rather than a lateral-position judgment per se, this may or may not have affected performance. Although this does raise the possibility that the maximal unambiguous ITD presented in that

case was 2.5 ms, rather than 3.9 ms, the arguments presented in the text remain unaffected.

Agrawal, S. S., Litovsky, R. Y., Jones, G. J., and van Hoesel, R. J. M. (**2006**). "Correlates of precedence effect in bilateral cochlear implant users under the conditions of direct stimulation and in free-field," Association for Research in Otolaryngology, 29th Midwinter Meeting, Baltimore, MD.

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. T., and Trahiotis, C. (**1996**). "Binaural beats at high frequencies: Listeners' use of envelope based interaural temporal and intensive disparities," J. Acoust. Soc. Am. **99**, 1670–1679.

Carlyon, R. P., and Deeks, J. M. (**2002**). "Limitations on rate discrimination," J. Acoust. Soc. Am. **112**, 1009–1025.

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.

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype perimodiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Cohen, L. T., Saunders, E., Knight, M. R., Cowan, R. S. C. (**2006**). "Psychophysical measures in patients fitted with Contour™ and straight nucleus electrode arrays," Hear. Res. **212**, 160–175.

Durlach, N. I., and Colburn, H. S. (**1978**). "Binaural phenomena," in *Handbook of Perception* (Academic, New York), Vol. **IV**, Chap. 10, pp. 373–406.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Foster, D. H., and Bischof, W. F. (**1991**). "Thresholds from psychometric functions: Superiority of bootstrap to incremental and probit variance estimators," Psychol. Bull. **109**, 152–159.

Freyman, R. L., Zurek, P. M., Balakrishan, Y., and Chiang, Y. C. (**1997**). "Onset dominance in lateralization," J. Acoust. Soc. Am. **101**, 1649–1659.

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. P. (**2002**). "Binaural cochlear implants placed during the same operation," Otol. Neurotol. **23**, 169–180.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Hafter, E. R., Buell, T. N., and Richards, V. N. (**1988**). "Onset-coding in lateralization: Its form, site and function," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 648–676.

Hafter, E. R., and Dye, R. H. Jr. (**1983**). "Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number," J. Acoust. Soc. Am. **73**, 644–651.

Hartmann, W. M., and Rakerd, B. (**1989**). "On the minimum audible angle-A decision theory approach," J. Acoust. Soc. Am. **85**, 2031–2041.

Hartung, K., and Trahiotis, C. (**2001**). "Peripheral auditory processing and investigations of the 'precedence effect' which utilize successive transient stimuli," J. Acoust. Soc. Am. **110**, 1505–1513.

Houtgast, T., and Plomp, R. (**1968**). "Lateralization threshold of a signal in noise," J. Acoust. Soc. Am. **44**, 807–812.

Javel, E., and Shepherd, R. K. (**2000**). "Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure," Hear. Res. **140**, 45–76.

Jones, G. J., Litovsky, R. Y., Agrawal, S. S., and van Hoesel, R. J. M. (**2006**). "Effect of stimulation rate and modulation rate on ITD sensitivity in bilateral cochlear implant users," Association for Research in Otolaryngology, 29the Midwinter Meeting, Baltimore, MD.

Klumpp, R. G., and Eady, H. R. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Laback, B., Majdak, P., and Baumgartner, W. (**2005**). "Interaural time differences in temporal fine structure, onset, and offset in bilateral electrical stimulation," Association for Research in Otolaryngology, 28th Midwinter Meeting, New Orleans, LA.

Lawson, D. T., Wilson, B. S., Zerbi, M., van den Honert, C., Finley, C. C., Farmer, J. C. Jr., McElveen, J. T., and Rousch, P. A. (**1998**). "Bilateral cochlear implants controlled by a single speech processor," Am. J. Otol. **19**, 758–761.

Lawson, D. T., Wolford, R., Brill, S., Schatzer, R., and Wilson, B. S. (**2001**). "Speech processors for auditory prosthesis," 12th Quarterly Progress Report, NIH Project N01-DC-8-2105.

Licklider, J. C. R., Webster, J. C., and Hedlun, J. M. (**1950**). "On the frequency limits of binaural beats," J. Acoust. Soc. Am. **22**, 468–473.

Litovsky, R. Y., Agrawal, S. S., Jones, G. J., Henry, B., and van Hoesel, R. J. M. (**2005**). "Effect of interaural electrode pairing on binaural sensitivity in bilateral cochlear implant users," Association for Research in Otolaryngology, 28th Midwinter Meeting.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., and Gongqiang, Y. (**2004**). "Bilateral cochlear implants in adults and children," Arch. Otolaryngol. Head Neck Surg. **130**, 648–655.

Loeb, G. E. (**2005**). "Are cochlear implant patients suffering from perceptual dissonance?," Ear Hear. **26**, 435–450.

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.

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (**2003**). "Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user," J. Acoust. Soc. Am. **114**, 1565-1574.

Loquet, G., Pelizzone, M., Valentini, G., and Rouiller, E. M. (**2004**). "Matching the neural adaptation in the rat ventral cochlear nucleus produced by artificial (electric) and acoustic stimulation of the cochlea," Audiol. Neuro-Otol. **9**, 144–159.

Majdak, P., Laback, B., and Baumgartner, W.-D. (**2006**). "Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing," J. Acoust. Soc. Am. **120**, 2190–2201.

McFadden, D., and Pasanen, E. G. (**1975**). "Binaural beats at high frequencies," Science **190**, 394–396.

McKay, C. M., and McDermott, H. J. (**1996**). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites," J. Acoust. Soc. Am. **100**, 1081–1092.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude modulated current pulse-trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

Müller, J., Schön, F., and Helms, J. (**2002**). "Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system," Ear Hear. **23**, 198–206.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Saberi, K. (**1996**). "Observer weighting of interaural delays in filtered impulses," Proc. Natl. Acad. Sci. U.S.A. **58**, 1037–1046.

Seeber, B. U., Baumann, U., and Fastl, H. (**2004**). "Localization ability with bimodal hearing aids and bilateral cochlear implants," J. Acoust. Soc. Am. **116**, 1698–1709.

Shannon, R. P. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.

Shepherd, R. K., Roberts, L. A., and Paolini, A. G. (**2004**). "Long-term sensorineural hearing loss induces functional changes in the rat auditory nerve," Eur. J. Neurosci. **20**, 3131–3140.

Smith, Z. M. (**2006**). "Binaural interactions in the auditory midbrain with electrical stimulation of the cochlea," Ph.D. thesis, Harvard-MIT Division of Health Sciences and Technology, Boston, MA.

Stecker, G. C., and Hafter, E. R. (**2002**). "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057.

Tollin, D. J., and Henning, G. B. (**1998**). "Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect," J. Acoust. Soc. Am. **104**, 3030–3036.

Tong, Y. C., and Clark, G. M. (**1985**). "Absolute identification of electric pulse rates and electrode positions by cochlear implant listeners," J. Acoust. Soc. Am. **77**, 1881–1888.

Tyler, R. S., Gantz, B. J., Rubinstein, J. T., Wilson, B. S., Parkinson, A. J., Wolaver, A., Preece, J. P., Witt, S., and Lowder, M. W. (**2002**). "Three-month results with bilateral cochlear implants," Ear Hear. **23**, 80S–89S.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," Percept. Psychophys. **101**, 1671–1680.

van Hoesel, R., Ramsden, R., and O'Driscoll, M. (**2002**). "Sound-direction identification, interaural time delay discrimination and speech intelligibility advantages in noise for a bilateral cochlear implant user," Ear Hear. **23**, 137–49.

van Hoesel, R. J. M. (**2004**). "Exploring the benefits of bilateral cochlear implants," Audiol. Neuro-Otol. **9**, 234–246.
van Hoesel, R. J. M., and Clark, G. M. (**1997**). "Psychophysical studies with two binaural cochlear implants subjects," J. Acoust. Soc. Am. **102**, 504–518.
van Hoesel, R. J. M., and Clark, G. M. (**1999**). "Speech results with a bilateral multichannel cochlear implant for spatially separated signal and noise," Aust. J. Audiol. **21**, 23–28.
van Hoesel, R. J. M., Tong, Y. C., Hollow, R. D., and Clark, G. M. (**1993**). "Psychophysical and speech perception studies: A case report on a bilateral cochlear implant subject," J. Acoust. Soc. Am. **94**, 3178–3189.
van Hoesel, R. J. M., and Tyler, R. S. (**2003**). "Speech perception and localization with bilateral cochlear implants," J. Acoust. Soc. Am. **113**, 1617–1630.
Wilson, B. S., Finley, C. C., Lawson, D. T., and Zerbi, M. (**1997**). "Temporal representations with cochlear implants," Am. J. Otol. **18**, S30–34.
Xu, J., Xu, S. A., Cohen, L. T., and Clark, G. M. (**2000**). "Cochlear view: Post-operative radiology for cochlear implantation," Am. J. Otol. **21**, 49–56.
Zeng, F.-G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.
Zurek, P. M. (**1980**). "The precedence effect and its possible role in the avoidance of interaural ambiguities," J. Acoust. Soc. Am. **67**, 952–964.

Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies

Yuxuan Zhang[a)]
Northwestern University Institute for Neuroscience, Northwestern University, Evanston, Illinois 60208

Beverly A. Wright
Department of Communication Sciences and Disorders, and Northwestern University Institute for Neuroscience, Northwestern University, Evanston, Illinois 60208

(Received 20 April 2005; revised 13 December 2006; accepted 21 December 2006)

Sound source localization on the horizontal plane is primarily determined by interaural time differences (ITDs) for low-frequency stimuli and by interaural level differences (ILDs) for high-frequency stimuli, but ITDs in high-frequency complex stimuli can also be used for localization. Of interest here is the relationship between the processing of high-frequency ITDs and that of low-frequency ITDs and high-frequency ILDs. A few similarities in human performance with high- and low-frequency ITDs have been taken as evidence for similar ITD processing across frequency regions. However, such similarities, unless accompanied by differences between ITD and ILD performance on the same measure, could potentially reflect processing attributes common to both ITDs and ILDs rather than to ITDs only. In the present experiment, both learning and variability patterns in human discrimination of ITDs in high-frequency amplitude-modulated tones were examined and compared to previously obtained data with low-frequency ITDs and high-frequency ILDs. Both patterns for high-frequency ITDs were more similar to those for low-frequency ITDs than for high-frequency ILDs. These results thus add to the evidence supporting similar ITD processing across frequency regions, and further suggest that both high- and low-frequency ITD processing is less modifiable and more noisy than ILD processing. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2434758]

PACS number(s): 43.66.Pn, 43.66.Qp [GDK] Pages: 2207–2216

I. INTRODUCTION

One of the key functions of sensory systems is to encode the spatial positions of environmental stimuli. In the auditory system, stimulus localization on the horizontal plane is primarily accomplished through the calculation of two disparities in the sound arriving at the two ears from a single source: interaural time differences (ITDs) and interaural level differences (ILDs). ITDs arise because a sound reaches the ear farther from the sound source slightly later than it reaches the closer ear. ILDs are generated by the sound shadow formed by the head, which attenuates the sound level reaching the farther ear more than that reaching the closer ear. Sound source position is primarily cued by ITDs for low-frequency sounds, and by ILDs for high-frequency sounds (Rayleigh, 1907; Feddersen *et al.*, 1957; Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). However, humans are also sensitive to changes in ITDs in high-frequency sounds with complex wave forms (Klumpp and Eady, 1956; Leakey *et al.*, 1958; Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976; Henning, 1980; Henning and Ashton, 1981; Bernstein and Trahiotis, 1994). Though such high-frequency ITDs are not usually the dominant localization cue (Blauert, 1982; Middlebrooks and Green, 1990; Wightman and Kistler, 1992; Eberle *et al.*, 2000; Macpherson and Middlebrooks, 2002), they can be used to separate concurrent sound sources, thereby aiding hearing in noise (Best *et al.*, 2005). Also, hearing-impaired listeners retain a relatively intact ability to detect ITDs in the amplitude envelope of high-frequency complex sounds (Buus *et al.*, 1984; Smoski and Trahiotis, 1986) while their ability to use ITDs in the fine structure of low-frequency sounds is compromised(Hawkins and Wightman, 1980; Buus *et al.*, 1984; Lacher-Fougere and Demany, 2005). Here, our goal was to gain insight into the processing of high-frequency ITDs by comparing behavioral performance on high-frequency ITD discrimination to that on low-frequency ITD and high-frequency ILD discrimination.

A few previous investigations have revealed similarities between high- and low-frequency ITDs in terms of both physiological encoding and behavioral performance, suggesting similar neural processing for ITDs across frequency regions. Physiologically, neurons sensitive to high-frequency ITDs are found in the medial superior olive, the nucleus traditionally regarded as being responsible for the initial encoding of low-frequency ITDs (Spitzer and Semple, 1995; Batra *et al.*, 1997a, b). These neurons have tuning characteristics to both monaural and binaural stimuli that are similar to those of neurons sensitive to low-frequency ITDs. Behaviorally, human listeners show similar sensitivity to high- and low-frequency ITDs when the envelopes of high-frequency stimuli are manipulated to produce neural representations in

[a)]Electronic mail: y-zhang6@northwestern.edu

auditory nerve fibers that are similar to those produced by low-frequency sounds (Bernstein, 2001; Bernstein and Trahiotis, 2002, 2003). In addition, similarities between low- and high-frequency ITD discrimination have also been observed in how performance changes with increases in signal duration (McFadden and Moffitt, 1977) and stimulus amplitude (Nuetzel and Hafter, 1976, but see McFadden and Pasanen, 1976). Overall, these results are consistent with Colburn and Esquissaud's (1976) proposal that, beyond the peripheral monaural channels in which temporal information is represented differently for low- and high-frequency sounds, ITDs are processed by the same mechanisms regardless of stimulus frequency.

Here, we further investigated the processing of high-frequency ITDs using two behavioral metrics that have not been examined before: patterns of training-induced learning and patterns of threshold variability. Our motivation for choosing these two metrics was twofold. First, these two performance patterns afford direct comparisons of ITD to ILD performance, an important component in testing the idea that low- and high-frequency ITDs are processed by similar neural mechanisms. There is considerable evidence that low-frequency ITDs and high-frequency ILDs are encoded by separate mechanisms (Irvine, 1986; Griffiths *et al.*, 1998). Thus, to ensure that an observed similarity between high- and low-frequency ITD performance reflects similar ITD processing across frequency regions rather than a feature common to the processing of both ITDs and ILDs, that similarity should be accompanied by a difference between ITD and ILD performance on the same measure. This comparison is particularly desirable because there is evidence that the majority of ILD-encoding neurons in the lateral superior olive, the nucleus traditionally regarded as the first encoding stage of ILDs, also respond to high-frequency ITDs, suggesting a possible contribution of high-frequency ILD mechanisms to high-frequency ITD sensitivity (Joris and Yin, 1995). However, we are aware of no direct comparison of high-frequency ITD and ILD discrimination performance, perhaps because of the complication of the different units used to measure threshold. The performance metrics used here, improvement patterns and variability patterns, instead describe changes or distributions of performance, and are thus directly comparable across ITD and ILD discrimination.

Second, both learning and performance variability have been used to reveal specific characteristics about the neural processes underlying behavior, and have been shown to differ for low-frequency ITD and high-frequency ILD discrimination. Patterns of learning have been used to gain insight into the organization and function, in addition to the modifiability, of the neural circuits governing performance on the trained task (e.g., (Karni and Sagi, 1991; Wright *et al.*, 1997; Demany and Semal, 2002; Fitzgerald and Wright, 2005). Consistent with the separate neural mechanisms for low-frequency ITDs and high-frequency ILDs, learning patterns differ between the two cues. Wright and Fitzgerald (2001) reported that, under identical training regimens, mean thresholds on low-frequency ITD discrimination reached asymptote by the end of an initial 2-h testing session, and subsequent multihour training did not add to the improvement. In contrast, for high-frequency ILD discrimination, listeners who received multiple hours of training improved significantly more than untrained controls on both the trained condition and an untrained condition with a different standard ILD. Performance variability can also reveal key aspects of neural processing by reflecting the internal noise in a neural process within individuals (e.g., Brunt *et al.*, 1983; Takahashi *et al.*, 2003) and the variation of that process across individuals. As in the case of learning, differences in the patterns of variability across tasks indicate differences in the underlying neural processes. Again consistent with the separate neural mechanisms for low-frequency ITDs and high-frequency ILDs, a reanalysis of Wright and Fitzgerald's data (reported in Sec. III) revealed different variability patterns for the two cues.

Given these cue-dependent patterns of performance, we reasoned that learning and variability could be used as two behavioral metrics to compare the processing of high-frequency ITDs to that of low-frequency ITDs and that of high-frequency ILDs. In such comparisons, a difference would signal a mismatch in the two processes involved, while a similarity would suggest, if not a common process, at least a common feature shared by the two processes. To enable this comparison, we evaluated performance on ITD discrimination with high-frequency amplitude-modulated tones in a total of 19 listeners using the same paradigm as Wright and Fitzgerald (2001).

II. METHODS

A. Organization of the experiment

We examined learning and performance variability on ITD discrimination with amplitude-modulated tones using a training paradigm identical to the one previously employed for ITD and ILD discrimination with pure tones (Wright and Fitzgerald, 2001). The experiment consisted of a pretest session, nine training sessions, and a posttest session. Each session took place on consecutive days (except weekends). All listeners participated in the pre- and posttests. Between these tests, only a subset of randomly assigned listeners, referred to as trained listeners, participated in the training sessions. The remaining listeners, referred to as controls, did not receive training.

In the pre- and posttests we tested each listener on five related discrimination conditions. During the training sessions, the trained listeners practiced on only one of the five pre- and posttest conditions, ITD discrimination with a 4-kHz carrier modulated at 0.3 kHz, referred to as the trained condition. We chose this combination of carrier and modulator because it was very similar to the stimuli previously reported to have yielded among the lowest discrimination thresholds with amplitude-modulated tones (Henning, 1974; Nuetzel and Hafter, 1976). The other four pre- and posttest conditions, referred to as untrained conditions, differed from the trained condition in either the carrier frequency (6 versus 4 kHz), the modulation rate (0.15 versus 0.3 kHz), the waveform type (a 0.3-kHz pure tone versus amplitude-modulated tones), or the interaural cue manipulated (ILD versus ITD). The testing order of the five conditions was randomized

across listeners, but was constant for each listener across the pre- and posttests. For each condition in the pre- and posttests, we collected five consecutive threshold estimates. During each of the nine training sessions, the trained listeners completed twelve threshold estimates (~1 h) on the trained condition only.

B. Task

We measured the ability of listeners to discriminate different values of ITD or ILD. To have independent control of these two cues, the stimuli were presented over headphones. Consequently, the listeners perceived a sound image inside the head, the lateral position of which was determined by the ITD and ILD in the stimulus. The stimuli were presented in a two-interval-forced-choice paradigm. Each trial consisted of two visually marked 300-ms observation intervals that were separated by a 660-ms silent period. In each interval, stimuli with the same spectra were presented to both ears. Across intervals, we manipulated only the ITD or ILD of the stimuli, which resulted in changes in the perceived lateral position of the sound image. In one interval randomly chosen on each trial, a standard stimulus was presented with an ITD of 0 μs and an ILD of 0 dB, placing the sound image approximately in the middle of the head. In the other interval, a signal stimulus was presented that differed from the standard stimulus only by a variable ΔITD or ΔILD that always favored the right ear (see also Wright and Fitzgerald, 2001). The listeners reported which interval contained the signal stimulus by pressing a key on a computer keyboard. Visual feedback was provided after each response throughout the entire experiment. To familiarize the listeners with the lateralization task, before starting each condition we asked them to indicate the lateral position of the perceived image of a repeatedly presented standard stimulus on a schematic diagram of a human head, and to report, verbally or visually, the distinctive lateral positions of samples of the standard and signal stimuli. The listeners were also instructed to listen to these samples before they started each 60-trial block throughout the whole experiment.

C. Procedure

A discrimination threshold for ITD or ILD was estimated from each 60-trial block using a 3-down-1-up adaptive procedure. Within each block of trials, we adjusted the ΔITD or ΔILD adaptively by decreasing its value after every three consecutive correct responses and increasing its value after each incorrect response (Levitt, 1971). The signal levels at which the direction of change switched from decreasing to increasing or from increasing to decreasing were denoted as reversals. The value of ΔITD or ΔILD for 79% correct responses, called threshold, was estimated by taking the average of an even number of reversals after discarding the first three (if the total number of reversals was odd) or four (if the total number of reversals was even) reversals. If there were fewer than four reversals left in a single 60-trial block after discarding the first three or four reversals, no threshold estimate was calculated, and the performance on that block was marked as "insufficient reversals." In the ITD conditions, the starting value of ΔITD was 1 μs, and the steps were multiplications or divisions by $10^{0.2}$ until the third reversal and by $10^{0.05}$ thereafter (Saberi, 1995). The maximum value of ΔITD was 650 μs, approximately the largest naturally occurring time delay between the two ears in humans (e.g., Feddersen *et al.*, 1957;Kuhn, 1977). In the ILD conditions, ΔILD started at 6 dB, was adjusted with a step size of 0.5 dB until the third reversal and 0.25 dB thereafter, and was never allowed to go below 0 dB. The starting values and step sizes for ΔITD and ΔILD were the same as those used in our previous experiments (Wright and Fitzgerald, 2001), and the differences in them between the two cue types appear to have had no influence on the effectiveness of threshold estimation (see Sec. II F).

D. Stimuli

Both pure tones and sinusoidally amplitude modulated (SAM) tones were used as stimuli. The SAM tones were synthesized by sinusoidally modulating the amplitude of a sinusoidal carrier to 100% depth. In the ITD conditions, we set the desired ongoing ITD by delaying the starting phase of the pure tone, or of both the carrier and the modulator for the SAM tones, in the left-ear stimulus relative to that in the right-ear stimulus (e.g., Henning, 1974). The sound level of the pure tones, or of the SAM tones before modulation, was 50 dB SPL to each ear. This sound level was low enough to avoid the influence of combination products in the SAM-tone stimuli (Plomp, 1965). In the ILD condition, the sound level was 50 dB SPL minus 0.5 times the desired ILD for the left-ear stimulus, and 50 dB SPL plus 0.5 times the desired ILD for the right-ear stimulus. The time structure of the whole wave forms was identical at the two ears. In all conditions, the stimuli to both ears started and ended simultaneously, and each stimulus had a total duration of 300 μs, including 10-ms cosine rise/fall ramps. The starting phase of the right-ear stimulus was randomized across intervals for pure tones. For SAM tones, the starting phases of both the carrier and modulation wave forms to the right ear were randomized across intervals, and were independent of each other.

We used a digital-signal processing board (Tucker-Davis Technologies, Gainesville, FL, AP2) to generate all stimuli. The stimuli to each ear were then delivered through separate 16-bit digital-to-analog converters (TDT DD1), antialiasing filters (8.5-kHz low-pass, TDT FT5), and programmable attenuators (TDT PA4). Finally, the stimuli were sent through a headphone buffer (TDT HB6) to headphones with circumaural cushions (Sennheiser, HD265). Listeners were tested in a sound-attenuated booth.

E. Listeners

Nineteen normal-hearing human volunteers (fourteen women) between the ages of 19 and 31 years (average of 23.1 years) served as listeners. Eight listeners were trained, and the remaining eleven were controls. All were paid for their participation. None of the listeners had previous experience in any psychoacoustic experiment.

F. Two issues in threshold estimation

Two issues in threshold estimation deserve comment. First, all of the listeners had considerably more difficulty discriminating high-frequency ITDs than low-frequency ITDs and high-frequency ILDs. As a consequence, quite often (18.8% out of the total number of threshold estimates) the value of ΔITD reached the 650 μs maximum before the end of a 60-trial block. Nevertheless, listeners were allowed to finish the full 60-trial block in order to gain exposure to the same number of trials as in a normal block. In such blocks, when the adaptive track called for a ΔITD value higher than 650 μs, we assigned that nominal value to the track, but kept the actual ΔITD value at 650 μs. When this happened, there usually were not enough reversals to yield a valid threshold estimate, and the block was noted as having "insufficient reversals" (15.5% out of the total number of threshold estimates). In a few of these cases there were enough reversals for threshold estimation, but the threshold estimate was higher than 650 μs, because they were calculated based on the nominal ΔITDs (3.3% out of the total number of threshold estimates). Because both of these circumstances reflected, to some degree, the listeners' inability to obtain a valid threshold estimate below the 650 μs limit, excluding these cases from the analyses would lead to an underestimation of average threshold. To help reduce this underestimation, in these instances we reevaluated these threshold estimates by alternative means. For those with insufficient reversals, we recalculated threshold estimates if there were at least six reversals in the block by averaging the last four reversals, and thus excluded only the very few cases that had fewer than six reversals ($<0.1\%$). The threshold estimates that were higher than 650 μs were replaced with 650 μs. We analyzed three different versions of the data set: one with insufficient reversals recalculated and threshold estimates higher than 650 μs replaced with 650 μs, another with both types of irregular thresholds excluded, and a third with both replaced with 650 μs. Further, each statistical analysis was performed both on the original data and its log transformation. All of our conclusions were the same for all analyses. We have chosen to report the reevaluated values because they appear to best represent the listeners' actual ability.

Second, although we used starting values in the adaptive tracks that were below the estimated thresholds for the ITD conditions (starting at 1 μs) and above them for the ILD conditions (starting at 6 dB), that difference appeared to have little influence on the ability of the adaptive procedure to estimate discrimination threshold. We analyzed all of the adaptive tracks from the pretests, the training phase, and the posttests separately for the high-frequency ITD (1403 tracks), low-frequency ITD (1445 tracks), and high-frequency ILD conditions (1967 tracks), emphasizing the reversals that were used for threshold estimation (referred to as usable reversals). On average, the first usable reversal occurred at the 32nd trial for the high-frequency ITD conditions, and the 29th trial for both the low-frequency ITD and the ILD conditions. Correspondingly, the average number of all usable reversals was 8.8, 10.3, and 10.6 in the three cases. Thus, the threshold estimates in the ITD and ILD conditions were based on approximately equal numbers of usable reversals, obtained over approximately equal numbers of trials. In addition, for all three cases, regression lines fitted through the signal levels at the usable reversals in each track were all approximately flat (the average slope was 0.057 for high-frequency ITD, 0.041 for low-frequency ITD, and −0.006 for ILD). Finally, the signal levels at the usable reversals in each track, if anything, varied more in the ILD than in the ITD conditions, as measured by the standard deviation of those signal levels expressed as a percentage of the estimated threshold (6.1% for high-frequency ITD, 7.6% for low-frequency ITD, and 11.0% for ILD). Thus, for all three cases, the usable reversals appeared to fluctuate within a small range of the final threshold estimates. Taken together, regardless of the different starting values, the adaptive procedure appeared to be equally effective for evaluating performance on high-frequency ITD, low-frequency ITD, and ILD discrimination.

III. RESULTS

A. Patterns of learning

1. Effect of practice on high-frequency ITD discrimination

Multihour training on ITD discrimination with high-frequency SAM tones did not help to improve listeners' performance beyond what was learned during the pretest on any condition tested. On the trained condition [ITD 4 kHz SAM at 0.3 kHz, Fig. 1(a)], the trained listeners did not improve more than controls between the pre- and posttests. According to a 2 group (trained versus control) $\times 2$ time (pretest versus posttest) analysis of variance (ANOVA) on the discrimination thresholds, with repeated measures on time, thresholds decreased significantly between the pre- and posttests (main effect of time, $F_{(1,16)}=8.128$, $p=0.012$).[1] However, the trained listeners did not differ from the controls either in overall thresholds (no main effect of group, $F_{(1,16)}=0.234$, $p=0.635$), or in the amount of improvement between the pre- and posttests (no group by time interaction, $F_{(1,16)}=0.147$, $p=0.706$). There also was no difference between the thresholds of the two groups on either the pretest ($t=-0.292, p=0.774$) or the posttest($t=-0.661, p=0.518$), indicating that the trained listeners both started and ended at the same level as the controls. Thus, both the trained listeners and controls improved on the trained condition, presumably due to the exposure during the pretest, but the trained listeners did not learn more than the controls despite their additional multihour training.

Supporting this conclusion, the trained listeners did not improve during the training sessions, either individually [Fig. 1(b)] or as a group [Fig. 1(a)]. We said that there was training-phase learning, for either an individual or all the trained listeners as a group, when a one-way ANOVA on the thresholds across the training sessions yielded a significant time effect *and* a linear regression of the thresholds fitted over the training sessions gave a significantly negative slope (see also Wright and Fitzgerald, 2001). Repeated measures on the training sessions were used when performing the one-

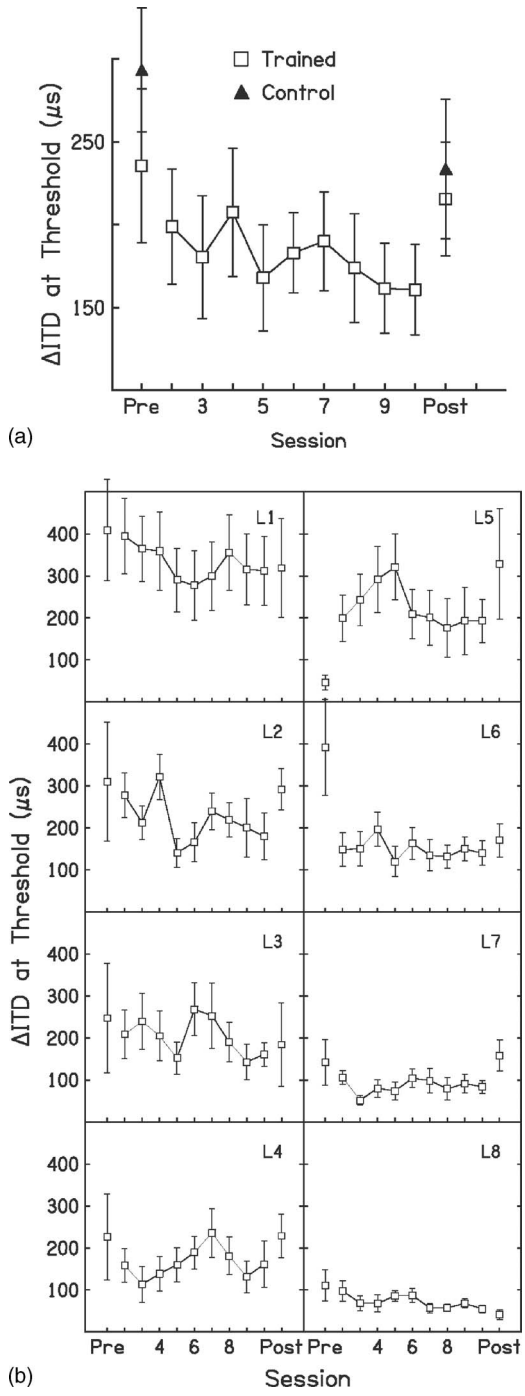


FIG. 1. Discrimination thresholds on the trained condition. (a) The mean ITD discrimination thresholds of eight trained listeners (open squares) and eleven controls (closed triangles) for the pretest, training sessions, and posttest. (b) Individual mean ITD discrimination thresholds of the trained listeners. Error bars represent ±1 standard error across listeners (a) or within individuals (b). Two 300-ms high-frequency complex stimuli (a 4-kHz sinusoidal carrier sinusoidally amplitude modulated at 0.3 kHz) were presented either with a standard ITD at 0 μs or with a signal ITD that equaled the standard ITD plus a ΔITD that favored the right ear. Threshold is defined as the value of ΔITD necessary for listeners to discriminate the signal from the standard ITD with 79% correct performance.

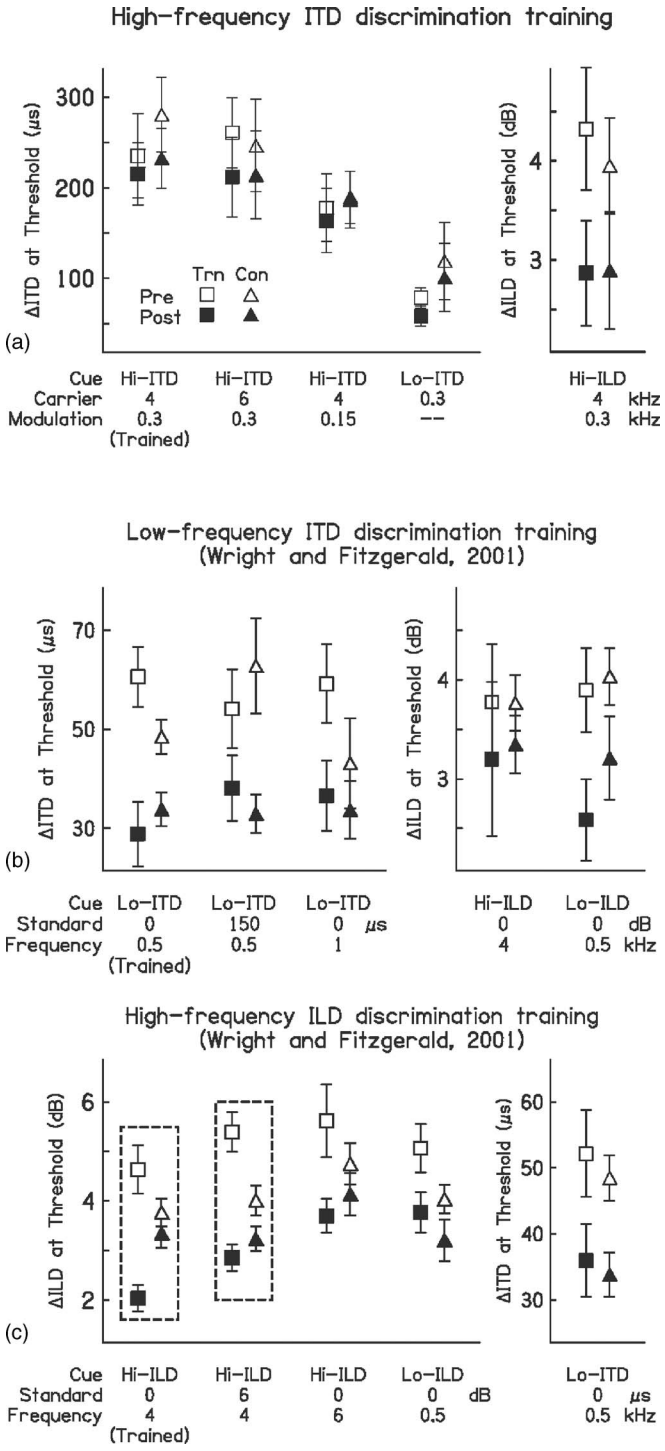


FIG. 2. Pre- and posttest discrimination thresholds in the high-frequency ITD (a), low-frequency ITD (b), and high-frequency ILD (c) training experiments. For (b) and (c), data are replotted from Wright and Fitzgerald (2001). Mean discrimination thresholds are shown for the trained listeners (squares) and controls (triangles) for both the pre- (open symbols) and posttests (closed symbols). The error bars represent ±1 standard error across listeners. Conditions are denoted on the abscissa by the interaural cue manipulated and the stimulus parameters (carrier frequency and modulation rate for AM stimuli, standard ITD or ILD and frequency for pure tones). For each panel, the trained condition is at the far left, and conditions on which trained listeners improved significantly more than controls are indicated by dashed rectangles.

way ANOVA on the thresholds of multiple listeners. We set α at 0.05 for all tests. According to these criteria, no training-phase learning was detected either for any individual trained listener or for all the trained listeners as a group.

On the untrained conditions, the trained listeners also failed to learn more than controls [Fig. 2(a)]. Separate 2 group $\times$2 time ANOVAs with repeated measures on time

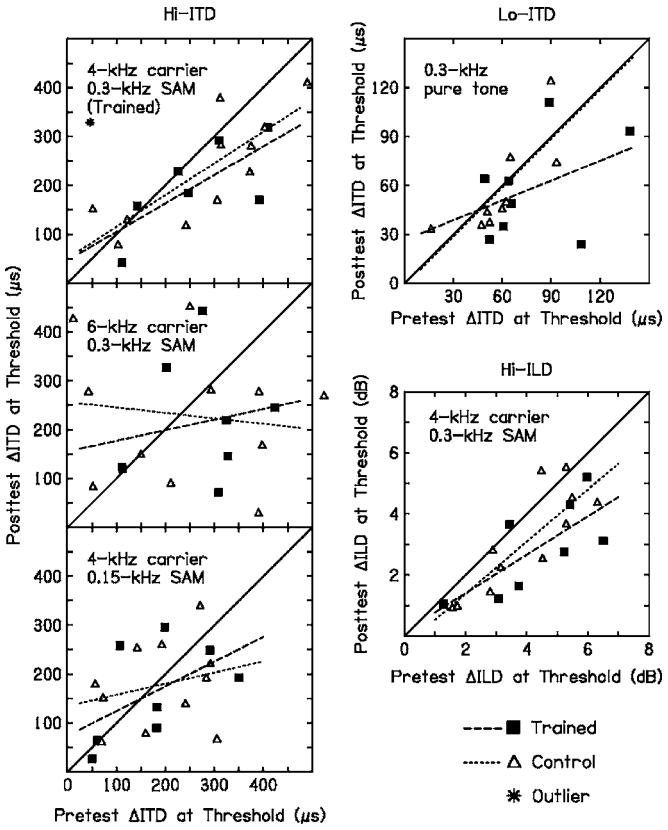


FIG. 3. The relationship between the pre- and posttest thresholds across individual listeners. Pretest (abscissa) and posttest (ordinate) thresholds for individual trained listeners (closed squares) and controls (open triangles) are shown for each of the five tested conditions (panels). Regression lines of the posttest on the pretest thresholds are fitted for the trained listeners (dashed lines) and controls (dotted lines). The diagonal lines (solid lines) represent equal performance on the pre- and posttests. One trained listener on the trained condition (asterisk) and two controls on the 0.3-kHz pure-tone condition (not shown, pretest: 287 and 490 μs, posttest: 122 and 464 μs) are excluded from the calculation of the regression lines because, though they have no influence on the relationship between the two groups, their performance exerted an out-of-proportion influence on the regression lines.

performed on each untrained condition showed that thresholds decreased significantly between the pre- and posttests for the ILD condition (main effect of time, $p<0.001$) but not for any untrained ITD condition (no main effect of time, $p \geqslant 0.119$). Most important, there was no difference between the trained listeners and controls, either in overall performance (no main effect of group, $p \geqslant 0.459$ on all untrained conditions) or in the threshold change between the pre- and posttests (no group by time interaction, $p \geqslant 0.173$). Thus, multihour training on high-frequency ITD discrimination with a SAM tone did not help the trained listeners to improve more than controls on discriminating ITDs with untrained stimulus characteristics, or ILDs with the trained standard stimulus.

The multihour training on high-frequency ITD discrimination also did not change the relationship between the pre- and posttest thresholds (Fig. 3). We performed linear regressions of the posttest on the pretest thresholds across individual listeners for both the trained and control groups. Tests of homogeneity of regression revealed no significant difference between the regression slopes of the trained and control groups on any condition ($p \geqslant 0.242$ for all conditions), indicating that no change in this relationship was introduced by the multihour training.

2. Comparison with learning patterns for low-frequency ITD and high-frequency ILD discrimination

We compared the learning patterns for high-frequency ITDs to those previously obtained for low-frequency ITDs and high-frequency ILDs (Wright and Fitzgerald, 2001). The training paradigm that was used in both the present and previous experiments allows the learning of the trained listeners to be divided into two distinguishable elements. Learning induced by multihour training is revealed by any additional improvement in the trained listeners relative to controls. Learning induced by the exposure during the pretest, on the other hand, is indicated by any improvement in controls. Here we based our comparison of learning patterns on the effects of multihour training and did not further consider those of the pretest. We did so for two reasons. First, multihour training led to distinct learning patterns for low-frequency ITD and high-frequency ILD discrimination, most likely by influencing the processing of these two cues when that processing is still separate. However, the learning patterns induced by the pretest were similar for both cues, suggesting that changes underlying these rapid improvements are common to the two cues (Wright and Fitzgerald, 2001). Thus, multihour training appears to be a more appropriate tool to explore the separate processing of different forms of ITD and ILD cues than is the brief pretest training. Second, due to the variety and randomized order of the conditions in the pretest, the pretest-induced learning on any condition could have resulted either from exposure to that condition or from generalization of learning from other conditions. With this uncertainty, no reliable comparisons between the pretest-induced learning patterns of the different forms of interaural cues could be obtained. Because only one condition was trained during the training phase, there is no such uncertainty for training-induced learning.

Comparing the present results with those reported previously (Wright and Fitzgerald, 2001), the learning pattern induced by multihour training for high-frequency ITDs was similar to that for low-frequency ITDs, but different from that for high-frequency ILDs [compare Fig. 2(a) to Fig. 2(b) and 2(c)]. For ITD discrimination with either low- or high-frequency stimuli [Figs. 2(a) and 2(b)], trained listeners improved no more than untrained controls on thresholds between the pre- and posttests across all conditions tested. The relationship between the pretest and posttest thresholds across individuals also did not differ between the two listener groups (not shown). For high-frequency ILD discrimination [Fig. 2(c)], however, trained listeners improved significantly more than controls on the trained condition and an untrained standard ILD. On these two conditions, the regression of the posttest over the pretest thresholds also yielded significantly shallower slopes for the trained listeners than for controls, indicating that multihour training was of greatest benefit to listeners who started most poorly (Wright and Fitzgerald,

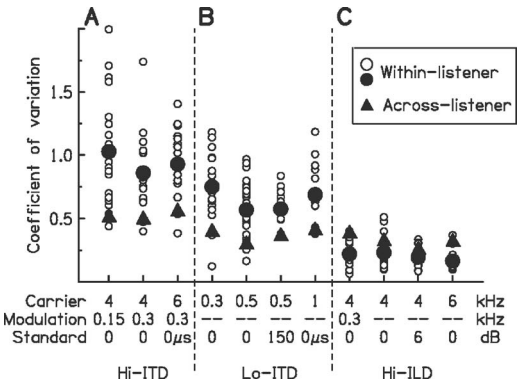


FIG. 4. Naïve within- and across-listener coefficients of variation (CV) for high-frequency ITD (a), low-frequency ITD (b), and high-frequency ILD (c) discrimination. Individual (open circles, each based on five pretest threshold estimates) and average (closed circles) within-listener CVs, as well as across-listener CVs (closed triangles), are shown for data pooled across the current and a previous (Wright and Fitzgerald, 2001) investigation. Conditions are denoted on the abscissa by carrier frequency, modulation rate, and standard ITD or ILD value.

2001, not shown here). Multihour ILD training did not lead to learning on conditions with untrained frequencies or the untrained cue (ITD).

B. Patterns of performance variability

1. Performance variability in high-frequency ITD discrimination

For all of the three ITD SAM conditions, the within-listener variability was significantly larger than the corresponding across-listener variability on the pretest. To enable comparisons of variability across conditions that yielded performance on different scales or even in different units, we calculated the coefficient of variation [CV, standard deviation/average, for example, see Hopkins (2000)] for both within- and across-listener variability, based on the five pretest threshold estimates obtained for each listener on each condition ($n=19$, Fig. 4). The average across-listener CV was approximately constant across the three ITD SAM conditions, ranging between 0.51 and 0.57 [Fig. 4(a), closed triangles]. In parallel, the within-listener CV on these high-frequency ITD discrimination thresholds did not vary significantly with different carrier frequencies or modulation rates, ranging between 0.86 and 1.03 (one-way ANOVA, $p=0.142$). However, on each ITD SAM condition, the CV value was significantly larger for within- than across-listener variability (one sample t test, all p values <0.001).

2. Comparison with variability patterns for low-frequency ITD and high-frequency ILD discrimination

Consistent with the learning results, the performance variability pattern observed for ITD discrimination with high-frequency AM sounds was more similar to that for ITD discrimination with low-frequency pure tones than to that for high-frequency ILD discrimination [compare Fig. 4(a) to Fig. 4(b) and 4(c)]. For this comparison, we calculated within- and across-listener CVs for the low-frequency ITD and high-frequency ILD conditions based on the pretest results combined across both the present and the previous (Wright and Fitzgerald, 2001) experiments. The across-listener CVs were similar in the four low-frequency ITD conditions (ranging from 0.31 to 0.42) and the four high-frequency ILD conditions (ranging from 0.27 to 0.40), and in both cases were somewhat smaller than those in the three high-frequency ITD conditions (ranging from 0.51 to 0.57). In contrast, the within-listener CVs varied considerably across the three cases, with the largest difference lying between the ITD (regardless of frequency region) and ILD conditions. The within-listener CVs differed significantly across the three cases, with the data collapsed across conditions within each case, according to a one-way ANOVA ($p<0.001$). Post hoc analyses revealed a significant difference between each pair of cases (all p values <0.001). The average CVs were smallest for high-frequency ILD (0.20, ranging from 0.17 to 0.23 across the four conditions), moderately large for low-frequency ITD (0.64, ranging from 0.57 to 0.75 across the four conditions), and larger yet for high-frequency ITD (0.94, ranging from 0.86 to 1.03 across the three conditions). In terms of effect size, the difference between low-frequency ITD and high-frequency ILD (2.5) was more than twice that between low- and high-frequency ITD (1.0).

Finally, the relationship between the within- and across-listener CVs for high-frequency ITD resembled that for low-frequency ITD, but differed from that for high-frequency ILD. For low- and high-frequency ITD discrimination, the within-listener CVs were 1.7 and 1.8 times as large as the across-listener CVs, respectively. For high-frequency ILD discrimination, this pattern was reversed, with the average *across-listener* CVs 1.6 times as large as the *with*in-*listener* CVs, mainly due to the markedly smaller within-listener variability in the ILD than the ITD conditions.

IV. DISCUSSION

The present results suggest that high-frequency ITD processing is more similar to low-frequency ITD than high-frequency ILD processing, in terms of both modifiability and internal noise level.

In terms of malleability, the learning data indicate that the processing of high-frequency ITDs resembles that for low-frequency ITDs, and differs from that for high-frequency ILDs. Identical multihour training paradigms yielded learning on high-frequency ILD discrimination, but not on ITD discrimination with either low- or high-frequency stimuli, indicating that ITD processing, regardless of the frequency region, is less malleable than ILD processing. The lack of training-induced learning on ITD discrimination suggests that, during the multihour training used here, no point along the entire neural pathway that underlies either high- or low-frequency ITD processing underwent modifications that had behavioral consequences for the current measures. In contrast, the presence of learning on ILD discrimination, induced by the same paradigm, indicates that at least one point in the ILD processing pathway is malleable. Therefore, high-frequency ITD processing differs from ILD processing at

least at one point, and resembles low-frequency ITD processing along the whole pathway in terms of the malleability revealed by the present multihour training.

It is worth noting that the observed lack of training-induced discrimination learning for high-frequency ITDs cannot be attributed to listeners being already highly practiced on this task through their daily sound exposure, as was previously proposed as a possible explanation for the absence of training-induced learning on low-frequency ITD discrimination (Wright and Fitzgerald, 2001). Because for high-frequency sounds, ILD, not ITD, is the dominant localization cue (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002), the experimentally naïve listeners would already have had more experience with sound localization based on high-frequency ILD than on high-frequency ITD. Thus, it is unlikely that high-frequency ITDs are over learned while high-frequency ILDs remain malleable.

Note also that the lack of training-induced learning on ITD discrimination under the current training paradigm does not preclude the possibility of improvement on ITD tasks (for a review of behavioral learning on binaural tasks, see Wright and Zhang, 2006). For example, there are reports of improvement on ITD discrimination following either an initial testing session (Ortiz and Wright, 2001; Wright and Fitzgerald, 2001) or a different training regimen (Rowan and Lutman, 2005). Though not consistently so (Litovsky *et al.*, 2000), learning has also been observed on an ITD-based precedence task (Saberi and Perrott, 1990), but may require an extraordinary amount of training ($>$20 h, Saberi and Antonio, 2003).

Consistent with the learning results, the present patterns of performance variability also suggest a larger difference between ITD and ILD processing than between ITD processing with low- and high-frequency stimuli. Performance variability was larger within than across listeners for ITD discrimination regardless of frequency region, but showed the reverse pattern for ILD discrimination. Across-listener variability, though slightly larger for high-frequency ITDs than for the other two cases, did not vary much across different cues (ITD and ILD), stimulus types (low-frequency pure tones, high-frequency pure tones, and high-frequency AM tones), and listener groups, and thus appeared to be determined by some factor that was common at least to both ITD and ILD processing regardless of stimulus parameters. In contrast, within-listener variability, reflecting the internal noise in the neural processing within a given individual, varied significantly across the three cases, with the clearest difference lying between ITDs and ILDs rather than between low- and high-frequency ITDs. Overall, ITD processing, whether for low or high frequencies, appears to be less malleable and to have greater internal noise than ILD processing.

The different learning and variability patterns in high-frequency ITD and ILD discrimination are unlikely to have resulted from the different stimulus types used in these two cases (pure tones for ILD and AM tones for ITD). Suggesting that the lack of training-induced learning is a feature of ITD processing rather than of AM stimuli, multihour training on ILD discrimination with high-frequency AM tones yielded significant learning (Zhang and Wright, 2005; Kumpik *et al.*, 2006), as was the case for pure-tone ILDs, while the same training yielded no improvement on ITD discrimination with either stimulus type. Similarly, suggesting that the large within-listener variability is attributable to ITD instead of AM stimuli, the magnitudes of the within- and across-listener variability in the present ILD condition with high-frequency AM tones were more similar to those in the pure-tone ILD, than to those in the AM-tone ITD, conditions (Fig. 4).

Given the apparent minimal influence of stimulus type, we interpret the present learning and variability data as supporting similar processing of ITDs across frequency regions and differential processing of ITDs and ILDs. The current learning results for high-frequency ITDs indicate that the different malleability previously observed between low-frequency ITDs and high-frequency ILDs (Wright and Fitzgerald, 2001) appears to be representative of ITD and ILD processing in general, regardless of stimulus frequency. Likewise, the systematic analyses of the within- and across-listener variability lead to the same conclusion in respect to internal noise level in neural processing. Thus, the present results on each of the two independent behavioral metrics reveal simultaneously both a similarity between low- and high-frequency ITDs and a difference between high-frequency ITDs and ILDs, making it evident that the restricted modifiability and large internal noise level are indeed characteristic of ITD processing at both low- and high-frequency regions, rather than of binaural processing in general. The current data may also help in evaluating the behavioral relevance of physiological observations. As noted in Sec. I, high-frequency ITD sensitive neurons have been found at the earliest stages of both the low-frequency ITD and the high-frequency ILD pathways (Spitzer and Semple, 1995; Batra *et al.*, 1997a, b). The present results suggest that human sensitivity to high-frequency ITDs is primarily mediated either by high-frequency neurons along the low-frequency ITD pathway, or by the ITD-sensitive neurons along the high-frequency ILD pathway but with the proviso that the ITD processing initiated by those neurons is less modifiable and more noisy than the ILD processing initiated by the same neurons. Taken together, the present results add to the behavioral evidence supporting Colburn and Esquissaud's (1976) proposal of similar ITD processing across frequency regions, and further reveal that ITD processing at both low and high frequencies is less modifiable and more noisy than ILD processing.

ACKNOWLEDGMENTS

We thank Rodrigo Cadiz and Andrew Sabin for technical support, and Karen Banai, Matthew Fitzgerald, Julia Huyck, Julia Mossbridge, Jeanette Ortiz, and Andrew Sabin for helpful suggestions on previous drafts of this paper. This paper was further strengthened by the thoughtful comments of Dr. Wesley Grantham and one anonymous reviewer during the review process. Matthew Fitzgerald performed a preliminary analysis of variability on the data from Wright and Fitzgerald (2001). This work was supported by NIH/NIDCD.

[1]The effect of time was significant only after excluding L5 from the analysis (with L5 included, the main effect of time was $F_{(1,17)}=1.753$, $p=0.203$). This listener showed an unusual deterioration in performance from the pretest to the following sessions. The significance obtained with L5 excluded held with any other single listener removed. Note that L5 was included in all of the remaining statistical analyses, because excluding her did not alter any of the conclusions.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997a**). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. II. Coincidence detection," J. Neurophysiol. **78**, 1237–1247.

Batra, R., Kuwada, S., and Fitzpatrick, D. C. (**1997b**). "Sensitivity to interaural temporal disparities of low- and high-frequency neurons in the superior olivary complex. I. Heterogeneity of responses," J. Neurophysiol. **78**, 1222–1236.

Bernstein, L. R. (**2001**). "Auditory processing of interaural timing information: New insights," J. Neurosci. Res. **66**, 1035–1046.

Bernstein, L. R., and Trahiotis, C. (**1994**). "Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise," J. Acoust. Soc. Am. **95**, 3561–3567.

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., and Trahiotis, C. (**2003**). "Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed stimuli'," J. Acoust. Soc. Am. **113**, 3335–3347.

Best, V., Carlile, S., Jin, C., and van Schaik, A. (**2005**). "The role of high frequencies in speech localization," J. Acoust. Soc. Am. **118**, 353–363.

Blauert, J. (**1982**). "Binaural localization: Multiple images and applications in room- and electroacoustics," in *Localization of Sound: Theory and Application*, edited by R. W. Gatehouse (Amphora, Groton), pp. 65–84.

Brunt, D., Magill, R. A., and Eason, R. (**1983**). "Distinctions in variability of motor output between learning disabled and normal children," Percept. Mot. Skills **57**, 731–734.

Buus, S., Scharf, B., and Florentine, M. (**1984**). "Lateralization and frequency selectivity in normal and impaired hearing," J. Acoust. Soc. Am. **76**, 77–86.

Colburn, H. S., and Esquissaud, P. (**1976**). "An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli," J. Acoust. Soc. Am. **59**, 23.

Demany, L., and Semal, C. (**2002**). "Learning to perceive pitch differences," J. Acoust. Soc. Am. **111**, 1377–1388.

Eberle, G., McAnally, K. I., Martin, R. L., and Flanagan, P. (**2000**). "Localization of amplitude-modulated high-frequency noise," J. Acoust. Soc. Am. **107**, 3568–3571.

Feddersen, W., Sandel, T., Teas, D., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.

Fitzgerald, M. B., and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," J. Acoust. Soc. Am. **118**, 3794–3803.

Griffiths, T. D., Elliott, C., Coulthard, A., Cartlidge, N. E., and Green, G. G. (**1998**). "A distinct low-level mechanism for interaural timing analysis in human hearing," NeuroReport **9**, 3383–3386.

Hawkins, D. B., and Wightman, F. L. (**1980**). "Interaural time discrimination ability of listeners with sensorineural hearing loss," Audiology **19**, 495–507.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Henning, G. B. (**1980**). "Some observations on the lateralization of complex waveforms," J. Acoust. Soc. Am. **68**, 446–454.

Henning, G. B., and Ashton, J. (**1981**). "The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay," Hear. Res. **4**, 185–194.

Hopkins, W. G. (**2000**). "Measures of reliability in sports medicine and science," Sports Med. **30**, 1–15.

Irvine, D. R. (**1986**). "The auditory brainstem: A review of the structure and function of auditory brainstem processing mechanisms," in *Progress in Sensory Physiology*, edited by D. Ottoson (Springer, Berlin), pp. 1–279.

Joris, P. X., and Yin, T. C. (**1995**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.

Karni, A., and Sagi, D. (**1991**). "Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity," Proc. Natl. Acad. Sci. U.S.A. **88**, 4966–4970.

Klumpp, R., and Eady, H. (**1956**). "Some measurements of interaural time difference thresholds," J. Acoust. Soc. Am. **28**, 859–860.

Kuhn, G. (**1977**). "Model of the interaural differences in the azimuthal plane," J. Acoust. Soc. Am. **49**, 157–167.

Kumpik, D. P., Kacelnik, O., Schnupp, J. W. H., and King, A. J. (**2006**). "Binaural perceptual learning: The effect of training on interaural level discrimination with amplitude modulated tones," Association for Research in Otolaryngology Abstracts **29**, 146.

Lacher-Fougere, S., and Demany, L. (**2005**). "Consequences of cochlear damage for the detection of interaural phase differences," J. Acoust. Soc. Am. **118**, 2519–2526.

Leakey, D., Sayers, B., and Cherry, C. (**1958**). "Binaural fusion of low- and high-frequency sounds," J. Acoust. Soc. Am. **30**, 222–223.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Litovsky, R. Y., Hawley, M. L., Fligor, B. J., and Zurek, P. M. (**2000**). "Failure to unlearn the precedence effect," J. Acoust. Soc. Am. **108**, 2345–2352.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

McFadden, D., and Moffitt, C. M. (**1977**). "Acoustic integration for lateralization at high frequencies," J. Acoust. Soc. Am. **61**, 1604–1608.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization of high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

Middlebrooks, J. C., and Green, D. M. (**1990**). "Directional dependence of interaural envelope delays," J. Acoust. Soc. Am. **87**, 2149–2162.

Nuetzel, J. M., and Hafter, E. R. (**1976**). "Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339–1346.

Ortiz, J. A., and Wright, B. A. (**2001**). "Rapid improvements on interaural-time-difference discrimination: Evidence for three types of learning," J. Acoust. Soc. Am. **109**, 2289.

Plomp, R. (**1965**). "Detectability threshold for combination tones," J. Acoust. Soc. Am. **37**, 1110–1123.

Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.

Rowan, D., and Lutman, M. (**2005**). "Generalisation of learning with ITD discrimination across frequency and type of cue," Association for Research in Otolaryngology Abstracts **28**, 257.

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 Antonio, J. V. (**2003**). "Precedence-effect thresholds for a population of untrained listeners as a function of stimulus intensity and interclick interval," J. Acoust. Soc. Am. **114**, 420–429.

Saberi, K., and Perrott, D. R. (**1990**). "Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate," J. Acoust. Soc. Am. **87**, 1732–1737.

Smoski, W. J., and Trahiotis, C. (**1986**). "Discrimination of interaural temporal disparities by normal-hearing listeners and listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **79**, 1541–1547.

Spitzer, M. W., and Semple, M. N. (**1995**). "Neurons sensitive to interaural phase disparity in gerbil superior olive: Diverse monaural and temporal response properties," J. Neurophysiol. **73**, 1668–1690.

Takahashi, C. D., Nemet, D., Rose-Gottron, C. M., Larson, J. K., Cooper, D. M., and Reinkensmeyer, D. J. (**2003**). "Neuromotor noise limits motor performance, but not motor adaptation, in children," J. Neurophysiol. **90**, 703–711.

Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.

Wright, B. A., Buonomano, D. V., Mahncke, H. W., and Merzenich, M. M. (**1997**). "Learning and generalization of auditory temporal-interval discrimination in humans," J. Neurosci. **17**, 3956–3963.

Wright, B. A., and Fitzgerald, M. B. (**2001**). "Different patterns of human discrimination learning for two interaural cues to sound-source location," Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312.

Wright, B. A., and Zhang, Y. (**2006**). "A review of learning with normal and altered sound-localization cues in human adults," Int. J. Audiol. **45**, 92–98.

Zhang, Y., and Wright, B. A. (**2005**). "Different interaural level difference processing with complex sounds and pure tones," J. Acoust. Soc. Am. **117**, 2562.

The effect of impedance on interaural azimuth cues derived from a spherical head model[a)]

Bradley E. Treeby,[b)] Roshun M. Paurobally, and Jie Pan
Centre for Acoustics, Dynamics and Vibration, School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

(Received 15 June 2006; revised 24 January 2007; accepted 26 January 2007)

Recent implementations of binaural synthesis have combined high-frequency pinna diffraction data with low-frequency acoustic models of the head and torso. This combination ensures that the salient cues required for directional localization in the horizontal plane are consistent with psychophysical expectations, regardless of the accuracy or match of the high-frequency cues, or the fidelity of experimental low-frequency information. This paper investigates the effect of a nonrigid boundary condition on the surface pressure and the resulting interaural cues used for horizontal localization. These are derived from an analytical single sphere diffraction model assuming a locally reacting and uniformly distributed impedance boundary condition. Decreasing the magnitude of a purely resistive surface impedance results in an overall decrease in the sphere surface pressure level, particularly in the posterior region. This produces nontrivial increases in both the interaural level and time difference, especially for sound source directions near the interaural axis. When the surface impedance contains a reactive component the interaural cues exhibit further changes. The basic impedance characteristics of human hair and their incorporation into the sphere diffraction model are also discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2709868]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.20.Fn [AK] Pages: 2217–2226

I. INTRODUCTION

Typical implementations of virtual audio systems combine acoustic signals with head related transfer functions (HRTFs) to artificially position sounds within a virtual three-dimensional (3D) environment. The HRTF itself is a measure of the combined diffraction properties of the exterior human acoustic system, and generally reflects the spectral changes that a plane sound wave from a particular source direction would undergo by the time it reaches the eardrums. Extensive localization testing in recent years has highlighted the importance of using HRTFs that are accurately matched to the individual users so that the auditory cues required for a realistic and spatially accurate acoustic environment are maintained. The use of nonindividualized HRTFs generally results in decreased spatial accuracy amongst other common error artifacts, although the extent depends on the degree of mismatch between the end user and the HRTF data utilized (Wenzel *et al.*, 1993; Bronkhorst, 1995). The availability of laboratory environments equipped to measure personalized HRTF sets is exceptionally limited; thus many recent implementations have instigated solutions which attempt to provide a better match between the end user and the nonindividualized HRTF set used. Amongst others these methods include selecting the HRTF set based on anthropometric similarities between the user and a database of previously measured users (Zotkin *et al.*, 2003), frequency scaling of the HRTF based on pinna characteristics (Middlebrooks, 1999), and allowing the user a choice of standard HRTFs based on personal preference of the quality of the virtual projection (Middlebrooks *et al.*, 2000). In particular, implementations combining an analytical low-frequency head, or head-and-torso, model with the high-frequency pinnae data of the HRTF have had particular success (Brown and Duda, 1998; Algazi *et al.*, 2002a; b). These composite HRTFs progressively replace cues with those derived from the analytical model for frequencies below 3000 Hz (below 500 Hz the model is used exclusively). The models can be modified to suit the anthropometry of the end user (Algazi *et al.*, 2001), and provide stable azimuth cues and trends consistent with psychophysical expectations regardless of the coherence of the low-frequency component of the HRTF, or the accuracy of the match between the spectral pattern provided by the users own pinnae and the high-frequency data used.

This paper provides a comprehensive investigation into the effect of altering the surface boundary condition of a low-frequency head model on the surface pressure and the salient azimuth cues [interaural time difference (ITD) and interaural level difference (ILD)]. These are derived from an analytical head model based on diffraction around a single sphere with a uniformly distributed and locally reacting impedance boundary. The analytical diffraction solution is presented and subsequently experimentally validated. The effect of varying the surface impedance on the corresponding surface pressure over the frequency range of interest for head-and-torso models (below 3000 Hz) is then investigated. These results are used to examine the general trends experienced by the interaural difference cues with impedance. Dis-

[a)]Portions of this work were presented in "Investigation of the effect of impedance on azimuth cues derived from spherical head models," Proceedings of the 12th International Conference on Auditory Display, London, UK, 20–23 June, 2006.

[b)]Electronic mail: treebs@mech.uwa.edu.au

cussion relating to the effect of human hair impedance on the interaural cues (based on the same boundary condition assumption) is also given.

Although mathematical solutions for a locally reacting and uniformly distributed impedance boundary condition have been previously published (Lax and Feshbach, 1948; Kear, 1959; Morse and Ingard, 1968), such treatments have generally not been utilized in virtual acoustics. In this regard derivations and discussions are overwhelmingly based on rigid boundary conditions and cannot be trivially expanded to include an acoustic impedance term (the assumption of an acoustically rigid boundary allows noticeable simplification of the analytical diffraction solution). Most commonly the models used are simply a single pinnae-less rigid sphere with its diameter matched to the inter-ear distance of the user, although more recent implementations have combined this with a separated spherical torso to include a low-frequency approximation of the secondary information introduced by upper torso reflections (Gumerov and Duraiswami, 2002; Algazi *et al.*, 2002a). The principal analytic results for the model utilized in the current study are thus also included in this paper.

Previous experimental studies have illustrated changes in the HRTF and interaural cues with varying the surface boundary conditions. Kuhn (1977) compared measurements from a mannequin (with a torso) with and without clothing and concluded that the interaural cues changed with this modification. For small angles of head rotation relative to the torso (which was held constant relative to the source angle) this study reported variability in the ILD and ITD over frequency (260–4000 Hz) of up to 5 dB and 50 μs, respectively. Riederer (2005) also investigated the effect of hair and hair style changes on the resultant HRTF using a mannequin fitted with a variety of wigs. The effect of the hair thickness and curliness appeared to be the most prominent (subsequent modification of hair style or changes in overall hair length reportedly had a much smaller effect) and produced an increased shadowing effect of the incident wave, with changes in the HRTF above 2 kHz on the order of 5 dB. Studies utilizing boundary element models incorporating nonrigid head surfaces have shown similar results; Katz (2001) reported up to 6 dB difference in the HRTF generated due to the inclusion of hair characteristics based on a locally reactive boundary condition, the effects again being most pronounced in the rear (shadowed) region of the head.

The lack of consideration for nonrigid acoustic boundary conditions in analytical auditory cue investigations may also be attributed to both the variability and the general absence of data on the acoustical properties of the various surfaces that constitute the human head and torso. This is in part due to the lack of experimental techniques available to take *in situ* measurements of impedance from human subjects. Limited previous studies have provided absorption coefficient data obtained using an impedance tube for both human skin (found to be approximately rigid in the frequency range of interest) and hair (Katz, 2000). In the case of hair, additional difficulty arises in accurately describing a boundary condition to embody the diverse range of hair densities, lengths, distributions, and fiber shapes. While the current work does not attempt to formulate a model to address this issue, it does give an indication of the level of changes expected from more detailed studies.

II. ANALYTICAL DIFFRACTION MODEL

A. Background to scattering problems

The scattering of sound by an object in a uniform and nonviscous medium is characterized by the corresponding pressure distribution, particularly in the region of the scatterer. This can be decomposed into an incident and a scattered field, both of which are governed by the scalar wave equation. In the case of a locally reactive sphere in an infinite medium, for a given incident wave the scattered field is entirely determined by the boundary condition imposed on the surface of the scatterer. This surface boundary condition can be examined using the notion of acoustic impedance (the ratio of surface pressure to radial velocity). The term is dependent on the acoustic properties of the boundary and surrounding fluid, and by definition is time invariant. Of particular interest is whether a material can be considered to be locally reactive. This requires that there is no significant wave motion in the direction tangent to the surface. The motion of the surface at any point can then be considered dependent only on the incident acoustic pressure at that location. While the angle of refraction within the material is dependent on the specific macroscopic material properties, it follows that if the surface material is reasonably dense (greater than $\sim$150 kg/m^3) then over the frequency range of interest a locally reactive boundary condition is a reasonable assumption (Ingard, 1981).

The boundary condition for a sphere of radius a with a locally reacting and uniformly distributed normal acoustic impedance $\mathrm{Z}(\omega)$ is given by

$$0 = \left[\frac{\partial p}{\partial r} + \left(i\rho_0 \omega \frac{1}{\mathrm{Z}(\omega)} \right) p \right]_{r=a}. \tag{1}$$

Here p is the sound pressure, ρ_0 is the mean medium density, ω is the angular frequency, and harmonic time dependence of the form $e^{-i\omega t}$ has been assumed. Impedance values are commonly discussed in their normalized form:

$$\zeta = \frac{Z}{\rho_0 c_0} = \theta + i\chi, \tag{2}$$

where ζ represents the specific acoustic impedance with resistive and reactive components θ and χ, c_0 the wave speed, and $\rho_0 c_0$ the characteristic impedance of the medium. Care must be taken when investigating impedance to ensure that consistent use of the harmonic time component is maintained between impedance measurements and analytical solutions. Using $e^{-i\omega t}$, most common acoustic materials will have a positive impedance phase angle over the frequency range of interest as the stiffness of the material dominates the response.

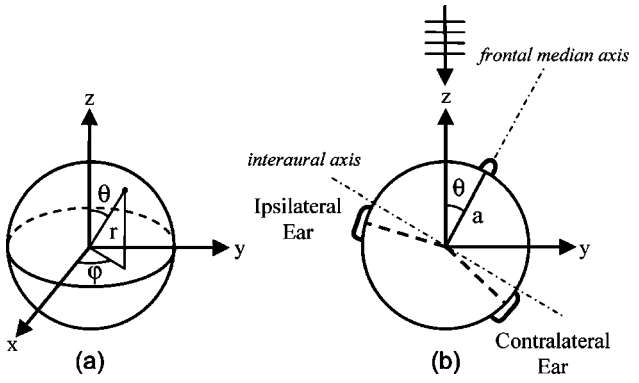


FIG. 1. (a) General spherical coordinate system. (b) Symmetrical coordinate system with incident wave in the negative z direction. The pinnae are offset from the frontal median axis by 100°.

B. Derivation of pressure equation

A general solution for the scalar wave equation (assuming a harmonic time dependence) in spherical coordinates as shown in Fig. 1(a) can be expressed as

$$p = p_0 e^{-i\omega t} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} G_n(r) Y_n^m(\theta, \varphi). \tag{3}$$

Here $G_n(r)$ represents a solution to the spherical Bessel differential equation (which constitutes a linear combination of spherical Bessel, spherical Neumann, or spherical Hankel functions), and $Y_n^m(\theta, \varphi)$ represents the spherical harmonic function. When considering scattering around a single sphere with a uniformly distributed boundary condition, the solution is axially symmetric. As a result the spherical harmonic component of the solution reduces to a Legendre polynomial that depends only on the cosine of the angle between the source and the evaluation position. The corresponding expression for a planar incident wave traveling in the negative z direction is

$$p_i = p_0 e^{-i\omega t} \sum_{n=0}^{\infty} (-i)^n (2n+1) j_n(kr) P_n(\cos\theta), \tag{4}$$

where θ and r define the evaluation position relative to the incident wave using the spherical coordinate system shown in Fig. 1(b), k is the wave number, P_n the Legendre polynomial, and j_n the spherical Bessel function of the first kind which is a solution to the Bessel equation that is finite at the origin. Similarly the expression for the scattered wave is

$$p_s = e^{-i\omega t} \sum_{n=0}^{\infty} A_n h_n^{(1)}(kr) P_n(\cos\theta), \tag{5}$$

where $h_n^{(1)}$ is the spherical Hankel function of the first kind. This represents a solution to the Bessel equation that satisfies the Sommerfeld radiation condition. The singularity of the spherical Hankel function at the origin implies that the sphere must have a finite size.

The constant A_n in Eq. (5) is solved using the boundary condition imposed on the surface of the sphere. Using the locally reactive and uniformly distributed boundary condition described in Eq. (1) yields the complete expression for the sound pressure due to the existence of the sphere in the propagation medium:

$$(p_i + p_s) = p_0 e^{-i\omega t} \sum_{n=0}^{\infty} (-i)^n (2n+1) P_n(\cos\theta) B_n, \tag{6}$$

where

$$B_n = j_n(kr) - \frac{h_n^{(1)}(kr)\left(j_n'(ka) + i\rho_0 c \frac{1}{Z(\omega)} j_n(ka)\right)}{h_n'^{(1)}(ka) + i\rho_0 c \frac{1}{Z(\omega)} h_n^{(1)}(ka)}, \tag{7}$$

and the $(\prime)$ operator denotes the radial derivative.

The interaural azimuth cues are derived from pressure values on the surface of the sphere and Eq. (7) simplifies significantly for $r=a$ using the recursion and cross product relationships for spherical Bessel functions. This yields the surface pressure function:

$$(p_i + p_s)_{r=a} = p_0 e^{-i\omega t} \frac{1}{(ka)^2} \sum_{n=0}^{\infty} \frac{(-i)^{n-1}(2n+1) P_n(\cos\theta)}{h_n'^{(1)}(ka) + i\rho_0 c \frac{1}{Z(\omega)} h_n^{(1)}(ka)}. \tag{8}$$

For a rigid sphere the impedance value becomes infinite and Eq. (8) reduces to yield the solution for the rigid boundary condition as presented and discussed by many previous authors (e.g., Blauert, 1997).

C. Extraction of azimuth cues

The two primary auditory cues for discerning sound source direction in the azimuth plane are the ILD and the ITD. These cues can be analytically derived from Eq. (8) using the known locations of the pinnae which are offset from the frontal median axis by 100° as shown in Fig 1(b). The ILD is obtained by examining the difference in sound pressure level between the two pinna locations:

$$\mathrm{ILD} = 20 \log_{10}\left(\left| \frac{(p_i + p_s)_{\text{ipsilateral ear}}}{(p_i + p_s)_{\text{contralateral ear}}} \right| \right). \tag{9}$$

Similarly the ITD for each angle of incidence is calculated from the interaural difference in arrival time. The actual analytical value of ITD differs slightly depending on the method used to derive this time delay. Interaural phase delay and group delay are frequently used and exhibit similar characteristics but both are frequency dependent and display an increase in ITD at low frequencies (Kuhn, 1977). Ray tracing algorithms (Woodworth and Schlosberg, 1954), amplitude threshold based time constants (Duda and Martens, 1998), or comparisons to minimum phase reconstructions (Kulkarni *et al.*, 1995) can alternatively be used to derive a single value for interaural time difference for each source angle. The ITD is calculated here using a frequency averaged interaural

phase delay. For a particular location on the sphere this phase delay is computed using the negative of the phase response $\Theta(\omega)$ divided by the angular frequency:

$$D(\omega) \triangleq -\frac{\Theta(\omega)}{\omega}. \tag{10}$$

The ITD is then computed from the difference in this delay (averaged across the frequency range of interest) between the two pinna locations. Since the monaural resonant effects of the pinna cavity begin to appear in an experimental HRTF above 3000 Hz, implementations of composite HRTFs generally utilize the analytical model only for frequencies up to this point. The phase delay average is hence calculated over a frequency range from 100 to 3000 Hz. This has the effect of proportionally increasing the ITD values when compared to those obtained from averages over a larger frequency range, or those found using some of the alternative measures mentioned previously. This increase is due to the augmented effect of the larger phase delay below 1000 Hz; however the observed trends due to the impedance inclusion are considered consistent regardless of this selection.

III. EXPERIMENTAL COMPARISON

A. Equipment and setup

The surface pressure function Eq. (8) was validated through a series of interaural measurements completed in an anechoic chamber. A rotating wooden sphere of radius 0.124 m (supported by a thin steel rod) was used which contained two internal microphone cavities (without pinnae), each offset from the frontal median axis by 100°. The angle of the frontal median sphere axis relative to the sound source was aligned using a laser level positioned at the base of the sound source (located approximately 3 m from the sphere) in conjunction with degree markings on the rotating sphere stand. The 0° marking was initially aligned such that an ITD of approximately 0 was obtained. Exposed areas of the stand were covered with a thick layer of highly absorbent material to minimize additional reflections.

A felt covering (190 kg/m^3, 10.3 mm thick) was used to investigate the changes in interaural cues derived from a sphere with a nonrigid boundary condition. This material was selected based on its varying acoustic characteristics throughout the frequency range of interest, its conformity to the locally reactive surface assumption, availability, and its ability to be crafted into a suitable form to use as a covering for the wooden test sphere. The covering was constructed such that the material fitted tightly but was not deformed to any significant degree. Small circular holes were cut in the material coincident with the microphone locations to enable microphone placement flush with the outer material surface (although repeat experiments showed that this placement was not critical). Comparison impedance measurements of both the wood and felt materials were conducted using the two microphone impedance tube method according to the ISO 10534-2 standard (ISO 10534-2, 1998). Figure 2(a) shows the absorption characteristics of the materials (including the rigid steel backing plate of the impedance tube), and Fig. 2(b) shows the resistive and reactive components of the specific impedance of the felt test material.

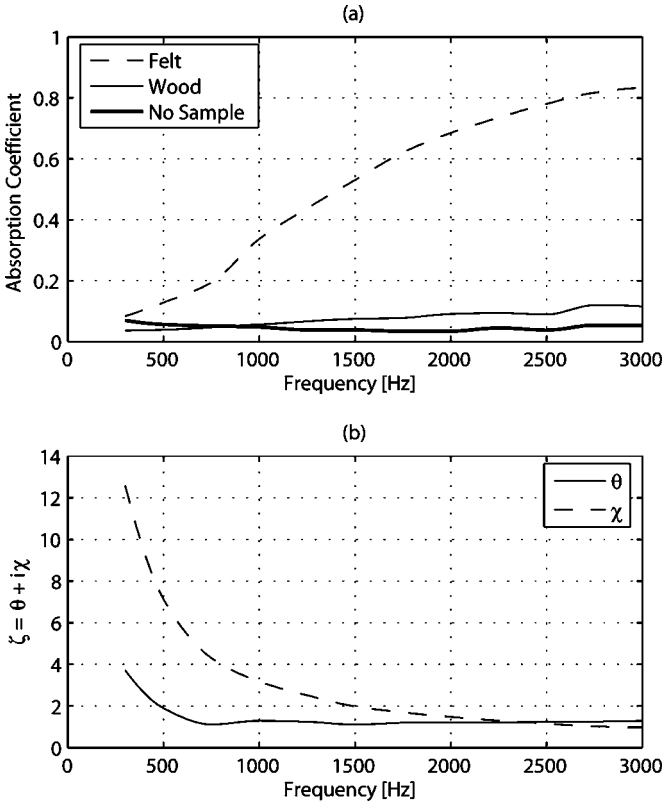


FIG. 2. Measured acoustical properties of the 190 kg/m^3 felt test material. (a) Absorption coefficient with reference to the properties of the wooden test material and steel impedance tube termination. (b) Resistive (θ) and reactive (χ) components of the specific acoustic impedance ζ as defined in Eq. (2).

For each test the microphones (BSWA Tech - MP205 tip, MA211 inline preamplifier) were positioned flush with the exterior of the outer sphere surface. Impulse response measurements were obtained at 5° increments of sphere rotation angle using maximum length sequences produced by the Brüel & Kjær DIRAC software and a Brüel & Kjær HP1001 unidirectional sound source. A sequence length of $2^{14}-1$ (the shortest available sequence length) with 10 averages and a sampling frequency of 48000 Hz was used. To remove the effects of the acoustic measurement environment the impulse response peak onsets were located and the tails truncated to 64 samples. The complete impulse responses were then shortened to 128 samples (with the timing information preserved) and converted to the frequency domain using a 256 point FFT.

B. Interaural comparisons

A comparison between ITD values derived analytically using the measured impedance values and those obtained experimentally is shown in Fig. 3. These results show a strong agreement for both the wooden and felt surfaces. The addition of the felt test surface to the wooden sphere causes an increased sphere radius, and approximately 55% of the difference in ITD between the wooden and felt surfaces is due to this change. The remainder is due to the impedance

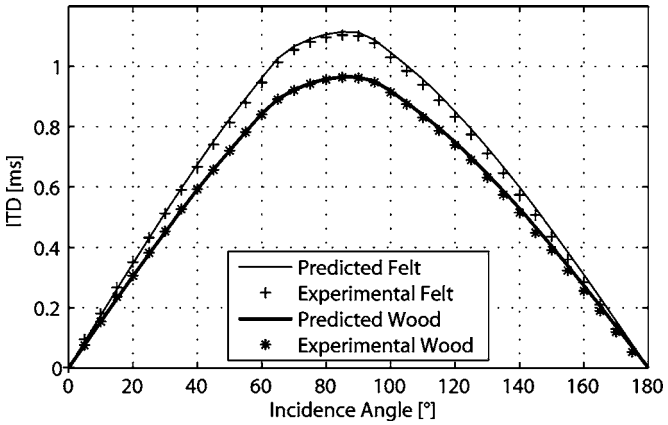


FIG. 3. Comparison between experimental and predicted values of interaural time difference (ITD) for the wood and felt sphere surfaces.

characteristics of the felt. Analytical comparisons for the felt test surface were made using this increased radius (0.135 m).

The experimentally measured change in ILD due to the addition of the felt covering is shown in Fig. 4(a). Again this result matches the predicted changes shown in Fig. 4(b) extremely well. The match between experimental and theoretical results suggests that Eq. (8) provides a strong analytical basis for the derivation of the azimuthal cues experienced by a single sphere with a uniformly distributed surface impedance.

IV. GENERAL TRENDS

A. Effect of impedance on the sphere surface pressure

Experimental comparisons shown and discussed in Sec. III validate the use of the analytical treatment presented in Sec. II for investigating the general effect of impedance on the interaural cues derived from a single sphere. The effects of varying two fundamental aspects of acoustic impedance (magnitude and phase) are investigated here independently. For convenience the general trends in derived interaural differences are examined using frequency independent values of impedance. Results and discussion relating to changes in azimuth cues due to the impedance of human hair (frequency dependent impedance) are given in Sec. V. A sphere radius of 0.0875 m is assumed throughout the remainder of the paper.

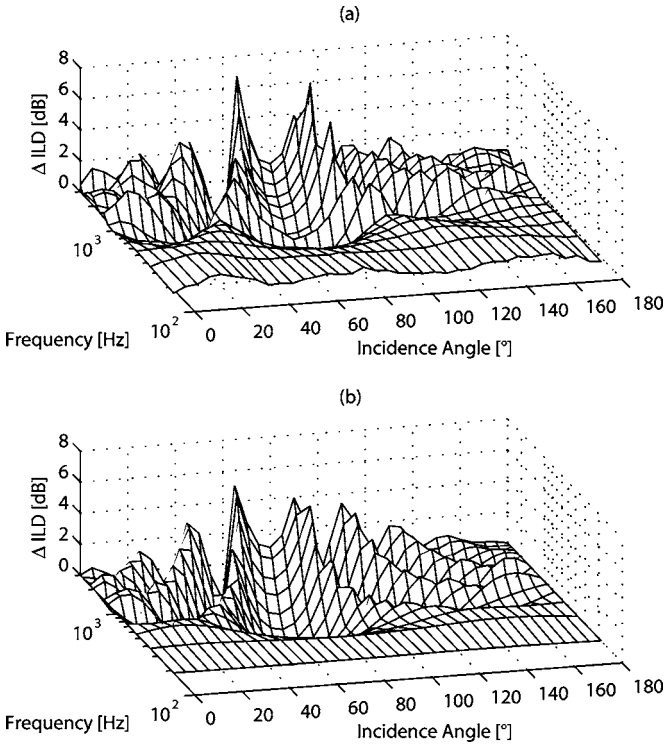


FIG. 4. Comparison between (a) experimental and (b) predicted values of the change in interaural level difference (ΔILD) due to the felt surface.

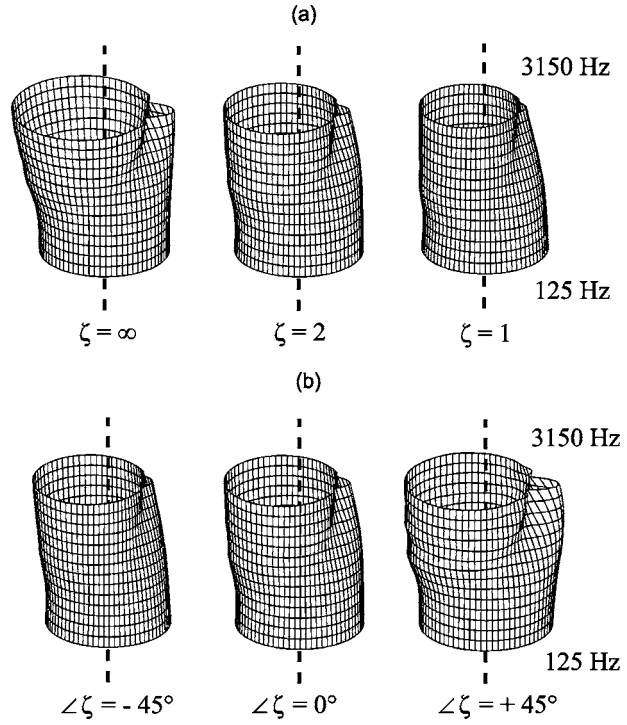


FIG. 5. Sphere surface pressure with changes in (a) resistive specific impedance ζ, and (b) complex specific impedance phase angle $\angle\zeta$ for a constant magnitude of $|\zeta|=2$. The incident wave approaches from the left and each horizontal layer represents the surface pressure at a particular frequency from 125 to 3150 Hz in 1/3 octave band increments. The dashed line indicates the polar origin and the bottom layer (125 Hz) is approximately the unit circle.

Figure 5(a) illustrates the effect of decreasing the magnitude of a purely resistive impedance from an infinite value ($\zeta=\infty$), to the characteristic impedance of air ($\zeta=1$), on the magnitude of the sphere surface pressure. The cylindrical shapes represent stacked polar plots, with each horizontal slice corresponding to the pressure distribution around the sphere circumference at a particular frequency. The frequency spacing corresponds to one-third-octave bands in the range from 125 to 3150 Hz, and the angular spacing is 5°. The incident wave approaches from the left, and the dashed line indicates the polar origin. Consistent with previous investigations of sphere scattering, the response at low frequencies is approximately unity (Morse and Ingard, 1968). This provides a convenient reference for examining the trends at higher frequencies. When the sphere is rigid ($\zeta=\infty$) as the frequency increases the surface pressure in the anterior region begins to approach double the free-field response. This is consistent with expectations of reflection from a rigid plane. In compliance, the response in the poste-

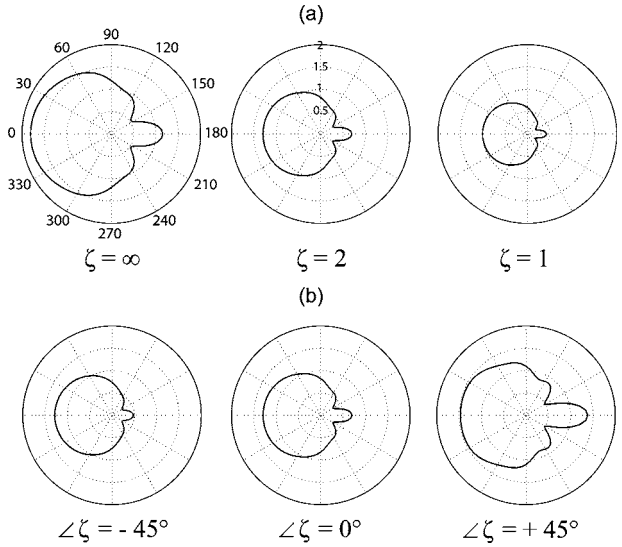


FIG. 6. Sphere surface pressure at $f=2000$ Hz with changes in (a) resistive specific impedance ζ, and (b) complex specific impedance phase angle $\angle\zeta$ for a constant magnitude of $|\zeta|=2$. These polar plots correspond to the third uppermost horizontal slice of the stacked polar plots shown in Fig. 5.

rior region reduces. The lobe of increased pressure at the rear of the sphere is a result of symmetrically diffracted waves arriving in phase creating a "bright spot" (Duda and Martens, 1998). At higher frequencies other small lobes of increased pressure can be seen on either side of the most prominent bright spot resulting from the same phenomena. The angular location of these ancillary lobes is dependent on frequency. Quantitative comparison of the changes in surface pressure for $f=2000$ Hz is possible using Fig. 6(a). These polar plots correspond to the third uppermost horizontal slice of the stacked polar plots shown in Fig. 5(a).

As the magnitude of the resistive surface impedance is decreased, there is an overall decrease in the sphere surface pressure level. There is a particularly significant reduction in the region of the bright spot, proportionally more so than for frontal regions. This is a result of the increased absorption of the sphere surface reducing the intensity of the reflected sound which has a compounding effect on the level of sound that is diffracted to the rear of the sphere. This trend is consistent with previous empirical observations regarding head scattering and impedance, particularly in relation to hair (Katz, 2001; Riederer, 2005).

When the surface impedance also contains a reactive component, the interaction of the incident wave with the sphere boundary causes a phase difference in the reflected wave. This change consequently alters both the resultant sound pressure on and around the sphere, and the corresponding interaural cues. Figure 5(b) illustrates the change in sphere surface pressure with impedance phase angle $\angle\zeta$ while keeping the overall impedance magnitude constant at a value of $|\zeta|=2$. The corresponding polar plots for $f=2000$ Hz are shown in Fig. 6(b). Again the most significant changes are in the posterior region of the sphere. For a negative impedance phase angle (positive resistive component, negative reactive component) the out-of-phase reflected waves produce a further reduction in the sound diffracted to the rear sphere surface. Conversely a positive impedance

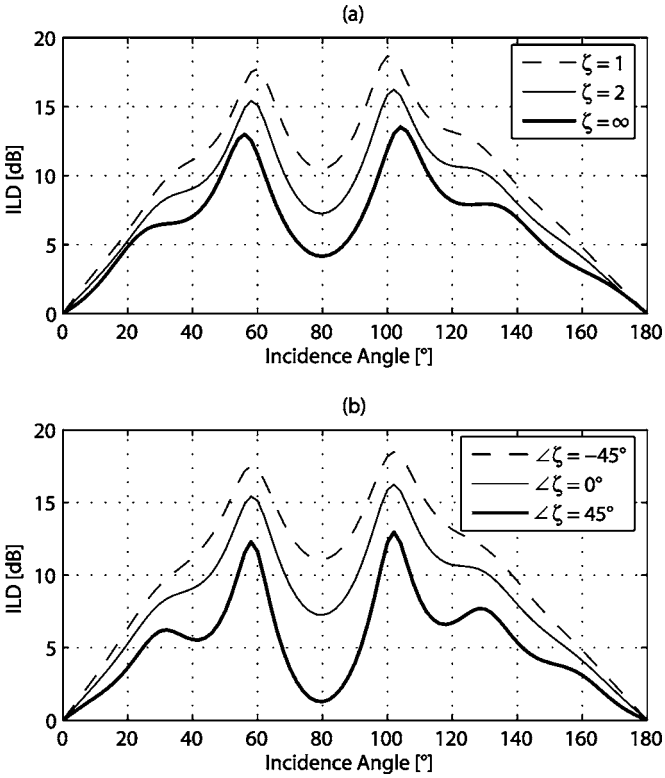


FIG. 7. ILD changes at 3000 Hz with (a) changes in resistive specific impedance and (b) changes in impedance phase angle for constant magnitude of $|\zeta|=2$.

phase angle increases the magnitude of the bright spot lobes in a particularly significant manner. For both cases the effect is less pronounced at lower frequencies.

B. Effect of impedance on the interaural level difference

Figure 7(a) illustrates the resulting changes in ILD at 3000 Hz with a reduction of resistive impedance magnitude. The general decrease in level difference at 80° for all curves is a result of the contralateral ear (offset from the frontal median axis by 100°) being coincident with the principal bright spot. As the impedance magnitude is decreased, the surface pressure reduction in the posterior region of the sphere results in a more pronounced level difference between opposing pinnae, particularly for source directions near the interaural axis. The same trend is exhibited at other frequencies; however below 500 Hz the effect is less significant because the augmented diffraction begins to reduce the ILD to a negligible amount. Figure 7(b) illustrates the resulting effect of the impedance phase angle for a constant magnitude $|\zeta|=2$ on the ILD. A negative phase angle produces an increased ILD in a similar manner to that shown in Fig. 7(a). A positive phase angle produces a decrease, however of particular interest is the amplified decrease for source angles where the contralateral ear coincides with a bright spot of increased magnitude.

Over the frequency range of interest the acoustic properties of most common materials will be governed by their stiffness, and as a result impedance values will have a positive phase angle. The overall decrease in the posterior sur-

face pressure due to the absorption characteristics of the material will be counteracted by the increase in the bright spot lobes due to the positive impedance phase angle. This effect can be seen in Fig. 4 where the change in ILD is positive for source angles where the contralateral ear coincides with a reduction in surface pressure (due to the decreased impedance magnitude), and negative for source angles where it coincides with a bright spot lobe of increased magnitude (due to the positive impedance phase angle). At higher frequencies (in particular those beyond the investigated frequency range) the reduction in posterior surface pressure due to the material absorption becomes dominant. Additionally at these frequencies many materials become governed by their mass properties (and hence the impedance will have a negative phase angle) which further consolidates this effect.

Previous studies into sound localization have shown that when ITD and ILD cues conflict, the ITD cue takes precedence on listeners' judgment of sound direction (Wightman and Kistler, 1992; Macpherson and Middlebrooks, 2002). This is particularly true for low frequencies where there is a limited ILD. It is generally believed that for wideband or low-pass stimuli, the ITD is used to establish the general azimuthal direction, and ILD and monaural spectral cues used to resolve any ambiguities (Shinn-Cunningham *et al.*, 2000). Under these conditions even when the ILD is biased by more than 10 dB, there is little effect on the overall perceived sound direction, although the number of front-back confusions substantially increases for very large ILD biases (Wightman and Kistler, 1997; Macpherson and Middlebrooks, 2002). The change in ILD due to the surface impedance is thus not considered significant in regards to a concrete change in apparent stimulus direction; however an ILD well matched to a modified ITD cue will increase the fidelity of any subsequent virtual 3D projection.

C. Effect of impedance on the interaural time difference

The interaural time difference cue is generally more impervious to system changes than the level difference [for example, for near-field sources (Duda and Martens, 1998)] and this trend continues here and is consistent with the primacy of the ITD as an azimuthal auditory cue when conflicts occur. Figure 8 illustrates the change in ITD with the magnitude of a purely resistive impedance (calculated from the frequency averaged time delay). A general increase in ITD is seen with a decrease in the impedance magnitude, particularly for source angles near the interaural axis. Here the difference in ITD between a completely rigid and highly absorptive sphere is around 100 μs. This difference is enough to shift the perceived source direction by as much as 20°, a value which is significantly more than previously published values for spatial resolution around the interaural axis (Blauert, 1997; Grantham *et al.*, 2003). For source directions closer to the median axis the change in ITD is reduced, however the minimum audible angle threshold in this region is also reduced. Altering the phase angle of the sphere impedance in either direction produces a reduction in ITD. This reduction is a maximum for impedance phase angles of ±90°, and for source angles near the interaural axis (40 μs

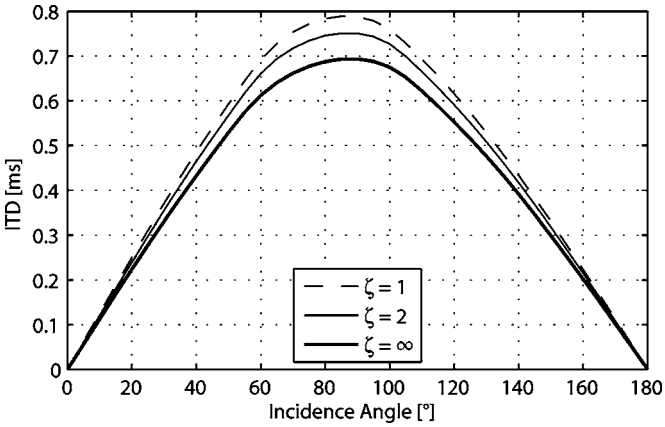


FIG. 8. Interaural time difference (ITD) changes with changes resistive specific impedance value.

for an impedance magnitude of $|\zeta| = 2$).

V. THE EFFECT OF HUMAN HAIR

A. Impedance characteristics of hair

Under natural conditions, human hair is generally of low density and is reasonably porous in composition. Using an impedance tube to measure normal impedance values at an approximated frontal hair surface results in calculated characteristics close to that of air (especially for sparsely populated samples). If the material has a particularly low bulk modulus, the speed of sound propagation through it can become less than c_0 and values of impedance $\xi < 1$ can occur. Using the acoustic properties of hair measured under these conditions with Eq. (8) will not give an accurate description of the surface boundary condition as the sphere (head) is not entirely constituted of the hair material (and if it were flow within the material could not be ignored). The impedance value used must account for the relatively thin layer of the hair material on an approximately rigid backing. To satisfy this requirement, measured values of hair impedance were assumed to be equal to the impedance at the rigid tube termination with the hair sample placed in front. A comparison between locally and nonlocally reacting limp porous layers on a rigid backing is given by Ingard (1981). This study suggests that over the frequency range investigated in the current work, locally and nonlocally reacting layers exhibit similar properties (with slightly improved absorption characteristics evident for the nonlocally reacting layer). Considering the relatively small thickness of the hair layer a locally reactive boundary condition is believed to be a valid assumption.

The properties of three separate hair samples of varying length (15–40 mm) and hair type [medium thickness circular cross section, fine thickness circular cross section, and medium thickness oval cross section (curly)] were measured using the two microphone methodology under the surface assumption described earlier. Samples were aligned arbitrarily in a sample holder (20 mm deep, 60 mm in diameter with a fine mesh front) at three different densities. Figure 9(a) shows the averaged results of absorption coefficient for each density. These match well with previously published

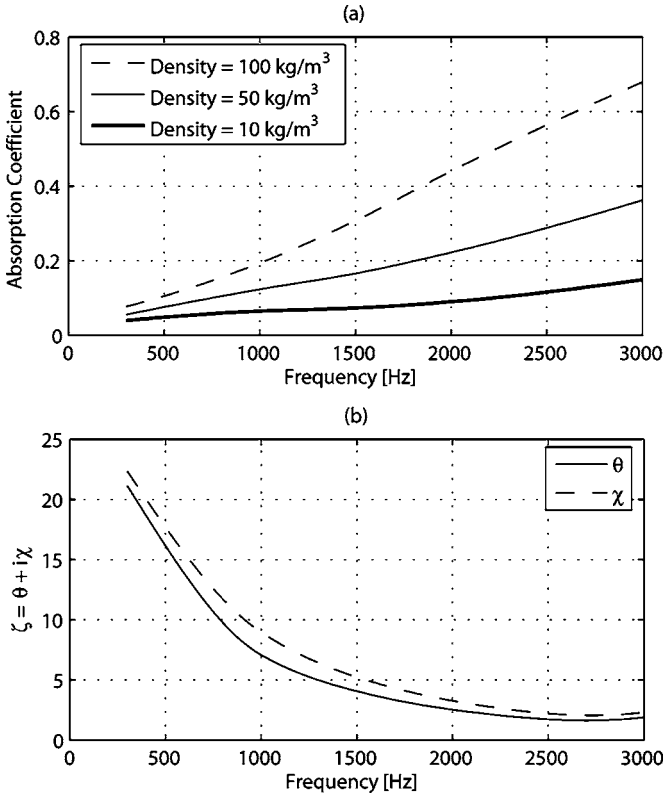


FIG. 9. Measured acoustical properties of human hair in a 20-mm-deep sample holder. (a) Absorption coefficient with varying density. (b) Resistive (θ) and reactive (χ) components of the specific acoustic impedance (ζ) of hair at 100 kg/m^3.

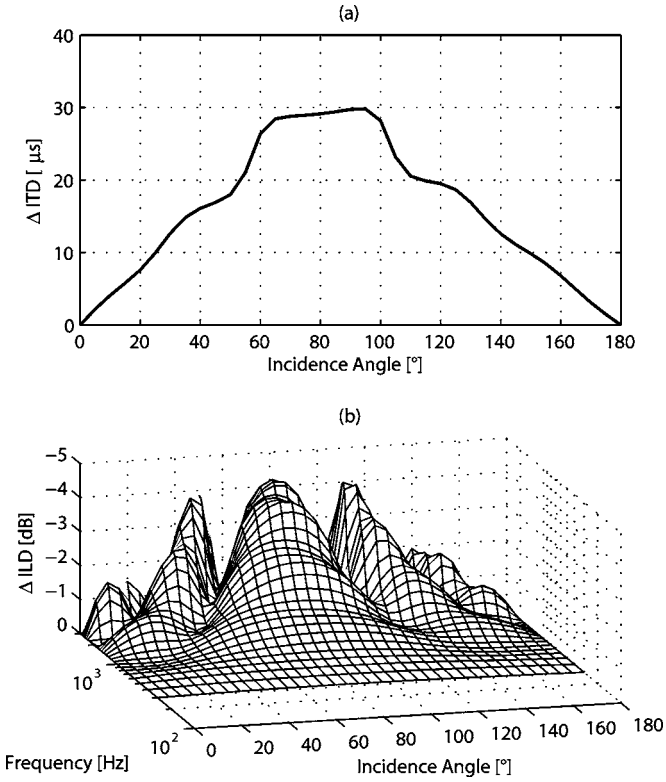


FIG. 10. Predicted changes in (a) ITD and (b) ILD due to the impedance of hair with 100 kg/m^3 density.

hair absorption coefficient data for sample densities of 86–255 kg/m^3 and a sample depth of 15 mm (Katz, 2000). Between sample comparisons provided remarkably consistent results suggesting that the length and type of hair have little effect on the acoustical properties. This is consistent with observations by Riederer (2005).

Figure 9(b) illustrates the resistive and reactive components of the specific impedance for a density of 100 kg/m^3 (again averaged across the three hair samples). The decrease in both the resistive and reactive components with frequency illustrates the increasing absorption. The similarity between the resistive and reactive components results in a relatively consistent positive impedance phase angle of approximately 50°.

B. Changes to interaural cues

Figure 10 illustrates the change in ITD and ILD due to the impedance characteristics of hair. The prediction is made under a locally reactive and uniformly distributed boundary condition assumption using the impedance of hair at a density of 100 kg/m^3 as shown in Fig. 9(b). For source directions near the interaural axis there is an increase in ITD of around 30 μs. Due to the frequency dependent nature of the impedance characteristics, little change is seen in the low-frequency component of the phase delay. This reduces the impact of the impedance on the ITD, which is calculated from the frequency averaged interaural phase delay. For sources lacking in low-frequency information (below 1000 Hz), the change in ITD is more prominent, although at higher frequencies the contribution of the ILD is also more significant.

Published values for just-noticeable-difference changes in the ITD vary depending on the frequency content of the source, and the reference ITD. These range from 10 to 100 μs, with the smaller values generally corresponding to reference ITDs closer to 0, or experiments using pure tones, particularly those at lower frequencies (Mossop and Culling, 1998). Depending on the nature of the hair surface and the listening experiment, including the effects of hair impedance may thus produce a small but noticeable shift in the perceived source direction. The change in ILD due to the inclusion of hair characteristics is shown in Fig. 10(b). Again this change is most significant for source directions close to the interaural axis, with differences on the order of 4 dB. While the magnitude of the changes in ITD and ILD exhibited due to the effect of hair are seemingly small, they are of similar order to changes reported from previous empirical studies into the effect of impedance (Kuhn, 1977; Katz, 2001; Riederer, 2005).

VI. CONCLUSION

The results of experimental validation show that a sphere diffraction model based on a locally reactive and uniformly distributed boundary condition provides a proficient foundation for predicting the changes to the surface pressure and interaural cues in the azimuth plane when the surface is not rigid. Using this model with varying surface impedance gives insight into the behavior of the surface pressure from

which the interaural cues are derived. Decreasing the magnitude of a purely resistive surface impedance results in an overall decrease in the sphere surface pressure level, particularly in the posterior region. This results in nontrivial increases in both the ITD and ILD, especially for sound source directions near the interaural axis. When the surface impedance is complex, the impedance phase angle dictates further changes to the surface pressure and interaural cues. For a negative impedance phase angle the surface pressure exhibits a further decrease. For a positive impedance phase angle the posterior surface pressure exhibits an increase in the magnitude of bright spot lobes (which result from symmetrically diffracted waves arriving in phase). This results in a decrease in the ILD, particularly when the contralateral ear coincides with the location of a bright spot lobe. The ITD is slightly decreased with modification of the impedance phase angle in either direction, with the maximum reduction occurring for a purely reactive surface impedance and source directions near the interaural axis.

Over the frequency range investigated in this study the impedance properties of most acoustic materials will be dictated by their stiffness. This results in a positive impedance phase angle and the decrease in surface pressure in the posterior region of the sphere due to the material absorption will be offset by the increase in the bright spot lobes. Both positive and negative shifts in ILD are therefore possible depending on the exact properties of the scattering surface. At higher frequencies (particularly beyond the range formally investigated in this study) the reduction in surface pressure due to absorption dominates. Additionally at these frequencies many materials are also commonly mass controlled and the decrease in the posterior surface pressure is consolidated by the negative impedance phase angle. This result may be significant for any modeling which utilizes the sphere diffraction solution at higher frequencies (for example, structural models which separately account for head, torso, and pinnae effects). Under the locally reactive and uniformly distributed boundary assumption the impedance characteristics of human hair of density 100 kg/m^3 produce changes in ILD and ITD on the order of 4 dB and 30 μs which are comparable with just-noticeable-difference thresholds.

While the analytical diffraction model presented in this paper is not adequately equipped to directly deal with surfaces of multifaceted impedances, the impetus and significance of the investigation was to examine whether modifying the surface boundary condition would modify the interaural cues to an extent that would warrant further investigation. As hair only covers the upper and rear portion of head surface, it is predicted that the actual changes due to the impedance characteristics of hair would be proportionally less. The presented model thus cannot necessarily provide an accurate direct replacement for the sphere diffraction models that are currently used to compute interaural parameters. A useful annexure and interesting physical study would be to solve the wave equation based on a sphere with position dependent impedance (as is possible for finite and boundary element computational models). This would provide a more accurate account of the effect of the surface impedance characteristics on the low-frequency interaural cues, and insight into the physical mechanisms of scattering around a human head with hair.

ACKNOWLEDGMENTS

B.E.T. would like to acknowledge the support of the Robert and Maude Gledden, and F S Shaw Memorial Postgraduate Scholarships. The insightful questions and comments by the anonymous reviewers and the associate editor were also appreciated.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001**). "Estimation of a spherical-head model from anthropometry," J. Audio Eng. Soc. **49**, 472–479.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002a**). "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064.

Algazi, V. R., Duda, R. O., and Thompson, D. M. (**2002b**). "The use of head-and-torso models for improved spatial sound synthesis," *Proceedings of the 113th AES Convention*, Audio Engineering Society, Los Angeles, CA, preprint 5712.

Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Bronkhorst, A. W. (**1995**). "Localization of real and virtual sources," J. Acoust. Soc. Am. **98**, 2542–2553.

Brown, C. P., and Duda, R. O. (**1998**). "A structural model for binaural sound synthesis," IEEE Trans. Speech Audio Process. **6**, 476–488.

Duda, R. O., and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

Grantham, D. W., Hornsby, B. W. Y., and Erpenbeck, E. A. (**2003**). "Auditory spatial resolution in horizontal, vertical, and diagonal planes," J. Acoust. Soc. Am. **114**, 1009–1022.

Gumerov, N. A., and Duraiswami, R. (**2002**). "Computation of scattering from *N* spheres using multipole reexpansion," J. Acoust. Soc. Am. **112**, 2688–2701.

Ingard, K. U. (**1981**). "Locally and nonlocally reacting flexible porous layers; A comparison of acoustical properties," J. Eng. Ind. **103**, 302–313.

ISO 10534-2 (**1998**). Determination of Sound Absorption Coefficient and Impedance in Impedance Tube. 2. Transfer-Function Method (ISO).

Katz, B. F. G. (**2000**). "Acoustic absorption measurement of human hair and skin within the audible frequency range," J. Acoust. Soc. Am. **108**, 2238–2242.

Katz, B. F. G. (**2001**). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparison to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455.

Kear, G. (**1959**). "The Scattering of Waves by a Large Sphere for Impedance Boundary Conditions," Ann. Phys. (N.Y.) **6**, 102–113.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.

Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (**1995**). "On the minimum-phase approximation of head-related transfer functions," *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, 1995*, (IEEE, New Paltz, New York), pp. 84–87.

Lax, M., and Feshbach, H. (**1948**). "Absorption and scattering for impedance boundary conditions on spheres and circular cylinders," J. Acoust. Soc. Am. **20**, 108–124.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Middlebrooks, J. C. (**1999**). "Virtual localization improved by scaling nonindividualised external ear transfer functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1510.

Middlebrooks, J. C., Macpherson, E. A., and Onsan, Z. A. (**2000**). "Psychophysical customization of directional transfer functions for virtual sound localization," J. Acoust. Soc. Am. **108**, 3088–3091.

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

Mossop, J. E., and Culling, J. F. (**1998**). "Lateralization of large interaural delays," J. Acoust. Soc. Am. **104**, 1574–1579.

Riederer, K. A. J. (**2005**). "HRTF analysis: Objective and subjective evalu-

ation of measured head-related transfer functions," Helsinki University of Technology, Helsinki.
Shinn-Cunningham, B. G., Santarelli, S., and Kopco, N. (**2000**). "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.
Wightman, F. L., and Kistler, D. J. (**1997**). "Monaural sound localization revisited," J. Acoust. Soc. Am. **101**, 1050–1063.
Woodworth, R. S., and Schlosberg, H. (**1954**). *Experimental Psychology* (Holt, New York).
Zotkin, D. N., Hwang, J., Duraiswami, R., and Davis, L. S. (**2003**). "HRTF personalization using anthropometric measurements," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2003* (IEEE, New Paltz, NY), pp. 157–160.

The role of the external ear in vertical sound localization in the free flying bat, *Eptesicus fuscus*

Chen Chiu and Cynthia F. Moss
Department of Psychology, Neuroscience and Cognitive Science Program, University of Maryland, College Park, Maryland 20742

(Received 16 November 2006; revised 21 December 2006; accepted 22 December 2006)

The role of the external ear in sonar target localization for prey capture was studied by deflecting the tragus of six big brown bats, *Eptesicus fuscus*. The prey capture performance of the bat dropped significantly in the tragus-deflection condition, compared with baseline, control, and recovery conditions. Target localization error occurred in the tragus-deflected bat, and mainly in elevation. The deflection of the tragus did not abolish the prey capture ability of the bat, which suggests that other cues are available used for prey localization. Adaptive vocal and motor behaviors were also investigated in this study. The bat did not show significant changes in vocal behaviors but modified its flight trajectories in response to the tragus manipulation. The tragus-deflected bat tended to attack the prey item from above and had lower tangential velocity and larger bearing from the side, compared with baseline and recovery conditions. These findings highlight the contribution of the tragus to vertical sound localization in the free-flying big brown bat and demonstrate flight adaptations the bat makes to compensate altered acoustic cues.
 [DOI: 10.1121/1.2434760]

PACS number(s): 43.66.Qp, 43.80.Ka, 43.66.Pn, 43.60.Jn, 43.80.Lb [JAS] Pages: 2227–2235

I. INTRODUCTION

Echolocating bats produce ultrasonic vocalizations and listen to echo returns to localize prey items and obstacles. They rely on biological sonar to accurately localize insects in a dynamic acoustic environment in which predator and prey are in continuous motion. Sound localization in bats, like other mammals, is accomplished largely via auditory computations on direction-dependent acoustic signals. Horizontal sound localization depends on binaural comparisons, such as interaural level difference (ILD) and interaural time difference (ITD), while vertical sound localization relies largely on spectral cues generated by the external ear.

The external ear of echolocating bats serves as a receiver to collect sound and is important for localizing auditory targets. The external ear of most bat species consists of two major parts, the pinna and the tragus (Fig. 1). The tragus is a piece of skin that stands in front of the ear canal and may affect the incoming acoustic signal. The size of the tragus varies across bat species but is typically a prominent structure, particularly compared with other mammalian ears.

It is generally believed that the tragus can generate spectral cues for vertical sound localization. Spectral notches in the head-related transfer function (HRTF) are elevation dependent, as reported in several bat species [*Phyllostomus discolor* (Firzlaff and Schuller, 2003), *Pteronotus parnellii* (Firzlaff and Schuller, 2004), *Antrozous pallidus* (Fuzessery, 1996), *Eptesicus fuscus* (Aytekin *et al.*, 2004; Müller, 2004; Wotton *et al.*, 1995; Wotton and Jenison, 1997)]. Previous studies have shown that spectral cues produced by the external ear are important for vertical sound localization in humans (Batteau, 1967; Bloom, 1977; Carlile *et al.*, 2005; Fisher and Freedman, 1968; Middlebrooks and Green, 1991; Oldfield and Parker, 1986) as well as other animal species (Heffner *et al.*, 1996; Parsons *et al.*, 1999).

Several studies have addressed the functional contribution of the tragus to elevation-dependent spectral cues. Grinnell and Grinnell (1965) removed the contralateral tragus of the ear of *Plecotus townsendii* and recorded the evoked potential from the inferior colliculus (IC). Wotton *et al.* (1995) measured elevation-dependent changes in acoustic signals at the tympanic membrane of the big brown bat, *E. fuscus*, both before and after tragus removal. These two studies each reported sound elevation effects of tragus deflection, which occur below the bat's eye-nostril plane. Aytekin *et al.* (2004) found that tragus removal produced no change in elevation-dependent spectral notches of the big brown bat's HRTF in the frequency range of 30 to 50 kHz, as Wotton *et al.* (1995) reported. Instead, they found that the tragus contributed to the gain and directionality of the HRTF at 70 to 90 kHz. A similar HRTF study on another species, *Phyllostomus discolor*, reported that tragus deflection produced a significant decrease in the depth of a spectral notch at about 55–60 kHz (Firzlaff and Schuller, 2003). All studies to date reported some degree of change in characteristics of the HRTF when the tragus is removed. However, the nature and extent of change varies across studies and bat species. No research findings suggest that tragus removal abolishes elevation-dependent spectral notches, indicating that other sources of spectral cues may play a role in vertical sound localization, even if they must be relearned following changes to the external ear.

Psychoacoustic experiments on *E. fuscus* have also suggested that the tragus contributes to vertical sound localization, particularly below the horizon. The bat's ability to discriminate vertical angle deteriorates when the tragus is deflected (Lawrence and Simmons, 1982). Vertical angle

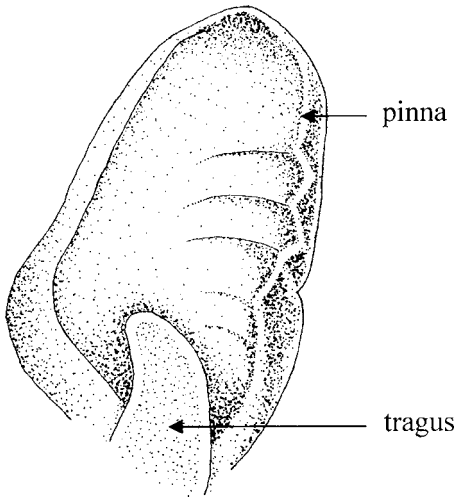


FIG. 1. Drawing of the external ear of *Eptesicus fuscus*, including the pinna and the tragus (drawn by Kweelen Lee).

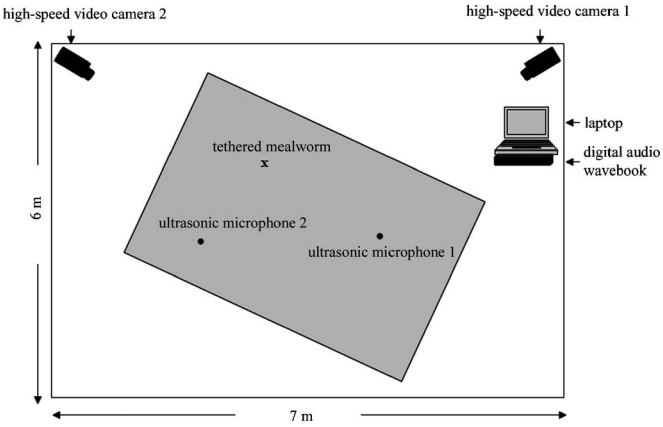


FIG. 2. Schematic of setup for video and sound recordings of tethered prey capture by echolocating bats. Two high-speed IR cameras (Kodak MotionCorder Analyzer, 240 frames per second) were mounted in the room to permit 3D reconstruction of the bat's flight path. Video recordings were synchronized with audio recordings taken with two ultrasonic microphones delivering signals to an IOTech Wavebook.

acuity (VAA) in tragus-deflected bats is impaired for positions below the horizon, but not above the horizon (Wotton and Simmons, 2000). While past studies on the role of the tragus on vertical sound localization are suggestive, none have directly examined its importance in natural behaviors, namely on the precise localization required for insect capture.

Another question that remains to be answered is the extent to which an animal can adapt to modifications of the external ear that alter the acoustic cues used for vertical sound localization. Plasticity of sound localization has been studied in a broad range of animal species, including humans. Several studies demonstrate that plasticity can take place in adulthood, as long as a sufficient practice period is allowed (Hofman *et al.*, 1998; King *et al.*, 2000; Knudsen *et al.*, 1994; Linkenhoker and Knudsen, 2002; Van Wanrooij and Van Opstal, 2005). In addition, the degree and time period of adaptation in spatial hearing depends on the sound localization task.

There are two purposes of this study, first to investigate the influence of tragus deflection on prey capture behavior, with a particular emphasis on target localization in the vertical plane, and second to measure adaptive motor behaviors in response to changes in the acoustic cues believed to contribute to vertical sound localization.

II. METHODS

A. Experimental animals

Six big brown bats, *E. fuscus*, were used in the experiment. They were housed in an animal colony room at the University of Maryland, College Park, MD. The temperature and humidity in the facility were maintained at 24–28 °C and 30–50%, respectively. The light/dark cycle was reversed and maintained at 12 h, with lights off at 7:00 am so that bats were run in experiments during their active period. Bats were housed in small groups with two to four individuals in one cage, with free access to fresh water. They were maintained at approximately 80% of *ad lib* feeding weight and ate only when they successfully took tethered mealworm during experimental trials.

B. Behavioral experiment

Experiments were run between May and September when *E. fuscus* were most active. All the experimental trials were conducted in a large carpeted flight room ($7\times6\times2.5$ m^3) with walls and ceiling lined with acoustic foam (Fig. 2). In order to eliminate the bat's use of visual cues, long-wavelength lighting (>650 nm) was used as the only light source in the flight room (Hope and Bhatnagar, 1979). Each bat was trained inside the flight room to catch tethered mealworms hung in random locations from the ceiling and with different string lengths (0.5, 0.75, 1, 1.25, and 1.5 m) to present insect prey at variable elevations. The data collection began after the bat performed the task at a minimum success rate of 75%.

C. Data collection

1. Audio recordings

Two ultrasonic microphones (UltraSound Advice, London) were placed on the floor to pick up vocalizations of the bat and stored digitally in a Wavebook (IOTech, sample rate 250 kHz per channel). These audio recordings were analyzed off-line using a custom MATLAB program to measure spectral and temporal features of echolocation calls produced by the bat performing the insect capture task.

2. Video recordings

Two high-speed video cameras (Kodak MotionCorder Analyzer, Model 1000, 240 frames per second) were mounted on two corners of the flight room to capture the motion of the flying bat. Video recordings from these two cameras were then digitized and analyzed off-line using commercial hardware and software (Peak Performance Technologies and MATLAB) to reconstruct the 3-D flight path of the bat.

3. Audio-video synchronization

Audio and video recordings were end-triggered simultaneously by the experimenter when the bat made or attempted contact with the mealworm and the preceding eight seconds of data were stored.

D. Tragus manipulation

Alteration of acoustic signals received at the bat's tympanic membrane was accomplished by gluing the tragus forward to the side of the head by Vetbond (3M) or Prosthetic Adhesive (Ben Nye). The glue was applied every day before the experiment started and served to hold down the tragus for approximately three hours (two hours after completion of experimental trials). There were four distinct experimental conditions: baseline, control, tragus-deflection, and recovery. Each condition was run over four successive days, except the control condition, which was run one day, and the entire experiment involved a total of 13 test days for each bat.

The behavioral task was identical in all four conditions. The baseline condition tested the prey capture performance of the individual bat with unmanipulated external ears. In the control condition, a drop of water was applied to the tragus, using the same procedures as the tragus-deflection condition without actually gluing down the tragus. The purpose was to determine if any change in the prey capture performance could be attributed to disturbance created by touching the bat's external ear. The tragus-deflection condition examined changes in the bat's prey capture performance when both tragi were glued down. The recovery condition was run after both tragi came up and documented the bat's behavior after the experimental manipulation to the external ears. The position of the tethered mealworm was changed every trial to prevent the bat's use of spatial memory rather than echolocation to perform the insect capture task.

E. Data analysis

Several parameters acquired from audio recordings were used to measure the bat's vocal behavior and are listed as follows: (1) spectral features of echolocation calls: start frequency (the highest frequency of the fundamental), end frequency (the lowest frequency of the fundamental), and bandwidth (the frequency range of the entire fundamental); (2) temporal features of echolocation calls: duration (the duration of the fundamental) and pulse interval (the time interval between the onset of two successive calls); and (3) terminal buzz duration, defined as the sound segment prior to insect capture or attempted capture with pulse intervals less than 8 ms.

Previous studies have shown that the tragus may play a role in vertical sound localization; thus the analysis of motor behavior was emphasized in the plane of elevation. Flight behavior was measured from video recordings and the following parameters were used: (1) trial time: from the moment the bat took off to when the bat made contact with the mealworm, (2) the elevation offset between the bat and the prey [Fig. 3(a)], (3) the tangential velocity of the bat in the vertical plane (side view) [Fig. 3(b)], and (4) the bearing in the vertical plane [Fig. 3(c)]. The bearing is the angle between two vectors, which are the vector of the bat's tangential velocity and of the bat-worm vector (vector from the bat to the mealworm). The first vector represents the actual direction the bat is heading, and the second one is the direction from the bat to the worm.

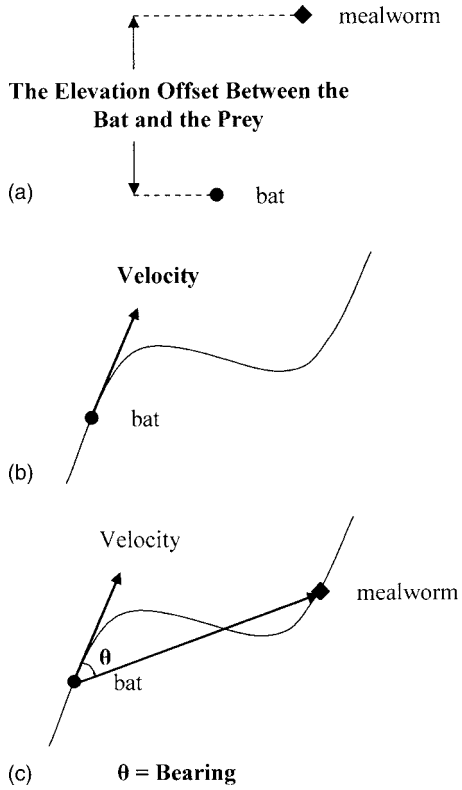


FIG. 3. Measurements of adaptive motor behavior. (a) The elevation offset between the bat and the prey; (b) the velocity of the bat from the side view; and (c) the bearing from the side view.

All vocal and motor behavior analyses were carried out for trial segments within one second before contact with the tethered mealworm. In addition, only the vocal and motor behaviors of the direct target hit trials were included to study adjustments of these behaviors following the tragus manipulation. Repeated measurement ANOVA was used to test statistical differences in data across conditions. Bonferroni adjustments were used to correct for additive errors associated with multiple tests in *post-hoc* analyses, e.g., $0.05/n$, where $n=10$.

III. RESULTS

A. Performance

Three insect capture behavior patterns were categorized from video recordings, i.e., direct target hit, target contact, and far miss. Direct target hit was the most typical pattern in the prey capture behavior. The bat approached the mealworm and used its tail membrane to scoop up the mealworm. Target contact was recorded when the bat attempted insect capture with a body part other than the tail membrane (such as left/right wing, mouth, etc.). The bat may successfully consume the tethered mealworm or drop it in the contact behavior described above, but in either case the bat made physical contact with the target. Far miss occurred when the bat failed

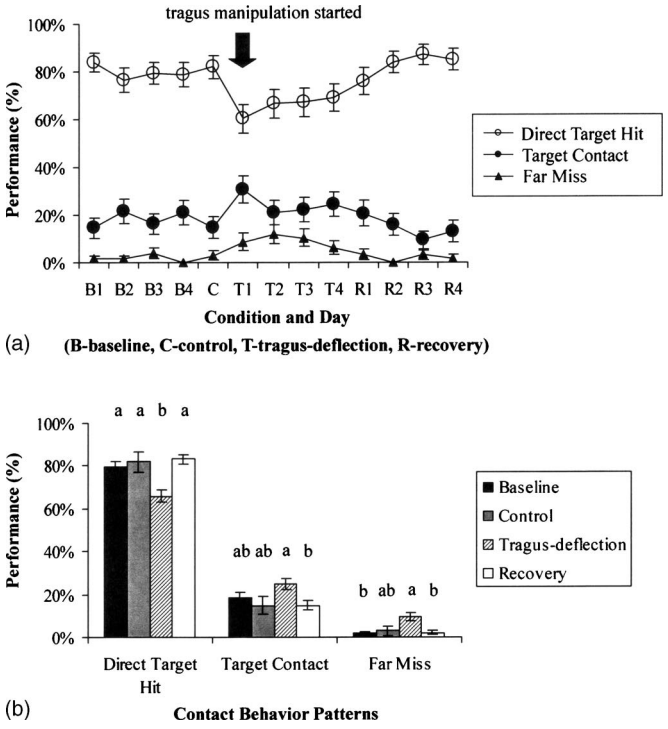


FIG. 4. Prey capture performance. (a) Prey capture performance under different conditions over repeated test days. The open circle summarizes direct target hits, the closed circle shows target contacts, and the closed triangle shows far misses. The x axis represents the conditions (B as baseline, C as control, T as tragus-deflection, and R as recovery) and the number refers to test days 1 to 4. (b) Prey capture performance under different conditions. The letters above the histograms represent the rank of the performance. The same letter means no significant difference.

to hit the actual target. The first pattern characterizes the bat's precise localization of its prey. The second and third patterns show localization errors of different magnitudes.

The prey capture performance of all six bats is shown in Fig. 4. The Fisher exact test (Zar, 1996) was used to analyze the performance change across days and conditions. Within the same condition, there is no significant difference in performance across different days [$p > 0.05$, Fig. 4(a)]. Direct target hit is the most frequent behavior pattern across all four experimental conditions, and target contact and far miss trials increase in the tragus-deflection condition. The direct target hit trials remain at around 80% in the baseline condition. The performance of the control condition is comparable to that of the baseline condition. There is a drop in the percentage of direct target hit trials and a rise in target contact and far miss trials on the first day of the tragus deflection condition; performance in the tragus manipulation condition gradually returns to the baseline level. The percentage of direct target hits is higher on the first day of recovery compared with the tragus-deflection condition but lower than in baseline trials. The performance of the following three days of recovery data is similar to the baseline condition. Collapsing data across days, the percentage of direct target hit trials in the tragus-deflection condition is the lowest, and the percentages of target contact and far miss trials are the highest [Fig. 4(b)].

We also analyzed the interaction position of the bat with respect to the insect across conditions. The moment the bat

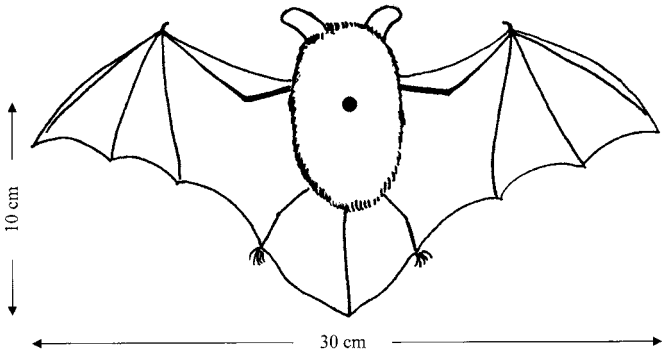


FIG. 5. The range of capture measurement in *E. fuscus*. The black dot on the bat's body is the center of the bat.

made contact with the mealworm is defined as interaction time. The bat's position at this time is referred to as the interaction position, and the distance between the bat and the prey at the interaction time is defined as the interaction distance. Because the bat can catch the mealworm using not only its tail membrane but also the wing, the range of capture is defined by the wingspan and body length of the bat (Fig. 5). The wingspan (30 cm) determines the horizontal range (x and y planes) and the length between the center of the body and the tip of the tail (10 cm) determines the vertical range (z plane) that the bat can reach. To examine in detail how the tragus manipulation influences interaction distance of the bat, the number of trials that exceed this range is shown in Table I across conditions. The interaction distance exceeds the range of capture in the z plane in significantly more trials when the tragus was glued down compared with baseline and recovery conditions. However, the tragus manipulation has no effect on the interaction distance in x and y planes.

B. Adaptive vocal behavior

The terminal buzz duration (Fig. 6) in both tragus-deflection and recovery conditions is significantly longer

TABLE I. The interaction distance under three different tragus conditions.

Dimension	Tragus condition	Trials exceed (x or $y > 15$ or $z > 10$)	%	p	*Post hoc* test
x **plane**	Baseline	2	0.74		
	Tragus-deflection	3	1.14		
	Recovery	1	0.41	n.s.	
y **plane**	Baseline	4	1.48		
	Tragus-deflection	3	1.14		
	Recovery	1	0.41	n.s.	
z **plane**	Baseline	3	1.11		b
	Tragus-deflection	12	4.56		a
	Recovery	5	2.07	<0.05	ab
Distance	Baseline	3	1.11		b
(3-D)	Tragus-deflection	14	5.32		a
	Recovery	5	2.07	<0.01	ab

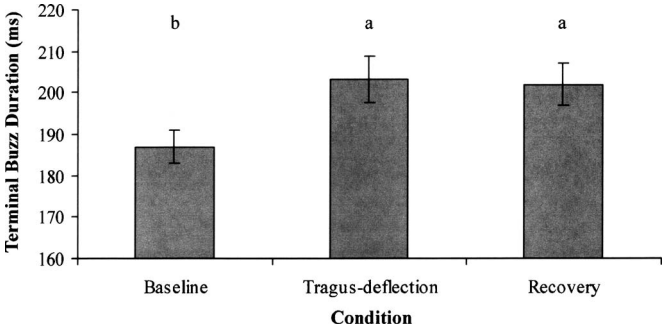


FIG. 6. Sonar buzz duration across the three different conditions, baseline, tragus-deflection, and recovery. The letter in the histogram represents the rank of the buzz length.

than in the baseline condition (one-way ANOVA, $p<0.05$). The features of vocalizations were analyzed in 100-ms time blocks during the final 1000 ms before the bat captured the prey item. Only direct target hit trials were included in the analysis of adaptive vocal behavior to examine if the bat modified its echolocation calls in order to catch the prey successfully. No reliable pattern of change in the vocalizations emerged from these analyses when comparing the baseline, tragus-deflection, and recovery conditions.

C. Adaptive motor behavior

There is no significant difference in trial time from release to capture across baseline (17.78±1.78 s), tragus-deflection (19.94±2.42 s) and recovery conditions (16.84±1.88 s). Although the tragus-deflection condition shows the largest average trial time compared with the other two conditions, the difference is not statistically significant (one-way ANOVA, $p>0.05$).

The adjustment of distance (between the bat and the prey) and bearing in the tragus-deflection condition is shown in Fig. 7. The magnitude of adjustment is computed from the distance and bearing difference between baseline and tragus-deflection conditions (the mean distance/bearing in the tragus-deflection condition subtract by the mean distance/bearing in baseline condition). The distance [Fig. 7(a)] and bearing [Fig. 7(b)] differences in the vertical plane are similar to differences in the horizontal plane in the last half second, but show larger differences in the vertical plane than in the horizontal plane before 0.5 s before contact. The modifications of flight path in the tragus-deflection condition are more prominent in the vertical than the horizontal plane.

The bat tended to attack the mealworm from above when tragi were glued down. The elevation offset between the bat and the prey in the tragus-deflection trials is significantly larger than in the baseline condition during the entire last second before contact [Fig. 8(a)]. The recovery condition shows the smallest elevation offset between the bat and the prey and even smaller than the baseline condition for half the time segments (five out of ten time segments). The bat flew slower in the tragus-deflection condition [Fig. 8(b)]. In the tragus-deflection condition, the bat first shows higher side tangential velocity than in the baseline condition and lowers it and then raises it again in the last 0.1 s before contact. The side tangential velocity in the recovery condition shows no

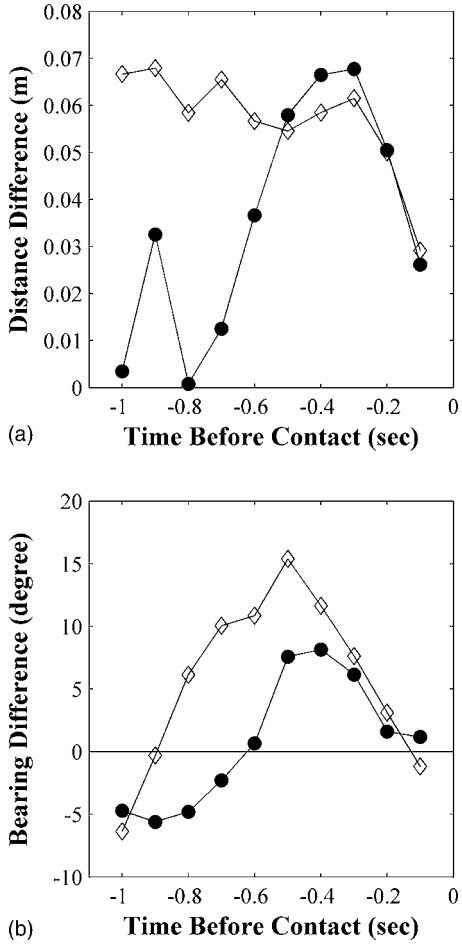


FIG. 7. The adjustment of flight path in different planes in the tragus-deflection condition. (a) Distance difference and (b) bearing difference in the horizontal (closed circle) and vertical (open diamond) planes. The difference is computed from the difference between mean values in baseline and tragus-deflection conditions in every time segment.

significant difference compared with the baseline condition in most time segments, except three (0.7, 0.6, and 0.1 s before target contact, $p<0.005$), and the differences between baseline and recovery conditions are not as large as the differences between tragus-deflection and recovery conditions. The bearing from the side view is larger in the tragus-deflection than in the baseline condition during 0.8 to 0.2 s before contact [$p<0.005$, Fig. 8(c)]. The recovery of the bearing is not complete and, in three time segments (0.4 to 0.2 s before contact, $p<0.005$), the bearing is significantly different from the baseline condition.

The prey capture performance dropped most dramatically on the first day of the tragus-deflection condition. Therefore, the motor behavior data from the first test day were analyzed in detail. The motor behavior of different attack patterns, direct target hit and target contact, was also compared here. Far miss trials were excluded from this analysis due to the small sample size. To simplify the description of the results on adaptive motor behaviors in the first day of tragus-deflection, we summarized the findings separately for the baseline condition direct target hit (B-DH) trials, the first day tragus-deflection condition direct target hit (1st T-DH) trials, and the first day tragus-deflection condition

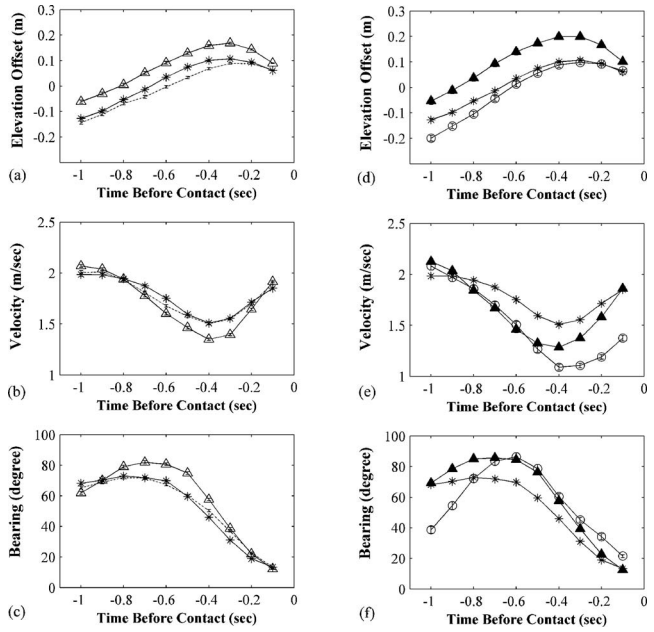


FIG. 8. The bat's adaptive motor behavior. (a) The elevation offset between the bat and the prey; (b) the velocity of the bat from the side view; and (c) the bearing from the side view in direct hit trials across the three conditions: baseline (asterisk), tragus-deflection (open triangle), and recovery (dot). The bat's adaptive motor behavior: (d) the elevation offset between the bat and the prey; (e) the velocity of the bat from the side view; and (f) the bearing from the side view, in the three conditions, baseline condition direct target hit (B-DH) trials (asterisk), the first day tragus condition direct target hit (1st T-DH) trials (closed triangle), and the first day condition target contact (1st T-C) trials (open circle). Error bars represent the standard error of the mean.

target contact (1st T-C) trials. Comparing these three different data sets provides information about how the bat modified its motor behaviors to enable insect capture. We hypothesize that the bat adapted its motor behaviors in response to changed acoustic input as a result of the tragus manipulation.

Following the tragus manipulation, the bat maintains almost the same elevation offset in 1st T-C trials, compared with B-DH trials in the last 0.7 s before prey capture [Fig. 8(d)]. On the other hand, 1st T-DH trials show significantly larger elevation offset between the bat and the prey than the other two conditions ($p<0.005$). This result is consistent with our hypothesis stated above. The bat shows significantly lower side tangential velocity in 1st T-C and 1st T-DH trials compared with B-DH trials in the last 0.7 s before capturing the prey [$p<0.005$, Fig. 8(e)]. The 1st T-C trials have the lowest side velocity across three data sets ($p<0.005$). In the last 0.1 s before prey capture, the bat shows the same side velocity in both 1st T-DH trials and B-DH trials. The tragus-deflected bat only made contact with the tethered mealworm when the side velocity at the last moment did not reach the baseline level. The 1st T-DH trials show significantly larger bearing from the side view than B-DH trials in the final second before prey capture ($p<0.005$), except for the beginning and end of this period [Fig. 8(f)]. The 1st T-C trials show smaller bearing in the beginning of the last 1 s before contact (1 and 0.9 s before contact) and the bearing increases significantly over B-DH trials ($p<0.005$), but is similar to 1st T-DH trials ($p>0.005$). The bearing in 1st T-DH trials is closer to B-DH trials than 1st T-C trials in the final 0.1 s before prey capture. The difference in bearing across conditions inthe final 0.1 s of a trial seems critical to the outcome of prey capture, i.e., direct target hit or off-axis contact of the prey item. Although these results on the velocity and bearing do not statistically support our hypothesis, adjustments of motor behaviors in the very last moment have immediate consequences on prey capture success.

IV. DISCUSSION

A. The influence of tragus deflection on prey capture performance and sound localization

Tragus deflection reduced sound localization accuracy and decreased successful prey capture performance of the big brown bat, with the largest effect on the first test day after the experimental manipulation of the external ear. Similar performance in control and baseline conditions demonstrates that the drop in the prey capture performance under the tragus-deflection condition is caused by changes in acoustic cues used for prey localization. Over test days, the bat adapted to the changes in acoustic cues introduced by tragus-deflection and successfully captured tethered prey after some experience with altered external ears. This result suggests that the bat can adapt quickly to altered acoustic cues for prey localization. The recovery and baseline conditions did not show significantly different performance, which suggests that the bat can switch back to using baseline acoustic cues for sound localization. These results are consistent with human studies: Introducing new spectral cues to the human ear via pinna molds increased sound localization error, particularly in the vertical plane (Fisher and Freedman, 1968; Hofman *et al.*, 1998; Oldfield and Parker, 1984; Van Wanrooij and Van Opstal, 2005). However, subjects regained the vertical sound localization ability after a few days of experience, and the newly learned cues did not interfere with the old ones (Hofman *et al.*, 1998; Van Wanrooij and Van Opstal, 2005).

In the present study, the percentage of trials exceeding the range of capture is used as an index of sound localization error. The more trials exceeding the range of capture, the more consistent is the error. In the vertical plane, the most trials exceeding the range of capture occurred in the tragus-deflection condition compared with baseline and recovery conditions. Tragus-deflection produced no effect on interaction distance in the other two planes. This indicates that the tragus-deflection has the largest effect on vertical sound localization. Previous behavioral studies of vertical localization in *E. fuscus* also came to similar conclusions with different experimental designs (Lawrence and Simmons, 1982; Wotton and Simmons, 2000).

The bat's prey capture performance decreased after tragi were glued down. The performance dropped significantly but did not drop below 50%, which suggests that prey capture ability of *E. fuscus* is not heavily dependent on the contribution of the tragus. This result is consistent with HRTF studies on the echolocating bat, which show some spectral changes following tragus deflection, but they are not very dramatic (Aytekin *et al.*, 2004; Firzlaff and Schuller, 2003; Grinnell and Grinnell, 1965; Müller, 2004; Wotton *et al.*, 1995).

Müller *et al.* (2006) demonstrate that the tragus, as well as the lower ledge of the pinna rim, introduces similar contributions to the directivity patterns in *Nyctalus plancyi*. It is suggested that the spectral cues introduced by the tragus can facilitate sound localization in the vertical plane. However, the contribution of the tragus is limited, and the present study demonstrates that the bat can adapt to changes in the filtering characteristics of the external ear. Although the big brown bat does not have a prominent lower ledge of the pinna rim, other parts of the external ear, such as the ridge along the pinna, may also contribute to sound localization. Human and bat studies have shown that auditory cues for horizontal and vertical sound source localization are not independent (human: Butler and Humanski, 1992; Gardner, 1973; bat: Aytekin *et al.*, 2004; Fuzessery, 1996). Therefore, changes in certain spectral cues caused by tragus-deflection may be compensated by other cues. Therefore, the tragus can contribute to the acoustic cues for vertical sound localization, but they are not exclusive.

B. Sensory-motor adaptation

Two highly interrelated systems, sensory and motor, are required for successful prey capture in the echolocating bat. The bat must localize the source of echoes reflected from prey and use this spatial information to guide motor systems to enable appropriate commands for prey capture. The bat relies upon precise sound localization of prey through binaural and monaural acoustic cues. The effect of the tragus on vertical sound localization has already been described above. Successful prey capture also depends on accurate motor control of the body. Distorted acoustic information about object location is expected to elicit errors in motor behaviors.

Since humans rely heavily on vision and bats on audition to perceive their spatial surroundings, there may be some relevant parallels to explore in sensory-motor adaptations. Several human studies have introduced distorted or rotated visual information to subjects who are required to produce movements to accomplish task-specific goals (Abeele and Bock, 2001; Cunningham, 1989; Cunningham and Welch, 1994; Imamizu *et al.*, 1998; Kagerer *et al.*, 1997; Marotta *et al.*, 2005; Martin *et al.*, 2002; Stratton, 1896, 1897a, b; Van Beers *et al.*, 2002; Yoshimura, 2002). Redding *et al.* (2005) indicate that prism exposure involved three adaptive processes, which are postural adjustments, strategic control, and spatial realignment. All these studies demonstrate that humans show plasticity in visual-motor control and are capable of selecting suitable locomotion to adapt to distorted visual cues. A related study on rhesus monkeys reported that nonhuman primates acquire and generalize visual-motor transformations as do humans (Paz *et al.*, 2005).

In the present study of altered sensory input, the big brown bat attacked from higher elevation in the tragus-deflection condition than the baseline condition. In addition, the trials in which the bat contacted the target show similar flight path characteristics to the baseline condition, suggesting that modifying the flight path can increase the prey capture performance of the bat. The bearing from the side view also shows a larger bearing in the tragus-deflection condition than the tragus-intact condition, including baseline and recovery. These flight path modifications are the most robust and consistently significant changes in the bat's motor behavior in response to altered acoustic cues for vertical sound localization in the bat. Similar trajectory modification has also been reported in human visual-motor adaptation studies (Abeele and Bock, 2001; Contreras-Vidal *et al.*, 2005; Cunningham, 1989; Seidler, 2005; Wolpert *et al.*, 1995a).

Altered acoustic cues for sound localization in this study bear the same relation to altered visual spatial cues in human studies. Human subjects wearing prisms that shift or rotate visual input showed hand trajectories that deviate from the original when asked to point to a target, but they also corrected the hand trajectory after some practice with feedback. Visual feedback is important for motor behavior adaptation (Redding and Wallace, 1994). A forward model predicts the outcome of the motor behavior and an inverse model records the signals, which are derived from the error between predicted and actual outcomes, used to select a motor command to reduce performance error. The trajectory change is the result of a motor learning process. The forward and inverse models are tightly coupled together and capable of explaining motor learning in humans (Kawato, 1999; Kawato and Wolpert, 1998; Wolpert and Kawato, 1998; Wolpert *et al.*, 1995b).

The same internal model can be applied to explain the bat's motor behavior adaptation in this study. The forward model in the bat predicts the target position and drives suitable motor commands for the animal to successfully intercept the mealworm. The bat typically captures the prey by positioning itself just above the prey item to scoop it up with the tail membrane. When a localization error occurs, the bat may still be able to make contact with the target, but with the wing or the mouth instead of the tail membrane. Through contact with the prey, the bat acquires information about the actual target position. The discordance between the estimated and actual target positions generates a motor error. The motor error signal is conveyed to the inverse model and permits further correction in the next motor command, by adjusting the flight path approach (the elevation offset between the bat and the prey) and angle (the bearing from the side view). Therefore, even when the bat makes an error in localizing the tethered mealworm position in the tragus-deflection condition, it can still use dynamic auditory feedback to correct its motor behavior and initiate a proper motor command to successfully intercept the target.

Other human visual-motor research shows that decreasing the reaction time increases the performance error. There is a trade-off between reaction time and accuracy of pointing to the target location (Fitts, 1966). Although the trial time of the bat in this study did not show any significant difference across baseline, tragus-deflection, and recovery conditions, the approaching side velocity did show significant differences across these three experimental conditions. The result of lowering the side velocity suggests a trade-off between speed and accuracy. A slower velocity may provide the bat with the additional time needed to compensate the alteration of information from the experimental manipulation. The

slower side velocity in the first day contact trials suggests that the bat slowed down to correct its approach for attempted insect capture.

Redding and Wallace (2002) proposed two adaptation processes in human visual prism experiments: strategic calibration and spatial alignment. Prism goggles disrupt the relationship between extrinsic and intrinsic space, and a new visual-motor transformation is needed for visually guided reaching or pointing. The strategic calibration is a faster motor modification to adjust to a change in visual-motor mapping. The spatial alignment is a slower process and requires remapping the visual and motor relation. Similar adaptation processes have been reported by Shinn-Cunningham (2001) for the auditory system. Short-term training changes the perceived sound source location and long-term training may activate a new neural pathway to extract spatial information from altered acoustic cues. The motor adaptation in the tragus-deflection condition of this study suggests that the bat applies a strategic calibration to adapt to new spectral cues introduced by the external ear manipulation. The spatial alignment between the auditory and motor mapping may take place after long-term training.

V. CONCLUSIONS

In conclusion, our results suggest that the tragus plays a role in vertical sound localization for prey capture in the free-flying big brown bat, but the bat can quickly adapt to altered acoustic cues for sound localization. Tragus-deflection does not completely disrupt prey capture ability of the echolocating bat, which suggests that other cues can be used to compensate the effect of changing acoustic cues for target localization in the vertical plane. This is consistent with the report by Aytekin *et al.* (2004) that binaural cues are available to the bat for estimates of vertical sound localization. Moreover, in this study we provide evidence that the bat adapts its flight path in response to altered acoustic cues for target localization. A big brown bat with defective external ears is occasionally found in the wild. Whether the defect is congenital or acquired, this study demonstrates that the animal could successfully compensate for altered acoustic cues for prey localization by modifying its motor behavior.

ACKNOWLEDGMENTS

This research was supported by NIMH Grant No. R01MH56366 and NIBIB Grant No. R01EB004750. We thank Amaya Perez, Ann Plantea, Kari Titcher, Tameeka Williams, and Wei Xian for their help with data collection and analyses.

Abeele, S., and Bock, O. (**2001**). "Mechanisms for sensorimotor adaptation to rotated visual input," Exp. Brain Res. **139**, 248–253.

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**(6), 3594–4605.

Batteau, D. W. (**1967**). "The role of the pinna in human localization," Proc. R. Soc. London, Ser. B **168**(1011), 158–180.

Bloom, P. J. (**1977**). "Creating source elevation illusions by spectral manipulation," J. Audio Eng. Soc. **25**(9), 560–565.

Butler, R. A., and Humanski, R. A. (**1992**). "Localization of sound in the vertical plane with and without high-frequency spectral cues," Percept. Psychophys. **51**, 182–186.

Carlile, S., Martin, R., and McAnally, K. (**2005**). "Spectral information in sound localization," Int. Rev. Neurobiol. **70**, 399–434.

Contreras-Vidal, J. L., Bo, J., Boudreau, J. P., and Clark, J. E. (**2005**). "Development of visumotor representations for hand movement in young children," Exp. Brain Res. **162**, 155–164.

Cunningham, H. A. (**1989**). "Aiming error under transformed spatial mappings suggests a structure for visual-motor maps," J. Exp. Psychol. Hum. Percept. Perform. **15**(3), 493–506.

Cunningham, H. A., and Welch, R. B. (**1994**). "Multiple concurrent visual-motor mappings: implications for models of adaptation," J. Exp. Psychol. Hum. Percept. Perform. **20**(5), 987–999.

Firzlaff, U., and Schuller, G. (**2003**). "Spectral directionality of the external ear of the lesser spear-nosed bat, *Phyllostomus discolor*," Hear. Res. **185**, 110–122.

Firzlaff, U., and Schuller, G. (**2004**). "Directionality of hearing in two CF/FM bats, *Pteronotus parnellii* and *Rhinolophus rouxi*," Hear. Res. **197**, 74–86.

Fisher, H. G., and Freedman, S. J. (**1968**). "The role of the pinna in auditory localization," J. Aud Res. **8**, 15–26.

Fitts, P. M. (**1966**). "Cognitive aspects of information processing: III. Set for speed versus accuracy," J. Exp. Psychol. **71**(6), 849–857.

Fuzessery, Z. M. (**1996**). "Monaural and binaural spectral cues created by the external ears of the pallid bat," Hear. Res. **95**, 1–17.

Gardner, M. B. (**1973**). "Some monaural and binaural facets of median plane localization," J. Acoust. Soc. Am. **54**, 1489–1495.

Grinnell, A. D., and Grinnell, V. S. (**1965**). "Neural correlates of vertical localization by echo-locating bats," J. Physiol. (London) **181**, 830–851.

Heffner, R. S., Koay, G., and Heffner, H. E. (**1996**). "Sound localization in chinchillas III: effect of pinna removal," Hear. Res. **99**, 13–21.

Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, A. J. (**1998**). "Relearning sound localization with new ears," Nat. Neurosci. **1**(5), 417–421.

Hope, G. M., and Bhatnagar, K. P. (**1979**). "Electrical response of bat retina to spectral stimulation: Comparison of four microchiropteran species," Experientia **35**, 1189–1191.

Imamizu, H., Uno, Y., and Kawato, M. (**1998**). "Adaptive internal model of intrinsic kinematics involved in learning an aiming task," J. Exp. Psychol. Hum. Percept. Perform. **24**(3), 812–829.

Kagerer, F. A., Contreras-Vidal, J. L., and Stelmach, G. E. (**1997**). "Adaptation to gradual as compared with sudden visuo-motor distortions," Exp. Brain Res. **115**, 557–561.

Kawato, M. (**1999**). "Internal models for motor control and trajectory planning," Curr. Opin. Neurobiol. **9**, 718–727.

Kawato, M., and Wolpert, D. (**1998**). "Internal models for motor control," Novartis Found Symp. 218, 291–307.

King, A. J., Parsons, C. H., and Moore, D. R. (**2000**). "Plasticity in the neural coding of auditory space in the mammalian brain," Proc. Natl. Acad. Sci. U.S.A. **97**(22), 11821–11828.

Knudsen, E. I., Esterly, S. D., and Olsen, J. F. (**1994**). "Adaptive plasticity of the auditory space map in the optic tectum of adult and baby barn owls in response to external ear modification," J. Neurophysiol. **71**, 79–94.

Lawrence, B. D., and Simmons, J. A. (**1982**). "Echolocation in bats: the external ear and perception of the vertical position of targets," Science **218**(4571), 481–483.

Linkenhoker, B. A., and Knudsen, E. I. (**2002**). "Incremental training increases the plasticity of the auditory space map in adult barn owls," Nature (London) **419**(19), 293–296.

Marotta, J. J., Keith, G. P., and Crawford, J. D. (**2005**). "Task-specific sensorimotor adaptation to reversing prisms," J. Neurophysiol. **93**, 1104–1110.

Martin, T. A., Norris, S. A., Greger, B. E., and Thach, W. T. (**2002**). "Dynamic coordination of body parts during prism adaptation," J. Neurophysiol. **88**, 1685–1694.

Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.

Müller, R. (**2004**). "A numerical study of the role of the tragus in the big brown bat," J. Acoust. Soc. Am. **116**(6), 3701–3712.

Müller, R., Lu, H., Zhang, S., and Peremans, H. (**2006**). "A helical biosonar scanning pattern in the Chinese Noctule, *Nyctalus plancyi*," J. Acoust. Soc. Am. **119**(6), 4083–4092.

Oldfield, S. R., and Parker, S. P. A. (**1984**). "Acuity of sound localization: A topography of auditory space. II. Pinnae cues absent," Perception **13**, 601–617.

Oldfield, S. R., and Parker, S. P. A. (**1986**). "Acuity of sound localization: A topography of auditory space. III. Monaural hearing conditions," Perception **15**, 67–81.
Parsons, C. H., Lanyon, R. G., Schnupp, J. W. H., and King, A. J. (**1999**). "Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets," J. Neurophysiol. **82**, 2294–2309.
Paz, R., Nathan, C., Boraud, T., Bergman, H., and Vaadia, E. (**2005**). "Acquisition and generalization of visuomotor transformations by nonhuman primates," Exp. Brain Res. **161**, 209–219.
Redding, G. M., and Wallace, B. (**1994**). "Effects of movement duration and visual feedback on visual and proprioceptive components of prism adaptation," J. Motor Behav. **26**(3), 257–266.
Redding, G. M., and Wallace, B. (**2002**). "Strategic calibration and spatial alignment: a model from prism adaptation," J. Motor Behav. **34**(2), 126–138.
Redding, G. M., Rossetti, Y., and Wallace, B. (**2005**). "Applications of prism adaptation: a tutorial in theory and method," Neurosci. Biobehav Rev. **29**, 431–444.
Seidler, R. D. (**2005**). "Differential transfer processes in incremental visuomotor adaptation," Motor Control **9**, 40–58.
Shinn-Cunningham, B. (**2001**). "Models of plasticity in spatial auditory processing," Audiol. Neuro-Otol. **6**, 187–191.
Stratton, G. (**1896**). "Some preliminary experiments on vision without inversion of the retinal image," Psychol. Rev. **3**, 611–617.
Stratton, G. M. (**1897a**). "Vision without inversion of the retinal image," Psychol. Rev. **4**(4), 341–360.
Stratton, G. M. (**1897b**). "Vision without inversion of the retinal image," Psychol. Rev. **4**(4), 463–481.
Van Beers, R. J., Wolpert, D. M., and Haggard, P. (**2002**). "When feeling is more important than seeing in sensorimotor adaptation," Curr. Biol. **12**, 834–837.
Van Wanrooij, M. M., and Van Opstal, A. J. (**2005**). "Relearning sound localization with a new ear," J. Neurosci. **25**(22), 5413–5424.
Wolpert, D. M., and Kawato, M. (**1998**). "Multiple paired forward and inverse models for motor control," Neural Networks **11**, 1317–1329.
Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (**1995a**). "Are arm trajectories planned in kinematic or dynamic coordinates? an adaptation study," Exp. Brain Res. **103**, 460–470.
Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (**1995b**). "An internal model for sensorimotor integration," Science **269**(5232), 1880–1882.
Wotton, J. M., and Jenison, R. L. (**1997**). "The combination of echolocation emission and ear reception enhances directional spectral cues of the big brown bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **101**(3), 1723–1733.
Wotton, J. M., and Simmons, J. A. (**2000**). "Spectral cues and perception of the vertical position of targets by the big brown bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **107**(2), 1034–1041.
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**(3), 1423–1445.
Yoshimura, H. (**2002**). "Re-acquisition of upright vision while wearing visually left-right reversing goggles," Jpn. Psychol. Res. **44**(4), 228–233.
Zar, J. H. (**1996**). *Biostatistical Anaylsis*, 3rd ed. (Prentice-Hall, Upper Saddle River, NJ).

Effects of carrier pulse rate and stimulation site on modulation detection by subjects with cochlear implants[a)]

Bryan E. Pfingst[b)]
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-0506

Li Xu
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701 and Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-0506

Catherine S. Thompson
Kresge Hearing Research Institute, Department of Otolaryngology, University of Michigan, Ann Arbor, Michigan 48109-0506

(Received 18 May 2006; revised 12 January 2007; accepted 17 January 2007)

Most modern cochlear-implant speech processors convey speech-envelope information using amplitude-modulated pulse trains. The use of higher-rate carrier pulse trains allows more envelope detail in the signal. However, neural response properties could limit the efficacy of high-rate carriers. This study examined effects of carrier rate and stimulation site, on psychophysical modulation detection thresholds (MDTs). Both of these variables could affect the neural representation of the carrier and thus affect perception of the modulation. Twelve human subjects with cochlear implants were tested. Phase duration of symmetric biphasic pulses was modulated sinusoidally at 40 Hz. MDTs were determined for monopolar stimulation at two carrier rates [250 and 4000 pulses/s (pps)], three stimulation sites (basal, middle, and apical), and five stimulus levels (10%, 30%, 50%, 70%, and 90% of the dynamic range). MDTs were lower for 250 pps carriers than for 4000 pps carriers in 71% of the 180 cases studied. Effects of carrier rate were greatest at the apical stimulation site and effects of stimulation site on MDTs depended on carrier rate. The data suggest a distinct disadvantage to using carrier pulse rates as high as 4000 pps. Stimulation site should be considered in evaluating modulation detection ability. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2537501]

PACS number(s): 43.66.Ts, 43.66.Mk, 43.66.Cb [RAL] Pages: 2236–2246

I. INTRODUCTION

Most modern cochlear-implant speech processors function by filtering the auditory stimulus into discrete bands of frequencies, extracting the envelopes of the filtered signals, and using these envelopes to amplitude modulate trains of interleaved pulses. The use of amplitude-modulated pulse trains for cochlear prostheses was developed to enable stimulation with interleaved pulse trains (Wilson *et al.*, 1991), which are needed to avoid problems with current interactions that occur with multichannel analog stimulation. The amplitude-modulation strategy is supported by the fact that subjects can achieve high levels of speech recognition using temporal-envelope cues, even with very limited spectral processing (Van Tasell *et al.*, 1987; Shannon *et al.*, 1995). Recent studies have demonstrated that modulation detection thresholds (MDTs), i.e., the just-detectable magnitudes of charge modulation, are strongly correlated with a subject's speech recognition with cochlear and auditory-brainstem implants (Fu, 2002; Colletti and Shannon, 2005).

In charge-modulated pulse trains, the carrier pulse rate limits the temporal detail with which the modulation waveform is sampled. Higher carrier pulse rates allow more accurate representation of analog modulation waveform. For this reason, it has been assumed that higher carrier pulse rates will result in better information transmission. However, in the case of a cochlear implant, where temporal information must be transmitted by the temporal response patterns of neurons, neural response features such as adaptation and refractory properties must be considered. In studies of the representation of cochlear-implant stimulation in guinea pig auditory cortex, Middlebrooks (2005) found that the representation of the modulation waveform (20, 40, or 60 Hz sinusoids) in the activity of cortical neurons was better when low-rate (254 pps) carriers were used than when high-rate (4069 pps) carriers were used. This result suggests the hypothesis that lower-rate carriers will result in better modulation sensitivity in human subjects with cochlear implants. We tested that hypothesis in the current study by measuring MDTs in human subjects with Nucleus® prostheses using 250 and 4000 pulses/s (pps) carriers.

[a)]Initial reports of these data were presented previously [Pfingst *et al.*, 2005 Conference on Implantable Auditory Prostheses, Asilomer Conference Grounds, Pacific rove, CA; Xu and Pfingst, The 25th Politzer Society Meeting, Seoul, Korea (2005)].

[b)]Author to whom correspondence should be addressed. Electronic mail: bpfingst@umich.edu

 0001-4966/2007/121(4)/2236/11/$23.00

The 250 and 4000 pps carrier rates were chosen to produce very different patterns of neural activity in the auditory nerve and central auditory pathways. At 250 pps, the auditory nerve fibers are capable of entrainment to the pulse train, giving a neural discharge to every pulse, with the same temporal pattern in most or all of the activated neurons (Wilson *et al.*, 1997). At pulse rates greater than 1000 pps, auditory nerve fibers cannot discharge to every pulse due to refractory properties. Because of across-fiber variation in time-to-recovery from refractory states, the across-fiber pattern of response to a high-rate pulse train will be more stochastic, which is a more natural across-fiber pattern. It has been suggested that this more stochastic pattern might result in better encoding of temporal information by cochlear implants (Rubinstein and Hong, 2003). However, a recent comparison of low and high carrier rates did not support this hypothesis (Galvin and Fu, 2005).

For the current study, we chose a 40 Hz modulation frequency. Most current cochlear prosthesis speech processors provide temporal envelope cues low-pass filtered at about 200–400 Hz (Wilson, 2004). For English phoneme recognition, the most important envelope information is below 16–20 Hz (Drullman *et al.*, 1994a, b; Fu and Shannon, 2000; Xu *et al.*, 2005). Higher-frequency periodicity cues (50–500 Hz) have been found to benefit lexical-tone recognition (Fu *et al.*, 1998; Xu *et al.*, 2002) and voice gender recognition (Fu *et al.*, 2004).

We hypothesized that effects of carrier rate on modulation detection might depend on stimulation site along the basal to apical extent of the electrode array and that stimulation site might interact with carrier rate. There are several reasons for this hypothesis. First, recent evidence suggests that the timing properties of spiral-ganglion neurons vary systematically along apical-basal dimension of the cochlea, with more apical neurons showing longer-latency, more slowly adapting responses than basal neurons (Adamson *et al.*, 2002; Liu and Davis, 2006). Thus, we might expect that the apical end of the cochlear-implant electrode array would stimulate fibers that are better equipped to process signals with lower carrier rates. In addition, pathology of the deaf, implanted cochlea could affect the temporal response properties of the neurons. This pathology could be somewhat systematic because the base of the cochlea is often more vulnerable to pathology, but there can also be unsystematic, subject specific, variation in pathology along the cochlear length (Hinojosa and Lindsay, 1980; Nadol, 1997). If variation in modulation-detection ability depends in part on peripheral physiology, not just cognitive processes, we would expect to find variation across stimulation sites in modulation detection thresholds and possibly interactions between stimulation site and temporal properties of the stimulus. We tested the hypothesis that effects of carrier rate vary across stimulation sites by assessing the effects of carrier rate at three locations in each subject: Basal, middle, and apical regions of the electrode array.

Modulation detection and modulation-frequency discrimination improve as a function of level throughout the dynamic range of electrical hearing (Pfingst *et al.*, 1994; Fu, 2002). Therefore, the level of the stimulus must be taken into account in comparing modulation detection across various stimulus conditions. Fu (2002) demonstrated that the mean MDTs averaged across five stimulus levels spaced throughout the dynamic range of electrical hearing correlated more highly with speech recognition than MDTs at only one level. Therefore, in the current study we assessed MDTs at five levels within the dynamic range of hearing for each condition.

II. METHOD

A. Subjects

Twelve postlingually deafened adults ranging in age from 37 to 71 years participated in this study. All had at least one year of experience with their cochlear implants. Five of the subjects had Nucleus 24M (straight array) implants and seven had Nucleus 24R(CS) (Contour) implants. Five of the subjects had participated in previous pilot studies involving modulation detection tasks. Details of the characteristics of each subject are summarized in Table I.

All subjects were paid an hourly rate for participation in the study plus travel expenses. The use of human subjects in this research was reviewed and approved by the University of Michigan Medical School Institutional Review Board.

B. Equipment and software

To assure uniformity in the external hardware, all listeners were tested with a laboratory-owned Sprint® processor (Cochlear Corporation) during the experiment. Communication with the processor was accomplished using an IF5 ISA card and a Processor Control Interface (PCI) from Cochlear Corporation. Sequences of frames were created and sent to the processor using the Nucleus Implant Communicator® (v. 3.7) software libraries. The software for the experiment was written locally and run on a personal computer.

C. Research design

The primary independent variables for this study were carrier rate and stimulation site. Two carrier rates were tested: 250 and 4000 pps. Three stimulation sites were tested in each subject. These sites were located in the apical (electrode 18), middle (electrode 11), or basal (electrode 4 or 6)[1] regions of the implant. Monopolar stimulation (MP 1+2) was used in all cases. The stimulation was between one scala tympani electrode and two external electrodes in parallel: (1) The plate electrode on the implanted receiver-stimulator and (2) the ball electrode implanted in the temporalis muscle. Psychophysical detection thresholds (T levels) and maximum comfortable loudness levels (C levels) were measured for each of the six conditions (2 pulse rates $\times$ 3 sites) to determine the dynamic range for testing modulation detection. Then MDTs were measured at five levels within the dynamic range for each of these six conditions. Details of the T level, C level, and MDT measurement procedures are given in the following.

Modern cochlear implants often use a combination of current-amplitude and phase-duration modulation to control the total charge per phase delivered. We used current ampli-

TABLE I. Subject characteristics.

Subject	Age (years)	Sex	Implant type	Duration of profound deafness in implanted ear prior to implantation (years)	Duration of implant use (years)	Participated in pilot study
S1	49	M	24R(CS)	<1	4	Yes
S2	58	F	24M	1	7	No
S3	70	F	24R(CS)	35	3	Yes
S4	56	F	24M	25	8	Yes
S5	66	M	24R(CS)	2	2	Yes
S6	37	F	24R(CS)	<1	1	Yes
S7	53	F	24M	12	16[a]	No
S8	66	M	24R(CS)	29	2	No
S9	45	F	24M	11	7	No
S10	64	M	24R(CS)	<1	5	No
S11	71	F	24R(CS)	30	5	No
S12	67	F	24M	3	4	No

[a]S7 was explanted and reimplanted after 9.5 years with the first implant. Duration of use of the second implant at the time of this experiment was 6.5 years.

tude to determine T and C levels and to set the stimulation level within the dynamic range. For determining MDTs, we used phase-duration modulation because the prosthesis design allows finer control of charge per phase using phase duration than is possible with current amplitude. Pulse phase duration was sinusoidally modulated at 40 Hz around a mean pulse duration of 50 μs per phase. Symmetric biphasic pulses were used with an 8 μs interphase gap. The gap was held constant while the durations of the positive and negative phases were modulated equally to maintain charge balance. The modulation index (m) was defined as

$$m = (\mathrm{PD}_{\max} - \mathrm{PD}_{\min})/(\mathrm{PD}_{\max} + \mathrm{PD}_{\min}),$$

where $\mathrm{PD}_{\max}$ and $\mathrm{PD}_{\min}$ are the maximum and minimum phase durations, respectively. We report modulation values in percent modulation ($m \times 100$) or in dB re 100% modulation ($20 \log m$). All stimuli were 600 ms in duration.

D. T and C levels and dynamic ranges

For each of the six conditions (2 pulse rates $\times$ 3 sites), T levels and C levels were obtained at 0% modulation and at 50% modulation (i.e., −6.02 dB re 100% modulation). T levels and C levels were obtained using the method of adjustment in which the subjects adjusted the level of the biphasic pulses on individual electrodes using the keyboard arrow keys and our custom software program. For T levels, the listeners set the level as just barely audible. For C levels, the subjects adjusted the level to the maximum comfortable loudness level at which the subjects felt they could listen for a long period of time without discomfort. T and C levels were measured for the unmodulated signal and for the 50% modulated signals for the six conditions in random order and the set was then repeated twice using a different randomization each time for a total of three measurements for each condition. The means of the sets of three measurements were used as the estimates of the T and C levels.

The variation of the repeated T and C level measurements was computed using the following procedures. First, for each set of the three measurements, the values were normalized to the median of the three. All normalized estimates pooled across pulse rates, modulation depths, sites, and subjects showed a normal distribution. The standard deviation of the normalized estimates was then obtained as a measure of the variation of the repeated T and C level measurements.

Across the 12 subjects times three sites tested in this experiment, the T levels for the unmodulated pulse train were an average of 1.31 dB higher than those for the 50% phase-duration modulated pulse train. C levels for the unmodulated pulse train were an average of 1.06 dB higher than those for the 50% modulated pulse train. Note that as modulation depths became shallower, T levels and C levels for modulated and unmodulated signals became indistinguishable, as detailed in Sec. III. To assure that the subject could hear all stimuli and that none would be too loud during the tracking procedure for determining MDTs, dynamic range for each condition was conservatively defined as the lower of the two C levels (modulated or unmodulated) minus the higher of the two T levels. MDTs were obtained at 10%, 30%, 50%, 70%, and 90% of these dynamic ranges, as detailed in the following.

E. Modulation detection thresholds

MDTs were obtained using a two-interval forced-choice paradigm with flanking cues. On each trial, subjects were presented with four sequential observation intervals marked by buttons on the computer screen which were highlighted in sequence. An electrical stimulus to the implant (a pulse train) was presented during each interval. The first and fourth in-

tervals contained identical unmodulated pulse trains which served as flanking cues. One of the other intervals (interval 2 or interval 3), chosen at random on each trial, also contained this unmodulated signal. The modulated pulse train occurred in the remaining interval and the subject was instructed to choose the interval that sounded different from the other three. Stimulus duration was 600 ms with a 600 ms interval between stimuli.

A two-down one-up adaptive tracking procedure (Levitt, 1971) was used, starting with a modulation depth of 50% and decreasing in steps of 6 dB to the first reversal, 2 dB for the next two reversals, and 1 dB for the next 10 reversals. The MDT was defined as the mean of the levels at the last 8 reversal points.

MDTs were measured in each subject for a total of 30 conditions (2 carrier rates $\times$ 3 stimulation sites $\times$ 5 levels). MDTs for these conditions were measured in random order and then the complete set of tests was repeated twice using a different randomization each time for a total of three estimates per condition. If trial-to-trial variability in any condition seemed high, additional MDTs for that condition were obtained. These data were later screened for outliers as described in the following.

After collecting the estimates of MDTs for each condition, the data were examined to determine if there were any outliers. First, for each set of the three estimates the values were normalized to the median of the three. All normalized estimates across carrier rates, levels, sites, and subjects showed a normal distribution. The standard deviation (s.d.) of the normalized estimates was obtained, which was equal to 2.37 dB. Then, the outliers were defined as estimates that were more than 7.1 dB (i.e., $3 \times$ s.d.) of the median of each set of the three estimates. Once an outlier was identified, which occurred only rarely (approximately 1.9% of all measurements), we excluded the outlier, replaced it with a fourth collected MDT, and took the mean of the resulting three values.

III. RESULTS

A. Dynamic ranges

As noted in Sec. I, modulation detection improves as a function of stimulus level throughout much or all of the dynamic range of electrical hearing. An important consideration in the current experiment is that the dynamic range for stimulation at 4000 pps was typically larger than that at 250 pps. Figure 1 (upper panel) shows the T and C levels of all 12 subjects. Stimulation at 4000 pps (open symbols) produced lower C levels and lower T levels than did stimulation at 250 pps (closed symbols). However, the effects of pulse rate were usually greater for T levels than for C levels on the decibel (dB) scale. That is, when the pulse rate was increased from 250 to 4000 pps, T levels decreased more than C levels. Thus, the dynamic range in dB of current for stimulation at 4000 pps was usually larger than that at 250 pps. In these data, the dynamic range was larger for the 4000 pps stimulus in all but one (the basal site of S8) of the 36 cases (12 subjects $\times$ 3 stimulation sites) studied (Fig. 1, lower panel). To compensate for these differences in dynamic range for the two carriers as well as large across-subject differences in dynamic range, we compared MDT versus level functions using percent of dynamic range as the measure of level. This scale was based on dynamic ranges in dB of current. This scale was chosen because it corresponds roughly to the scale used in delivering current in the subjects' normal everyday speech processors.

B. MDT-versus-level functions

MDTs for the 12 subjects are shown in Fig. 2. For each subject, MDT-versus-level functions for the two carrier rates at each of three stimulation sites are shown. There is considerable variability across subjects and conditions in the details of the MDT-versus-level functions but there are some clear and consistent trends. MDTs improve as a function of level in almost all cases. In most cases, the MDTs for the 250 pps carrier are better than those for 4000 pps carrier when compared at equal levels in percent of dynamic range. Differences between MDTs obtained for the 4000 pps carrier and the 250 pps carrier averaged over all three stimulation sites and all five levels are shown for each subject in the lower right corner of each panel. These values are positive for 9 of the 12 subjects. The positive values indicate better average modulation detection when the 250 pps carrier was used. Note that the MDT data for the three subjects with negative values (4000 pps carrier better) were generally poor compared to those for the other subjects in the study. They were particularly poor, relative to those for other subjects, at higher levels for both carriers and for all levels for the 250 pps carrier. The stimulation site itself did not produce a consistent effect on the MDTs. However, if the effects of stimulation site were considered together with level, as detailed in Sec. III D, we found some interesting interactions of the two variables affecting the MDTs.

A three-way ANOVA was performed to compare the means of the MDTs produced by the three variables, i.e., level, carrier rate, and stimulation site. The results indicated that both level and carrier rate produced significantly different mean MDTs (level: $F=63.4$, d.f.$=4$, $p<0.000$; carrier rate: $F=48.5$, d.f.$=1$, $p<0.000$) whereas stimulation site did not yield significantly different mean MDTs ($F=0.44$, d.f. $=2$, $p=0.644$).

C. Effects of carrier rate

From the data in Fig. 2, effects of carrier rate on MDTs (MDT for the 4000 pps carrier minus MDT for the 250 pps carrier) were computed for the 180 cases (12 subjects $\times$ 3 sites per subject $\times$ 5 levels per site). Of these 180 cases, MDTs were better for the 250 pps carrier in 70.8% of the cases. The differences between MDTs for the two carriers ranged from +25.3 dB (MDT for the 250 pps carrier better) to −12.8 dB (MDT for the 4000 pps carrier better). The mean difference was 5.5 dB. The distribution of these differences is shown in the left panel of Fig. 3.

We also computed mean MDTs, averaged across the five stimulus levels, for the 36 cases (12 subjects $\times$ 3 sites). Mean MDTs averaged across multiple stimulation levels have been shown by Fu (2002) to be highly correlated across subjects

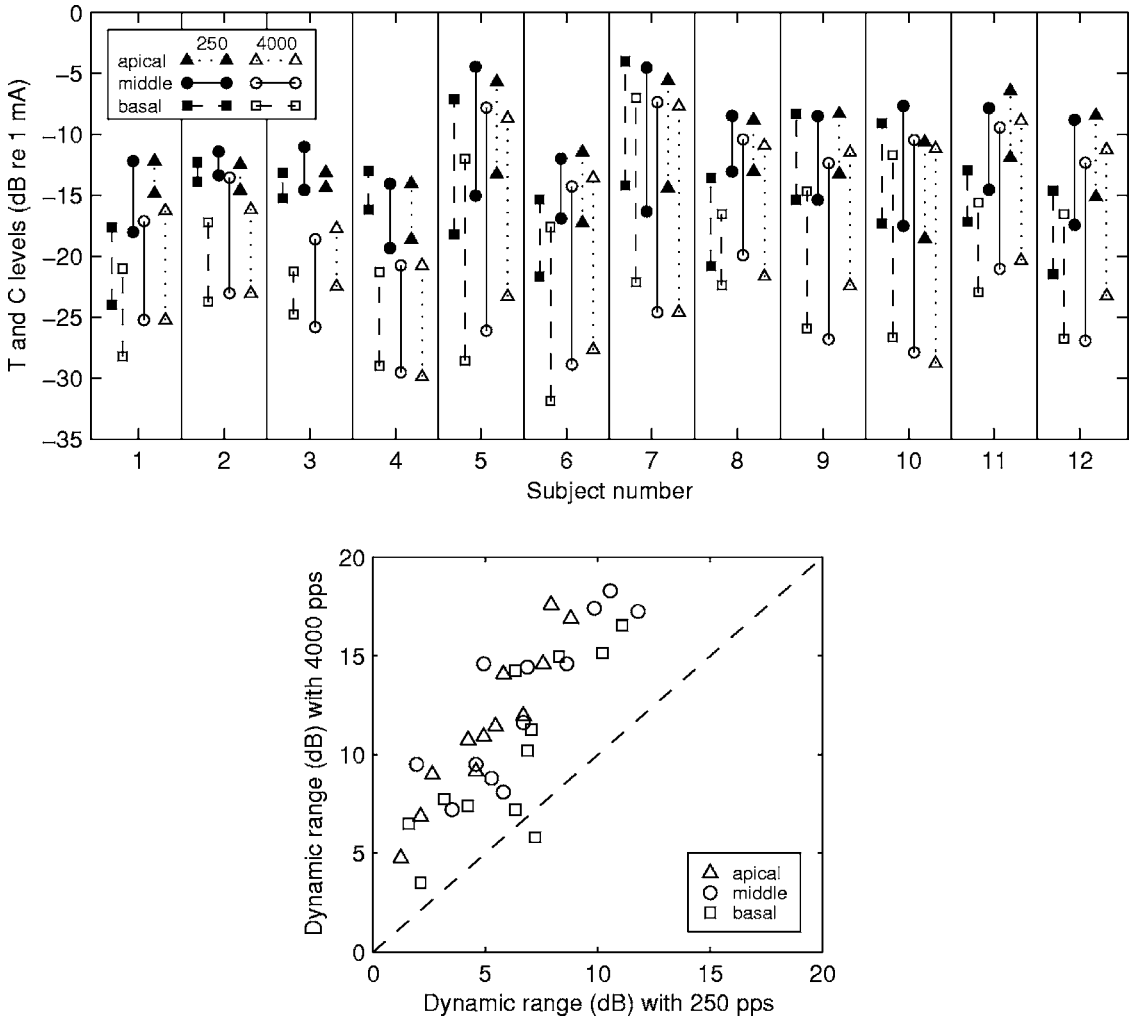


FIG. 1. T and C levels and dynamic ranges. Upper panel: T and C levels obtained from the 12 subjects. In each pair of points connected by a vertical line, the upper point is the C level and the lower point is the T level. Means for three repeated measurements are shown. The standard deviations, calculated as described in Sec. II, were 1.39 and 0.86 dB for repeated T and C level measurements, respectively. Lower panel: Comparison of the dynamic ranges obtained with 250 pps carriers to those obtained with 4000 pps carriers. Each point indicates the dynamic ranges for the 4000 pps carrier (ordinate) and the 250 pps carrier (abscissa) for one stimulation site in one subject, calculated from the T and C levels shown in the upper panel. Points above the diagonal indicate that dynamic ranges were larger for the 4000 pps carrier than for the 250 pps carrier.

with consonant recognition and well correlated with vowel recognition. Mean MDTs for the 250 pps carrier were better than those for the 4000 pps carrier in 28 of the 36 cases tested. The distribution of these differences is shown in the right panel of Fig. 3. The mean of this distribution was 5.5 dB.

Group-mean MDT-versus-level functions averaged across all 12 subjects are shown in Fig. 4 for each of the two carrier rates at each of the three stimulation sites. In these group-mean data, MDTs for the 250 pps carrier (closed symbols) were lower (better) than those for the 4000 pps carrier (open symbols) in the corresponding conditions across three tested stimulation sites and across all five tested levels with the only exception being the middle site at 10% of the dynamic range.

D. Effects of stimulation site

On average, the best MDTs for the 250 pps carrier were at the apical stimulation site and the best MDTs for the 4000 pps carrier were at the basal site (Fig. 4). The poorest MDTs were for the 4000 pps carrier at the apical site. Thus, the largest difference between MDTs for the 250 pps carrier and the 4000 pps carrier was at the apical stimulation site. The data for the differences in MDTs between the two carriers were organized in a matrix with stimulus level being one factor and stimulation site being another factor. A two-way ANOVA of the data revealed a statistically significant effect of stimulation site on differences between MDTs for the two carriers ($F=7.7$, d.f.$=2$, $p=0.0007$) but no statistically significant effect of stimulus level on these differences ($F=2.4$, d.f.$=4$, $p=0.056$). No interaction between levels and sites was found to be statistically significant ($F=0.7$, d.f.$=8$, $p=0.675$).

Interaction between stimulation site and carrier rate in determining MDTs is also illustrated in Fig. 5, which compares MDTs for the three possible pairings of the three stimulation sites (apical, middle, and basal) at the two carrier rates (250 and 4000 pps). The left column of Fig. 5 compares MDTs at the apical and basal sites for 60 cases (12

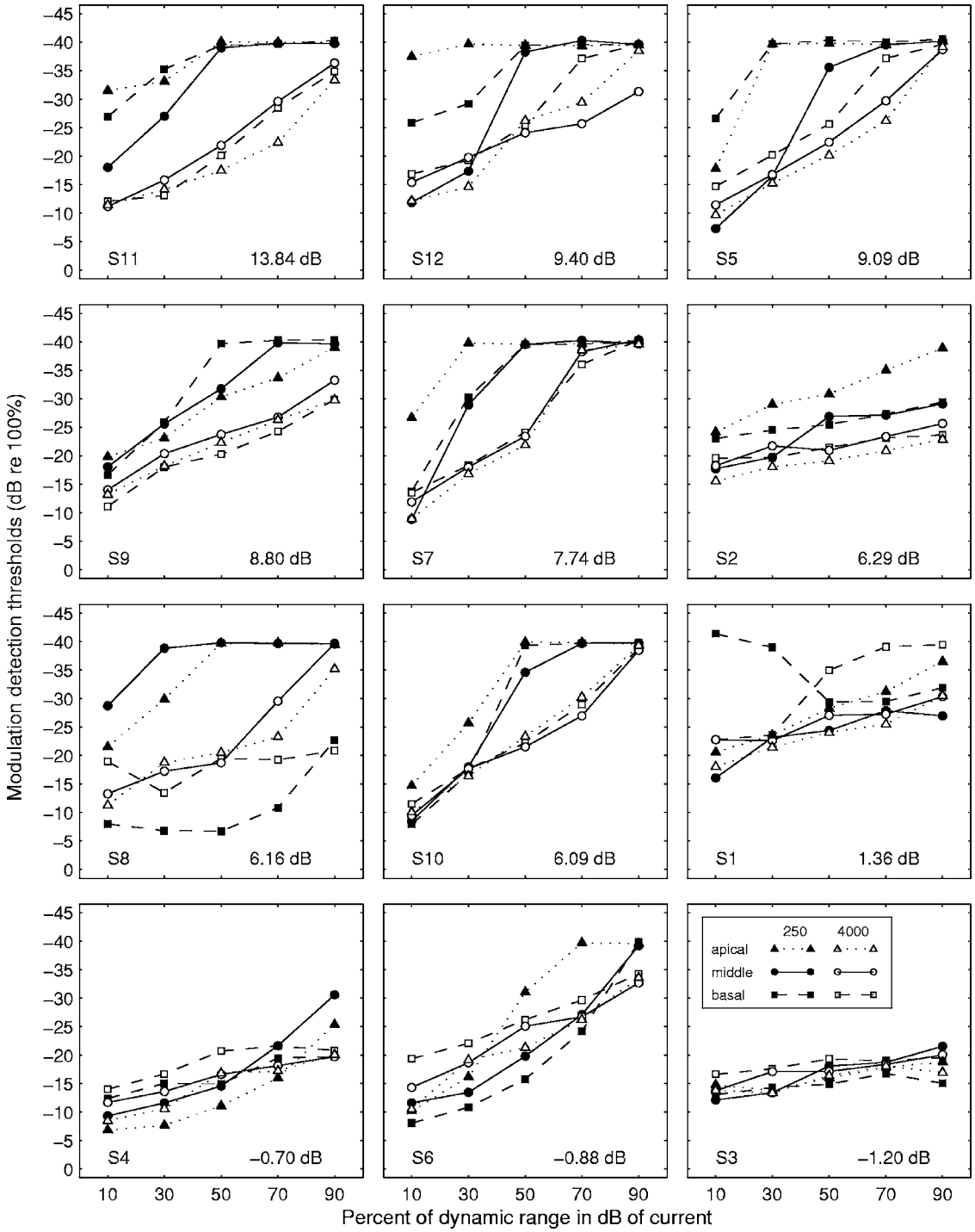


FIG. 2. Modulation detection threshold vs level functions for the 12 subjects. Each panel shows MDT-versus-level functions for a single subject for three stimulation sites (basal, middle, and apical) at two carrier rates (250 and 4000 pps). The legend is shown in the lower right panel. Subject numbers are indicated in the lower left corner of each panel. For each subject, the mean effect of carrier rate, calculated as the mean difference in MDTs (MDT at 4000 pps minus MDT at 250 pps) for all 15 conditions (3 sites × 5 levels), is shown in the lower right corner of the panel. The panels are arranged in order from highest mean difference value (upper left panel) to the lowest mean difference value (lower right panel). The abscissa gives the stimulus level in percent of dynamic range where dynamic range is in dB of current.

subjects × 5 stimulation levels) (gray circles) and the mean MDTs across the 5 levels for each of the 12 subjects (closed squares). At 250 pps (Fig. 5, upper left panel), MDTs were smaller (better) at the apical site in the majority of the cases: 63.3% of the 60 cases fell above the diagonal and the average apical-basal difference was −3.00 dB. At 4000 pps (Fig. 5, lower left panel), MDTs were better at the basal site in the majority of the cases: 76.7% of the 60 cases were below the diagonal and the average apical-basal difference was +2.39 dB. The differences between the stimulation sites in the means across the 5 levels (closed squares) in the upper panel were compared with those in the lower panel (leftmost column of Fig. 5). It was found that the apical-basal differences in the two panels were statistically significant (paired t

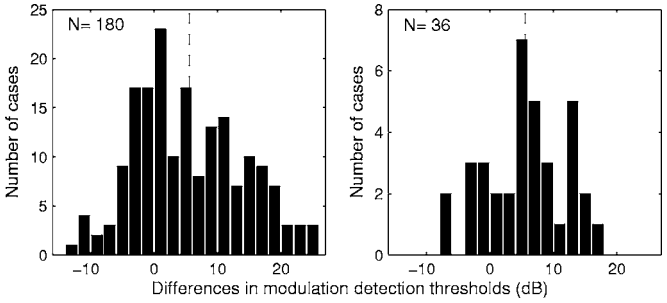


FIG. 3. Distribution of the differences between MDTs obtained with the two carriers (MDT for the 4000 pps carrier minus MDT for the 250 pps carrier). In the left panel, the differences were calculated for 180 cases (12 subjects × 3 sites/subject × 5 levels/site). In the right panel, the differences were calculated using mean MDTs (averaged across the five tested levels) for 36 cases (12 subjects × 3 sites/subject). Positive values indicate that MDTs were lower (better) for the 250 pps carrier. The vertical dashed lines indicate the means of the distributions.

test, $t=3.075$, $p=0.0106$). Thus, the effect of stimulation site (apical versus basal) on MDTs depended on the carrier rate.

Comparison of the apical and middle sites (middle column in Fig. 5) showed results similar to the comparison of apical and basal sites. Statistically significant differences were found between the apical-middle differences in the two panels in the middle column of Fig. 5 (paired t test, $t=3.555$, $p=0.0045$). Comparison of the middle and basal sites (right column in Fig. 5) showed little consistency in effects of site on MDTs. The middle-basal differences were not statistically significant (paired t test, $t=0.382$, $p=0.7095$).

E. Relation between MDTs and dynamic ranges

As noted earlier, the dynamic ranges for the 250 pps carrier were smaller than those for the 4000 pps carrier in all but one of the 36 cases studied (Fig. 1), and the MDTs for the 250 pps carrier were smaller (better) than those for the 4000 pps carrier in 28 of these 36 cases (Fig. 3, right panel). Thus, under these circumstances, it seems that small dynamic ranges were associated with better MDTs. To determine if small dynamic ranges were associated with better MDTs as a general rule, we compared mean MDTs to dynamic ranges across subjects, utilizing the commonly observed variation in dynamic ranges across implanted subjects. We make these comparisons in the 12 subjects for each of the three stimulation sites and two carrier rates (Fig. 6). These data do not support the hypothesis that smaller dynamic ranges are associated with better modulation detection under these circumstances. In fact, across subjects, there is a trend in the opposite direction. As shown in Fig. 6, regression lines for these data had a positive slope in all cases, i.e., larger dynamic ranges tended to be associated across subjects with better MDTs. However, the correlations were statistically significant at the $p<0.05$ level in only one case (the apical site for the 4000 pps carrier).

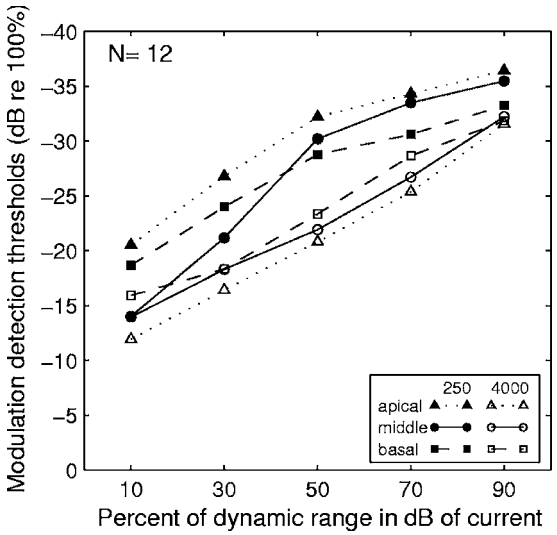


FIG. 4. Mean MDT-vs-level functions for two carrier rates and three stimulation sites. Means were averaged across the 12 subjects. The abscissa gives the stimulus level in percent of dynamic range where dynamic range is in dB of current.

F. Effects of level

From the data in Figs. 2 and 4, it is evident that the shapes of the MDT-versus-level functions are often different for the two carriers. On average, MDTs (in dB) obtained with the 250 pps carrier improved more rapidly as a function of level (in percent of dB dynamic range) than those for the 4000 pps carrier. Given this consideration, it is not obvious precisely what relative levels are appropriate for comparison of MDTs across conditions. A practical approach to this issue is to determine how much one would need to change the presentation level in one condition to obtain the same MDT as that obtained at a given level in another condition. For example, for subject S11 at the middle stimulation site (the solid lines in the upper left panel of Fig. 2), the MDT at 30% of the dynamic range with a 250 pps carrier was −27 dB re. 100% modulation. To achieve this MDT with a 4000 pps carrier at the same stimulation site, the stimulus level would need to be raised to about 60% of the dynamic range, i.e., an increase in level of about 30% of the dynamic range. In the group-average data in Fig. 4, the level for stimulation with the 4000 pps carrier needed to achieve similar MDTs to those obtained with the 250 pps carrier at 30% of the dynamic range would be about 22%, 16%, and 45% of the dynamic range higher for the basal, middle, and apical sites, respectively. Thus, minor errors in estimating the appropriate levels for comparison would not be sufficient to account for the differences in MDTs observed between data for the two carriers.

G. Loudness cues

Amplitude modulation of pulse trains results in pitch-like perceptions and other perceptual cues that may be useful to subjects in recognizing speech signals. As noted in Sec. II, phase-duration modulation of pulse trains can produce a slight lowering of T levels and C levels, suggesting the possibility that loudness cues can contribute to modulation detection. In order to estimate the possible contribution of the loudness cues to modulation detection, we measured the difference in T levels and C levels between unmodulated sig-

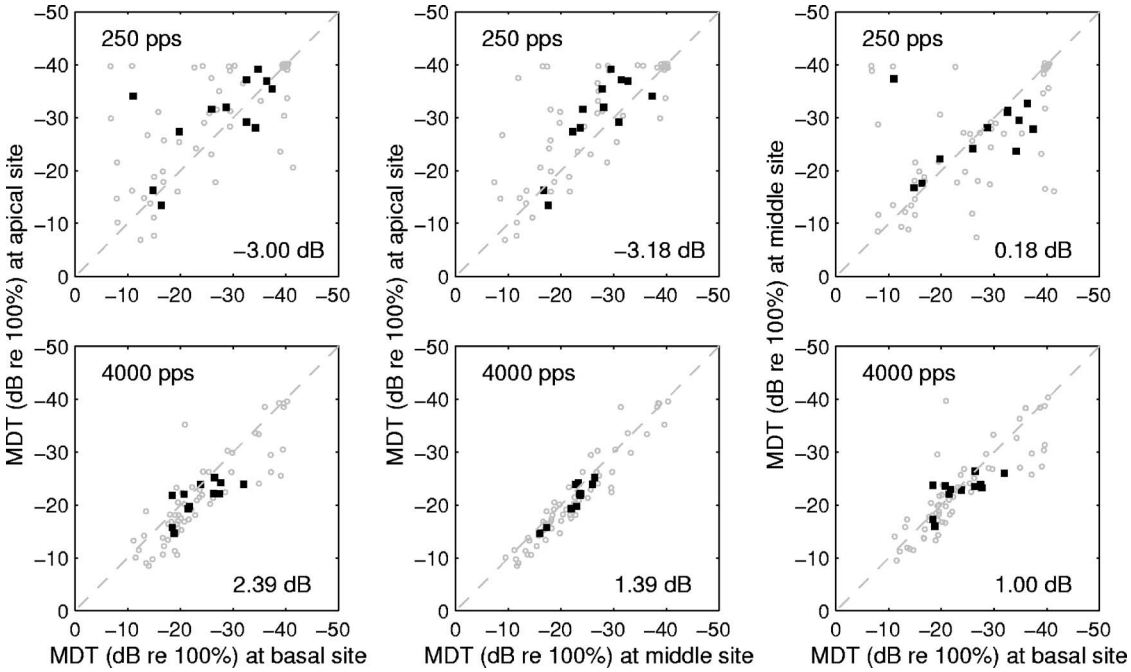


FIG. 5. Comparisons of MDTs vs stimulation site. MDTs for the 60 cases (12 subjects × 5 stimulation levels per subject) are plotted using gray circles and the means across the 5 levels for the 12 subject are plotted using closed squares. The top and bottom rows represent the 250 and the 4000 pps carriers, respectively. Comparisons are shown for the three possible pairings of the three stimulation sites: Apical vs basal in the left column, apical vs middle in the center column, and middle vs basal in the right column. Mean differences in MDTs between the pairs of sites are shown in the lower-right corner of each panel.

nals and modulated signals at various modulation depths (0%–50%) using six of the subjects from this study (S1, S3, S5, S6, S10, and S12). Results of these measurements are shown in Fig. 7. The differences in T and C levels between the unmodulated signal and the 50% modulated signal ranged between 1.25 and 1.12 dB, similar to the mean differences for all 12 subjects reported in Sec. II. However, for modulation depths of 25% or less, these mean differences between the unmodulated and modulated signals were less than 0.3 dB. These differences are well below the range of variation in repeated T and C level measurements for the unmodulated signals and the modulated signals, so if any loudness differences exist at these modulation depths they are below the resolution of our measurement procedures.

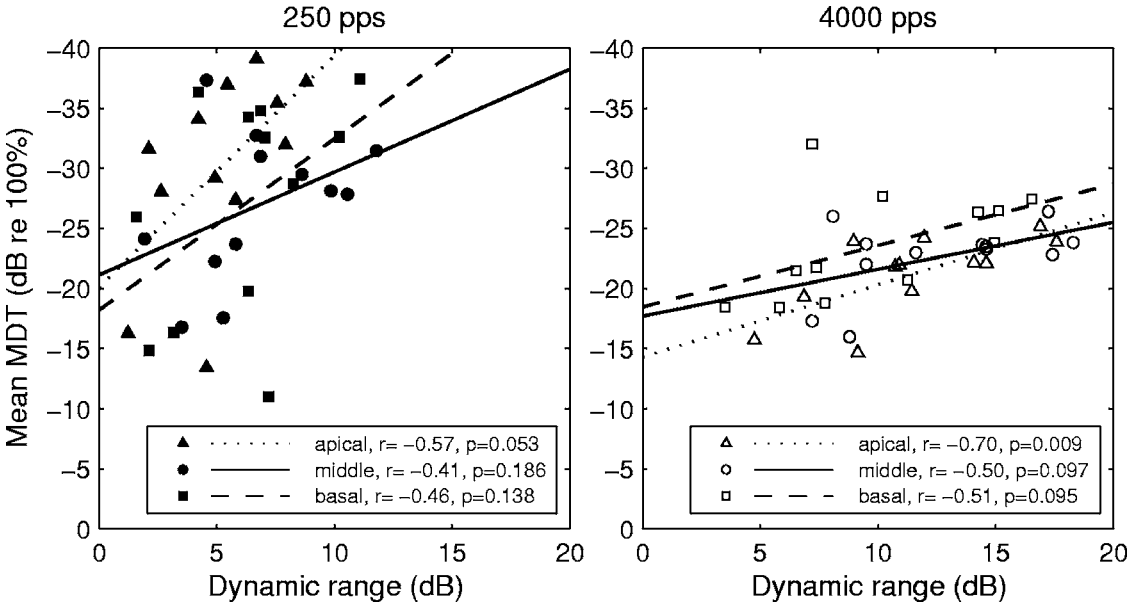


FIG. 6. Relationship between mean MDTs and dynamic ranges for the three stimulation sites (apical, middle, and basal). The left and right panels represent data for the 250 and 4000 pps carriers, respectively. The dotted, solid, and dashed lines represent the least-square fits of the data for the apical, middle, and basal stimulation sites, respectively. The legend shows the correlation coefficients and the statistical significance of the test results.

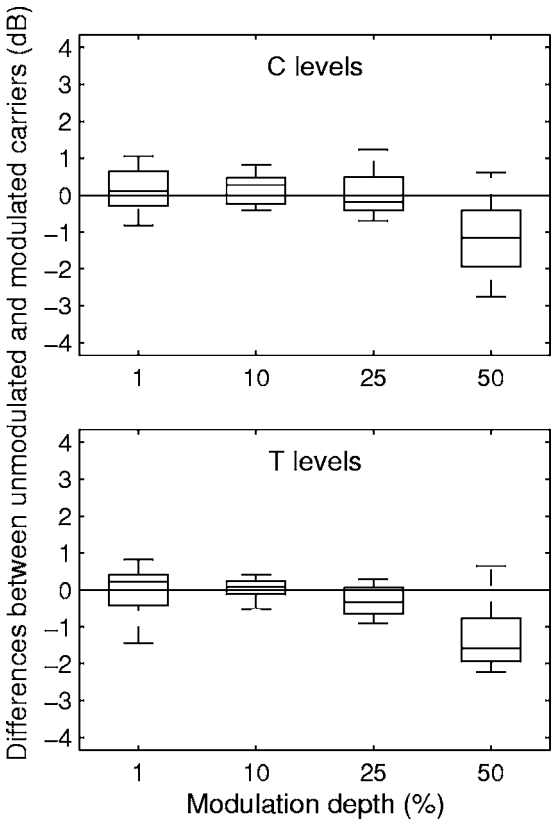


FIG. 7. Differences of T and C levels between modulated and unmodulated carriers. The data were pooled across sites and carrier rates since no systematic differences were found in the data among different stimulation sites or between the low and high carrier rates. The boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data.

IV. DISCUSSION

A. Summary of results

The major findings in this study were (1) that MDTs were usually better when a low rate (250 pps) carrier was used than when a high rate (4000 pps) carrier was used; (2) that the effects of carrier rate on MDTs are affected by the apical to basal location of the electrical-stimulation site, and (3) the effects of stimulation site on MDTs depend on carrier rate. We also confirmed, as expected, that MDTs improved as a function of stimulus level and found that the shapes and slopes of the MDT-versus-level functions varied as a function of carrier rate in most subjects. Finally, we found that the relationship between dynamic range and MDTs depended on the variable (carrier rate or subject) associated with differences in the dynamic range.

B. Effects of carrier rate

The increase in MDTs (reduced sensitivity) resulting from the higher carrier rate averaged 5.5 dB (almost a factor of 2 in percent of modulation) and ranged up to 25.3 dB (a factor of about 18 in percent of modulation) (Fig. 3). The effects of carrier rate were not as large as the effects of level, which averaged 18.0 dB over levels from 10% to 90% of the dynamic range and ranged up to 32.8 dB. Nevertheless, as noted in Sec. III F, very large increases in level of the 4000 pps carrier signal would be needed in many cases to achieve MDTs that were equivalent to those achieved with the 250 pps carrier.

There was considerable variability in effects of pulse rate from subject to subject. A few subjects showed little or no effect of carrier rate or showed small effects in the opposite direction to that observed for the majority. In general, the subjects with small effects of carrier rate had poor MDTs for both carriers. Similar across-subject variability in effects of pulse rate on psychophysical performance has been observed in a study of intensity discrimination in the context of interleaved multichannel stimulation (Drennan and Pfingst, 2006). This across-subject variability in effects of pulse rate on basic psychophysical functions suggests a basis for across-subject variability in effects of pulse rate on speech recognition, which has been observed in several studies (Vandali *et al.*, 2000; Holden *et al.*, 2002).

Some of our subjects showed the largest effects of carrier rate at the intermediate levels within the dynamic range. In these cases, the smaller effects at the highest and lowest levels were probably due to ceiling effects and floor effects, respectively.

Effects of carrier rate similar to those seen in our study were demonstrated in a recently published study by Galvin and Fu (2005). That study used six subjects with Nucleus-22 or Nucleus-24 implants. There were a number of minor differences between their study and ours: They compared 250 and 2000 pps carriers and used a 20 Hz modulation frequency. They used primarily bipolar configurations (BP+3 and BP+13) and one stimulation site (i.e., one reference-electrode site) per subject. One interesting, but probably minor procedural difference was that their study compared MDTs for the two carrier rates at levels that were matched in loudness to various levels in the dynamic range of a 1000 pps pulse train on a common BP+3 electrode pair, whereas in our study the levels were matched in terms of the percent of the dynamic range for each carrier and each stimulation site. While it is clear that perceived loudness increases as a function of stimulus level and modulation detection thresholds decrease as a function of level, the precise relationship between perceived loudness and modulation detection is not known. The shapes of the MDT versus percent of dynamic range functions reported by Galvin and Fu (2005) were not noticeably different from the functions reported in our study. In both cases the intersubject variability was much larger than any subtle differences that might be based on the two level scales.

C. Effects of electrode location

Our study showed some interesting interactions between carrier rate and stimulation site. These relationships have not been studied previously to our knowledge. On average, the best MDTs at 250 pps were found at the apical stimulation site and the greatest effects of carrier rate on MDTs were also found at this apical site. However, not all subjects showed the same pattern. Thus, we must consider both systematic and more seemingly random variation in MDTs across stimulation sites.

The finding of interactions between carrier rate and the apical versus basal location of the stimulation site is consistent with the finding of apical-basal differences in temporal properties of auditory nerve fibers (Adamson *et al.*, 2002; Liu and Davis, 2006). Alternatively, across-site variation in the effects of carrier rate on MDTs could be due to variation in pathology along the length of the cochlea. The diversity of effects across subjects and stimulation sites suggests that there are interactions between carrier pulse rate and other variables that are specific to individual subjects and individual stimulation sites within subjects. Possible candidates for the variables underlying this observation include the nerve survival pattern and the condition of the implanted scala tympani. Nerve survival pattern in a hearing-impaired patient with a cochlear implant is never complete and the pattern of nerve loss as well as the condition of the surviving neurons varies considerably from patient to patient (Hinojosa and Lindsay, 1980; Nadol, 1997). In addition, fibrous tissue and new bone frequently grow near the implant in the scala tympani, potentially resulting in alterations in the pathways from individual electrodes to the excitable neural elements (Kawano *et al.*, 1998). Finally, the radial position of the electrode array with respect to the modiolus varies along the length of the cochlear implant in ways that are not always predictable (Saunders *et al.*, 2002). These three variables (nerve survival pattern, pattern of tissue growth, and pattern of electrode location with respect to the modiolus) can combine to create seemingly random variation in the number and position of neural elements excited by individual stimulation sites along the length of the electrode array. The pathology that influences these factors could vary systematically from base to apex contributing to the apical-basal differences that we observed and they could also be responsible for the less systematic across-site variation that we also observed.

The magnitudes of differences in MDTs between the apical and basal site for the 250 pps carrier averaged 3.0 dB and were as large as 33.0 dB (Fig. 5). Thus, when comparing modulation detection ability across subjects, it is important to sample MDTs at several sites in order to get an accurate estimate of each subject's relative ability.

D. Relation to dynamic range

Dynamic ranges for electrical stimulation of cochlear implants are typically small and are highly variable across subjects. The smallest dynamic ranges have been found to be associated with poor electrode discrimination, poor place-pitch perception, and poor speech recognition (Blamey *et al.*, 1992; Pfingst *et al.*, 1999; Donaldson and Nelson, 2000). In this study, the subjects with dynamic ranges greater than 7.2 dB all had relatively good MDTs when the 250 pps carrier was used, while those with dynamic ranges less than 7.2 dB showed a range of performance from good to poor. Dynamic ranges were usually much larger for the 4000 pps carrier rate, but MDT performance was typically much poorer for this carrier compared to that for the 250 pps carrier. Thus, large dynamic range per se does not assure better MDT performance.

Galvin and Fu (2005) found no significant correlation between dynamic ranges and mean MDTs. However, three factors that affect dynamic range (subjects, electrode configuration, and pulse rate) were confounded in that analysis. As noted in our study, the relationship between MDTs and dynamic range depends on the variable (carrier rate or subject) that is associated with differences in the dynamic range. The relation between MDTs and dynamic ranges across carrier rates was opposite to the relation between MDTs and dynamic range across subjects with carrier rate held constant, suggesting that at least two different mechanisms underly these relationships.

V. CONCLUSIONS

(1) High carrier pulse rates pose a distinct disadvantage for detection of amplitude-modulation.
(2) Modulation detection ability depends to some extent on the location of the stimulation site along the tonotopic axis of the cochlea.
(3) Presentation of stimuli in the upper regions of the dynamic range results in better modulation detection.
(4) Subjects with larger dynamic ranges tend to have better modulation detection thresholds. However, increasing dynamic range by using higher carrier rates is detrimental to modulation detection.

ACKNOWLEDGMENTS

We express appreciation to our research subjects for their cheerful participation in these studies, to John Middlebrooks for his helpful suggestions and insightful comments, and to Chenfei Ma and Rose Burkholder for assistance with data analysis and presentation. We also express appreciation to Charles Kowalski at the University of Michigan Center for Statistical Consultation and Research for statistical consultation and to Thyag Sadasiwan for programming. This work was supported by NIH NIDCD Grant Nos. R01 DC03808, R01 DC04312, and P30 DC05188.

[1]For the basal site, the default was electrode 4, but that electrode was unusable in two subjects (S2 and S3), so electrode 6 was used in those cases.

Adamson, C. L., Reid, M. A., Mo, Z. L., Bowne-English, J., and Davis, R. L. (**2002**). "Firing features and potassium channel content of murine spiral ganglion neurons vary with cochlear location," J. Comp. Neurol. **447**, 331–350.

Blamey, P. J., Pyman, B. C., Gordon, M., Clark, G. M., Brown, A. M., Dowell, R. C., and Hollow, R. D. (**1992**). "Factors predicting postoperative sentence scores in postlinguistically deaf adult cochlear implant patients," Ann. Otol. Rhinol. Laryngol. **101**, 342–348.

Colletti, V., and Shannon, R. V. (**2005**). "Open set speech perception with auditory brainstem implant?," Laryngoscope **115**, 1974–1978.

Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.

Drennan, W. R., and Pfingst, B. E. (**2006**). "Current-level discrimination in the context of interleaved multichannel stimulation in cochlear implants: Effects of number of stimulated electrodes, pulse rate and electrode separation," J. Assoc. Res. Otolaryngol. **7**, 308–316.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech perception," J. Acoust. Soc. Am. **95**, 1053–1064.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of reducing

slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.
Fu, Q-J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.
Fu, Q-J., Chinchilla, S., and Galvin, J. J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users." J. Assoc. Res. Otolaryngol. **5**, 253–260.
Fu, Q-J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by Nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.
Fu, Q-J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.
Galvin, J. J. III, and Fu, Q-J., (**2005**). "Effects of stimulation rate, mode and level on modulation detection by cochlear implant users," J. Assoc. Res. Otolaryngol. **6**, 269–279.
Hinojosa, R., and Lindsay, J. R. (**1980**). "Profound deafness: Associated sensory and neural degeneration," Arch. Otolaryngol. Head Neck Surg. **106**, 193–209.
Holden, L. K., Skinner, M. W., Holden, T. A., and Demorest, M. E. (**2002**). "Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy," Ear Hear. **23**, 463–476.
Kawano, A., Seldon, H. L., Clark, G. M., Ramsden, R. T., and Raine, C. H. (**1998**). "Intracochlear factors contributing to psychophysical percepts following cochlear implantation," Acta Oto-Laryngol. **118**, 313–326.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liu, Q., and Davis, R. L. (**2006**). "From apex to base: How endogenous neuronal membrane properties are distributed in the spiral ganglion," J. Assoc. Res. Otolaryngol. Abst. **29**, 305.
Middlebrooks, J. C. (**2005**). "Transmission of temporal information from a cochlear implant to the auditory cortex," J. Assoc. Res. Otolaryngol. Abst. **28**, 91.
Nadol, J. (**1997**). "Patterns of neural degeneration in the human cochlea and auditory nerve: Implications for cochlear implantation," Otolaryngol.-Head Neck Surg. **117**, 220–228.
Pfingst, B. E., Holloway, L. A., Poopat, N., Subramanya, A. R., Warren, M. F., and Zwolan, T. A. (**1994**). "Effects of stimulus level on nonspectral frequency discrimination by human subjects," Hear. Res. **78**, 197–209.
Pfingst, B. E., Holloway, L. A., Zwolan, T. A., and Collins, L. M. (**1999**). "Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants," Hear. Res. **134**, 105–115.
Pfingst, B. E., Xu, L., Thompson, C. S., and Ma, C. (**2005**). "Effects of carrier pulse rate on modulation detection in subjects with cochlear implants," 2005 Abst. Conf. Implant. Aud. Prost., p. 139.
Rubinstein, J. T., and Hong, R. (**2003**). "Signal coding in cochlear implants: Exploiting stochastic effects of electrical stimulation," Ann. Otol. Rhinol. Laryngol. **112**, 14–19.
Saunders, E. *et al.* (**2002**). "Threshold, comfortable level and impedance changes as a function of electrode-modiolar distance," Ear Hear. **23**, 28S–40S.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M., (**2000**). "Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.
Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.
Wilson, B. S. (**2004**). "Engineering design of cochlear implants," in *Cochlear Implants: Auditory Prostheses and Electrical Hearing*, edited by F.-G. Zeng, A. N. Popper, and R. R. Fay (Springer, New York), pp. 14–52.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Wilson, B. S., Finley, C. C., Lawson, D. T., and Zerbi, M. (**1997**). "Temporal representations with cochlear implants," Am. J. Otol. **18**, S30–S34.
Xu, L., and Pfingst, B. E. (2005). "Effects of carrier pulse rate on modulation detection in subjects with cochlear implants," The 25th Politzer Society Meeting, Seoul, Korea.
Xu, L., Thompson, C., and Pfingst, B. E. (**2005**). "Relative contributions of spectral and temporal cues for phoneme recognition," J. Acoust. Soc. Am. **117**, 3255–3267.
Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). "Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses," J. Acoust. Soc. Am. **112**, 247–258.

Sensitivity of a continuum vocal fold model to geometric parameters, constraints, and boundary conditions

Douglas D. Cook[a)]
School of Mechanical Engineering, Purdue University, 140 S. Intramural Drive, West Lafayette, Indiana 47907-2031

Luc Mongeau[b)]
School of Mechanical Engineering, Purdue University, 140 S. Intramural Drive, West Lafayette, Indiana 47907-2031

(Received 10 July 2006; revised 10 January 2007; accepted 10 January 2007)

The influence of key dimensional parameters, motion constraints, and boundary conditions on the modal properties of an idealized, continuum model of the vocal folds was investigated. The Ritz method and the finite element method were used for the analysis. The model's vibratory modes were determined to be most sensitive to changes in the anterior-posterior length of the vocal fold model, due to the influence of three-dimensional stress components acting in the transverse plane. Anterior/posterior boundary conditions were found to have a significant influence on the vibratory response. Overestimation of modal frequencies resulted when vibration of the structure was restricted to the transverse plane. The overestimation of each modal frequency was proportional to the ratio of longitudinal to transverse Young's modulus, and was significant for ratio values less than 20.
 [DOI: 10.1121/1.2536709]

PACS number(s): 43.70.Bk [BHS] Pages: 2247–2253

I. INTRODUCTION

Computational models of the human vocal folds are often used to gain a better understanding of the dynamics of flow-induced oscillations involved in voice production. Such models are usually two-dimensional or three-dimensional. Two-dimensional models approximate the human vocal folds as planar structures. This simplification reduces the computational expense involved in solving the equations of motion. Examples include Ishizaka and Flanagan (1972), Horacek and Svec (2002), and Thomson *et al.* (2005). Three-dimensional models are computationally more expensive, but yield more detailed representations of the vocal fold motion and their structural characteristics. For example, a three-dimensional structural model (without fluid-structure coupling) was used to simulate the adduction and abduction of the vocal folds by Hunter *et al.* (2004). Quasi-three-dimensional or hybrid models have also been investigated by Alipour *et al.* (2000).

Computational models often use geometrical approximations in order to simplify the analysis. To this end, the vocal fold structure was idealized as a solid rectangular parallelepiped by Berry and Titze (1996) (see Fig. 1). The material was assumed to be transversely isotropic to account for presumably greater stiffness in the direction of muscle fibers. Planar displacements of the structure were also assumed, i.e., every point of the structural model was required to move in a plane. Modal analysis of the idealized rectangular structure showed that for a nearly incompressible material formulation, the second and third modes of vibration occurred at nearly the same frequency across a wide range of tissue properties. Superposition of these two modes was reported to result in a converging/diverging orifice geometry. It was conjectured that both these modes might respond to aerodynamic loading, as postulated based on theoretical arguments (Titze, 1988).

In the present study, the continuum model of Berry and Titze (1996) was used to perform a geometric sensitivity study of the idealized vocal fold structure. The objectives were to determine the influence of spatial dimensions on vibratory response, and to investigate the validity of the planar displacement assumption. This information may help in determining the model complexity required to achieve an optimal compromise between computational cost, and model details. Because the mechanical structure of the human vocal folds is deceivingly complex, the characterization of simpler, idealized models constitutes a logical first step towards the development of more realistic structural models of the human vocal folds.

II. METHODS

A. Model geometry and material properties

The material properties and nominal model dimensions used in this study were the same as those used by Berry and Titze (1996) and are listed in Table I. A sketch of the solid domain and the boundary conditions is shown in Fig. 1. No motion was allowed for nodes on the anterior/posterior and the lateral faces. The inferior/superior and medial faces were free. A linearly elastic, transversely isotropic solid was used to characterize the mechanical properties of the vocal fold. The *xz* plane was designated as the transverse plane, with the longitudinal direction aligned with the *y* axis.

[a)]Author to whom correspondence should be addressed. Electronic mail: ddcook@purdue.edu
[b)]Present address: Mechanical Engineering Department, McGill University, 817 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 2K6.

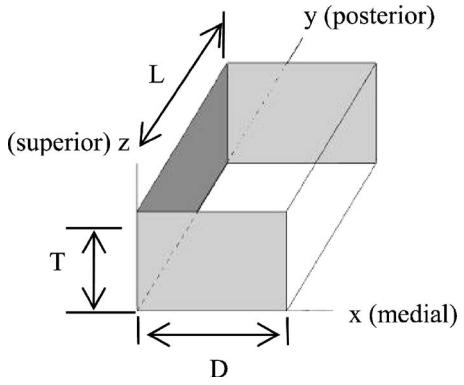


FIG. 1. Schematic diagram of the continuum model (adapted from Berry and Titze, 1996) shaded faces indicate fixed boundary conditions.

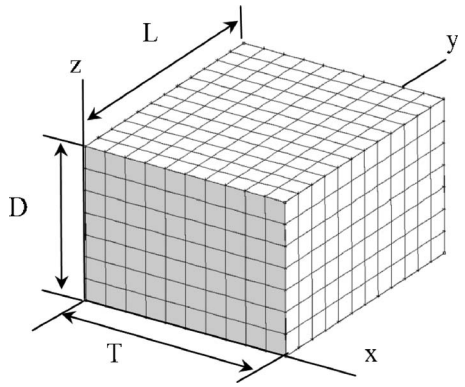


FIG. 2. Schematic of the computational grid. Length, *L*, thickness, *T*, and depth, *D*, shown. Shaded face indicates fixed boundary condition.

B. Modal analysis procedures

Modal analysis was performed using both the Ritz method and the finite element method. The use of the Ritz method allowed for direct comparisons with the results of Berry and Titze (1996), and for verification of the finite element method implementation. The finite element method was then used to examine the effects of the planar displacement assumption, because it more readily allowed for the implementation of three-dimensional displacements.

The Ritz method implementation was based on that described by Berry and Titze (1996). Polynomial displacement functions were used in the x and z directions, and a sinusoidal displacement function was used in the y direction, between anterior and posterior boundaries. No displacement was allowed in the y direction.

For the finite element analysis, model geometry and boundary conditions were consistent with Table I and the Ritz formulation. The rectangular domain, shown in Fig. 2, was meshed using 840 quadratic (27 node) three-dimensional solid finite elements for a total of 7875 nodal points. Repeated calculations using varying mesh densities confirmed that this configuration represents a mesh-converged solution. A commercially available finite element code was used (ADINA R&D Inc., 2005).

Modal analysis was performed using the continuum model of Fig. 1 and the transversely isotropic material properties given in Table I. The first five mode shapes and modal frequencies obtained using the Ritz method are shown in Fig. 3. All resulting modal frequencies were within 2% of those obtained by Berry and Titze (1996). Mode shapes and convergence behavior were also comparable to the results of Berry and Titze (1996), thus cross validating the Ritz method and finite element method implementations.

III. GEOMETRIC SENSITIVITY ANALYSIS

A geometric sensitivity study was conducted using the Ritz method. The three spatial dimensions (thickness, T; depth, D; and length, L) were each varied by approximately 30%. Only one dimension was varied at a time, with all other model parameters held constant.

A. Results

The first five natural frequencies are shown for a range of thickness values in Fig. 4(a). The second and third modal (or natural) frequencies were within 4 Hz of each other. The first, second, and third modal frequencies were nearly insensitive to changes in thickness. Relative changes in frequency, calculated by using the parameters of Table I as reference values, are shown in Fig. 4(b). The first three modal frequencies changed by less than 1% for thickness variations of almost 60%. The fourth and fifth modal frequencies decreased by approximately 15% over the same range.

The influence of depth is illustrated in Fig. 5(a). Relative values are shown in Fig. 5(b). The second and third modal frequency values were within 3 Hz of each other for depths ranging between 0.8 and 1.2 cm. The modal frequencies of the second, third, fourth, and fifth modes decreased by 10–15% for a 40% increase in depth. The first modal frequency remained nearly constant, changing by only 3% over the same range.

TABLE I. Mode dimensions and tissue properties.

	CGM Units	SI Units
Lateral depth, D	1.0 cm	0.01 m
Longitudinal (anterior-posterior) length, L	1.2 cm	0.0012 m
Vertical thickness, T	0.7 cm	0.007 m
Tissue density, ρ	1.03 g/cm^3	1030 kg/m^3
Transverse Young's modulus, E_{trans}	10^5 dyne/cm2	10^4 Pa
Longitudinal shear modulus, μ_{trans}	10^5 dyne/cm2	10^4 Pa
Transverse Poisson's ratio, ν_{trans}	0.9999	0.9999
Longitudinal Poisson's ratio, ν'	0	0

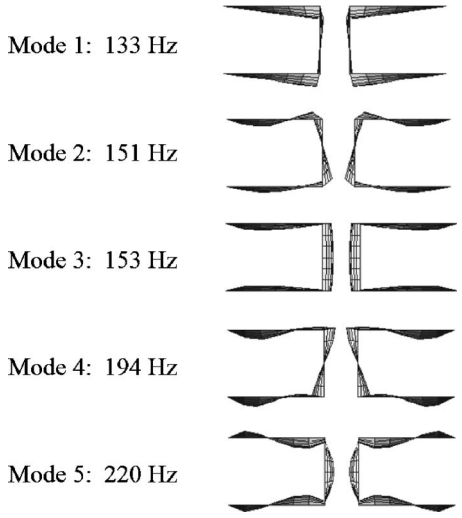


FIG. 3. Modal frequencies and mode shapes of the first five modes obtained by the Ritz method. Mode shapes are viewed as coronal cross sections.

Changes in modal frequencies are shown as a function of length in Fig. 6(a). Relative values are shown in Fig. 6(b). The second and third modal frequencies were within 3 Hz of each other over a length variation of about 30%. All modal frequencies decreased with increased length. The modal frequency rate of decrease was noticeably higher than for the other dimensions. The frequency of the least sensitive mode (mode five) decreased by 12% for a 30% change in length. The frequency of the most sensitive mode (mode one) decreased by 33% for a 30% change in length.

B. Discussion

The modal frequencies were least sensitive to changes in thickness, *T*. The model was more sensitive to changes in depth than to changes in thickness, as shown by a comparison between Figs. 5 and 6. Finally, the model was most sensitive to changes in length, *L*. The least sensitive modal frequency in Fig. 6 decreased more rapidly with length than that of any of the modes for the other dimensions, shown in Figs. 4 and 5.

In general, modal frequencies increase with stiffness, and decrease with mass. Structural stiffness depends upon

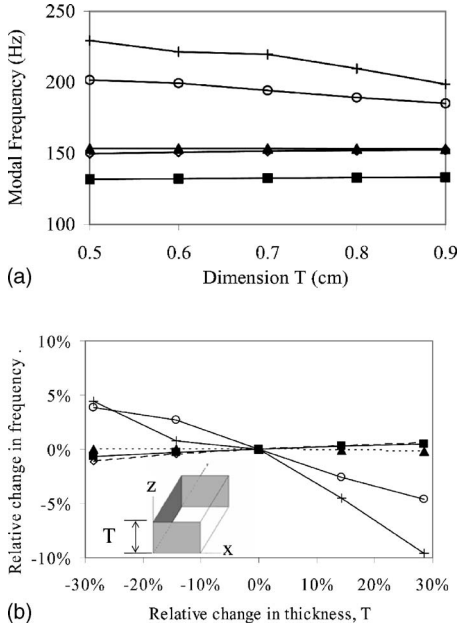


FIG. 4. (a) Modal frequencies as a thickness, *T*. (b) Relative change in modal frequencies. ■: Mode 1; ◇: Mode 2; ▲: Mode 3; ∘: Mode 4; +: Mode 5.

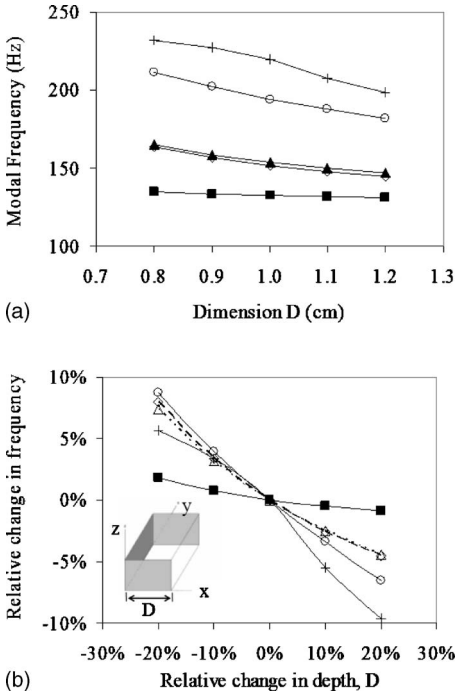


FIG. 5. (a) Modal frequencies as a function of depth, *D*. (b) Relative change in modal frequencies. ■: Mode 1; ◇: Mode 2; ▲: Mode 3; ∘: Mode 4; +: Mode 5.

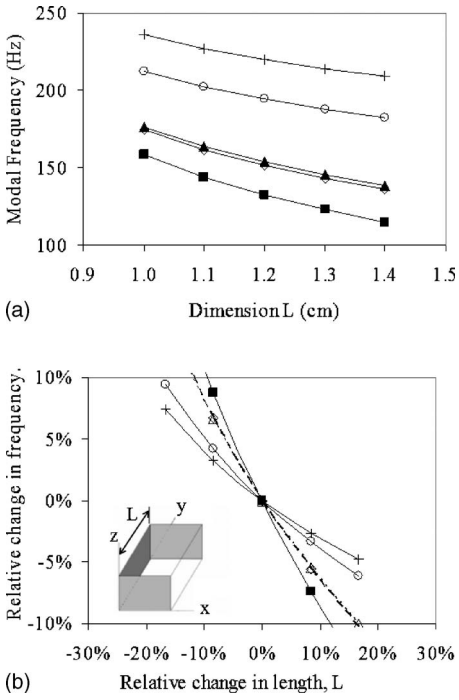


FIG. 6. (a) Modal frequencies as a function of length, *L*. (b) Relative change in modal frequencies. ■: Mode 1; ◇: Mode 2; ▲: Mode 3; ∘: Mode 4; +: Mode 5.

both tissue properties and boundary conditions. Fixed (zero displacement) boundary conditions tend to stiffen the structure while free boundary conditions tend to reduce the stiffness. An increase in length increases the distance between the anterior/posterior fixed boundaries and thus adds mass to the structure while decreasing stiffness. Changes to the model thickness increase mass and also increase the fixed boundary area on three sides. This adds mass and stiffness in approximately equal quantities, resulting in little change in modal frequencies.

Previous studies have reported that most of the vibratory energy is found within the first three modes of vibration (Alipour *et al.*, 2000; Berry *et al.*, 2001). Hence, it has been proposed that these modes are critical to the self-oscillation of the vocal folds. It is interesting to note that the second and third modes were within 5 Hz of each other over a wide range of model dimensions. This suggests that these modes could possibly converge to similar frequencies for large displacements in situations where the material behaves nonlinearly and the modal frequencies are amplitude dependent. Such mode merging is one possible self-oscillation mechanism (Berry and Titze, 1996). It is also worth noting that the fourth and fifth modes may also be important in voice production since they occur near the first formant frequency (resonance of the vocal tract), thus possibly contributing to both radiated sound and acoustic coupling.

C. Effect of boundary conditions and out-of-plane stresses

The influence of anterior/posterior boundary conditions was investigated by modifying the model length values. The modal frequencies for large variations in length are shown in Fig. 7. Two distinct vibratory behaviors may be identified. The modal frequencies are independent of length for L greater than six centimeters. This occurs because the anterior/posterior boundary conditions are less influential as the distance between these boundaries is increased. As a result, the model behavior is governed by the remaining two spatial dimensions (thickness and depth) for $L>6$ cm. The model behavior in this region may be described as two dimensional. In fact, modal frequencies of a corresponding two-dimensional plane strain model (described in Sec. 6) were found to be identical to those calculated for an extremely long ($L=10^6$ cm) three-dimensional model.

The modal frequencies were highly dependent on length for values between $L=0$ cm and $L=4$ cm (Fig. 7). This is because the anterior/posterior boundary conditions become increasingly influential as the structure length decreases. All three spatial dimensions (length, depth, and thickness) affected model behavior for $0<L<4$ cm. This is identified as the three-dimensional region. A transitional region of moderate dependence on length exists for values of L between about four and six centimeters. A continuum model representative of human vocal folds would have a nominal length of about 1.2 cm, clearly within the three-dimensional region.

The model behavior in the three-dimensional region is caused by anterior/posterior boundary conditions which induce shear and normal stresses on the transverse (xz) plane.

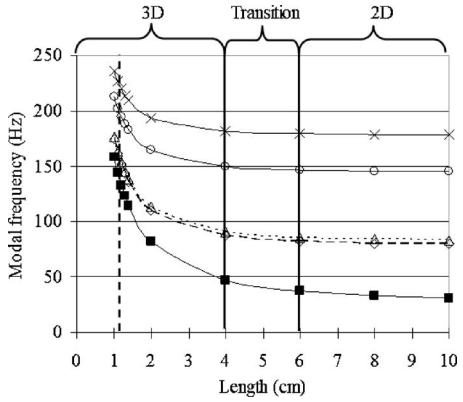


FIG. 7. Modal frequency vs. model length, L. The nominal length is designated by a dashed line at $L=12$ cm. ■: Mode 1; ◇: Mode 2; ▲: Mode 3; ∘: Mode 4; ×: Mode 5.

Boundary conditions and associated stresses act to stiffen the structure. The three-dimensional model responded at higher frequencies than the two-dimensional model because of this stiffening effect. This observation is consistent with Thomson *et al.* (2005) who reported that a three-dimensional physical model of the vocal folds exhibited higher modal frequencies than a corresponding two-dimensional numerical model.

The original continuum model proposed by Berry and Titze (1996) accounted for all stress components. In contrast, two-dimensional vocal fold structural models that rely upon the plane strain assumption do not properly account for out-of-plane stresses and boundary conditions. This is because two-dimensional formulations entirely neglect the anterior/posterior boundary conditions. Because anterior/posterior boundary conditions likely have a significant effect on modal frequencies, two-dimensional representations may not be suitable for modeling the human vocal folds.

IV. INFLUENCE OF THE PLANAR DISPLACEMENT CONSTRAINT

The planar displacement assumption is commonly used in structural models of the human vocal folds (Alipour *et al.*, 2000; Tao *et al.*, 2006). This assumption is related to observations of excised canine larynx vibrations (Berry *et al.*, 2001), where longitudinal displacements of canine vocal folds were reportedly one order of magnitude smaller than displacements in the other two directions. But small displacements may cause stresses which significantly affect vibratory response. The continuum model has been shown to be sensitive to stresses that act on the xz plane. In order to quantitatively investigate the influence of the planar displacement condition, a three-parameter formulation of the constitutive laws was used.

A. Three-parameter formulation for the constitutive equations

The standard stress/strain relations for a transversely isotropic material depend upon five independent material properties. The material properties in the transverse (xz) plane are the Young's modulus, E, and either Poisson's ratio ν, or shear

modulus μ. The remaining material properties are Poisson's ratio, ν', and shear modulus μ', both perpendicular to the transverse plane, as well as the longitudinal Young's modulus, E' (Lekhnitskii, 1963). When an additional assumption of incompressibility is imposed, the stress-strain relations can be represented as

$$[S] = \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{bmatrix} = \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{bmatrix}. \tag{1a}$$

$$[S] = \begin{bmatrix} \frac{1}{E} & -\frac{1}{2E'}\left(\frac{1}{2E'}-\frac{1}{E}\right) & 0 & 0 & 0 \\ \frac{1}{2E'} & \frac{1}{E'} & -\frac{1}{2E'} & 0 & 0 & 0 \\ \left(\frac{1}{2E'}-\frac{1}{E}\right) & -\frac{1}{2E'} & \frac{1}{E} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu'} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu'} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{4}{E}-\frac{1}{E'} \end{bmatrix} \tag{1b}$$

where σ_i and ε_i represent normal stresses and strains in the i direction, while τ_{ij} and γ_{ij} represent shear stresses and strains in the ij plane. This formulation is advantageous because it involves only three independent material properties (E, E', and μ') instead of five (see Sec. VI B for detailed information). The remaining properties, Young's moduli and the shear modulus, are relatively easy to measure, and values for these properties are often available in the literature. Because empirical values for several other material properties are not available, many studies involving three-dimensional vocal fold models have resorted to estimations of these parameters or parametric studies (Alipour *et al.*, 2000; Berry and Titze, 1996). The incompressibility assumption reduces the number of material constants, thereby facilitating the characterization of the material behavior based on experimental data. However, to date, differential measurement of longitudinal and transverse Young's modulus of laryngeal tissues has not been accomplished.

The model of Sec. III utilized the assumption that no displacements occur in the y direction (Berry and Titze, 1996). This assumption yields the following compliance matrix:

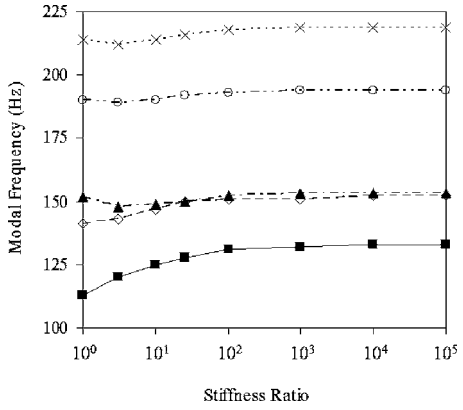


FIG. 8. Modal frequencies as a function of stiffness ratio. ■: Mode 1; ◇: Mode 2; ▲: Mode 3; ∘: Mode 4; ×: Mode 5.

$$[S] = \begin{bmatrix} \frac{1}{E} & 0 & -\frac{1}{E} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{E} & 0 & \frac{1}{E} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu'} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu'} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{4}{E} \end{bmatrix}. \tag{2}$$

The same matrix is also obtained by taking the limit of Eq. (1) as E', approaches infinity. Thus, the planar displacement assumption is equivalent to assuming that the material is infinitely stiff in the longitudinal direction. The validity of this assumption was investigated, as described below.

B. Variation of longitudinal stiffness

The planar displacement assumption was relaxed in order to determine its influence on the vibratory behavior of the continuum model. The model parameters were identical to those used in the Ritz method implementation. However, displacements in the y direction were allowed. The finite element method was used to perform modal analysis.

The longitudinal stiffness, E', was varied between $E'=E$ (isotropy), and a value 100,000 times greater (to approximate infinite longitudinal stiffness). This corresponds to a stiffness ratio, $n=E'/E$, ranging from $n=1$ to $n=10^5$. The stiffness ratio was used to enforce the condition of incompressibility through appropriate values of Poisson's ratio (see Sec. 5.2). The remaining model parameters were those given in Table I.

C. Results and Discussion

The first five modal frequencies are shown as functions of the stiffness ratio, n, in Fig. 8. For $n<10^3$, some variation was observed, especially in the range $1<n<10^2$. However, the model's vibratory characteristics were relatively independent of stiffness ratio for $n>10^3$. In fact, the frequencies

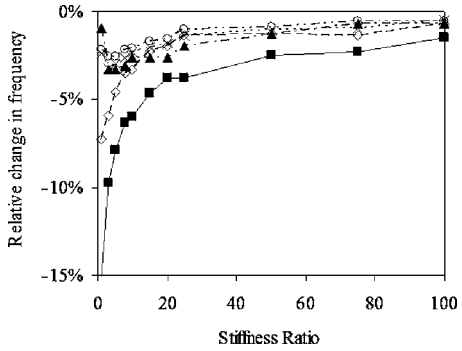


FIG. 9. Relative differences in modal frequency as a function of stiffness ratio. ■: Mode 1: ◇: Mode 2; ▲: Mode 3; ○: Mode 4; ×: Mode 5.

obtained for $n=10^5$ were identical to those predicted by the Ritz method. This is because the longitudinal stiffness corresponding to $n=10^5$ was large enough for the displacements to be planar, as assumed in the Ritz method model.

The effects of constraints on transverse motion are illustrated by normalizing the data shown in Fig. 8. Each modal frequency was divided by its respective limiting value (as obtained by the Ritz method). For example, each modal frequency value for Mode 1 was divided by 133 Hz, and each value for Mode 2 curve was divided by 151 Hz, etc. The results are shown in Fig. 9, for stiffness ratio values between 0 and 100. Negative relative values were obtained in all cases. This is because modal frequencies corresponding to a finite longitudinal stiffness are naturally less stiff than the planar displacement case. The modal frequencies approach the limiting value as n is increased. The relative difference in modal frequencies for the first five modes was less than 5% for $n>20$. Thus, the planar displacement assumption may be considered appropriate over that range. In general, it may be concluded that the validity of the planar displacement assumption depends upon the ratio of the longitudinal to transverse stiffness of the vocal fold tissue.

V. CONCLUSIONS

A sensitivity analysis of the continuum model of Berry and Titze (1996) was performed. The results showed that the modal frequencies of an idealized continuum model were most sensitive to changes in the model length. The planar displacement assumption was found to be equivalent to the assumption of infinite stiffness in the y direction. This assumption did not introduce serious errors in the modal frequency values for stiffness ratio values greater than 20. However, there is very little empirical data concerning this ratio. The actual validity of the planar displacement assumption as it pertains to the human vocal folds will hopefully be clarified as further empirical data become available.

The results also indicated that out-of-plane stresses have a significant effect on the vibratory characteristics of the continuum model. Because these stresses are not accounted for by two-dimensional structural models, the vibratory response of such models may not accurately represent the dynamic response of the human vocal folds. More research is needed to ascertain the exact role of out-of-plane stresses in the human vocal folds. This may be accomplished by the use of detailed finite element modal analyses that more accurately represent the layered structure and geometry of the human vocal folds. In particular, comparisons between two-dimensional and three-dimensional fluid-structure interaction models may provide valuable insight into the relative importance of elastic and flow-induced stresses.

ACKNOWLEDGMENTS

This study was supported by Grant No. R01 DC005788 from the National Institute on Deafness and Other Communication Disorders. The authors would like to thank Professors Arvind Raman and Eric Nauman of Purdue University, and Rosaire Mongrain of McGill University, for their valuable help and suggestions.

APPENDIX

Two-dimensional model

The two-dimensional model described in Sec. III C consisted of a rectangular domain of thickness T and depth D. The domain was assumed to be a two-dimensional plane-strain representation of the continuum model. All parameters were the same as in Table I. One edge of length T was fixed while the three other sides were free to vibrate. Both Ritz method (using polynomial displacement functions) and finite element method modal analyses were performed. Results from both methods were within 1%. This model may be thought of as a single xz plane of the three-dimensional continuum model (Fig. 1), but without the anterior/posterior boundary conditions.

Reduction of independent material properties

Characterization of a transversely isotropic material requires five independent material properties: E, E', ν', μ', and either ν, or μ (Lekhnitskii, 1963). By applying the condition of incompressibility, symmetry of elastic coefficients, and symmetry of Poisson's ratios, the following relations are obtained (Itskov and Aksel, 2002):

$$\nu_{xy}=\nu_{zy}=\frac{1}{2}, \tag{1A}$$

$$\nu_{yz}=\nu_{yx}=\frac{E}{2E'}, \tag{2A}$$

$$\nu_{zx}=\nu_{xz}=1-\frac{E}{2E'}. \tag{3A}$$

Where $E=E_x=E_z$ is the transverse Young's modulus, and $E'=E_y$ is the longitudinal Young's modulus. The Poisson's ratios, ν_{ij}, are those given in the stress/strain relations for a general orthotropic material:

$$\begin{bmatrix} \frac{1}{E_x} & -\frac{\nu_{xy}}{E_y} & -\frac{\nu_{xz}}{E_z} & 0 & 0 & 0 \\ -\frac{\nu_{yx}}{E_x} & \frac{1}{E_y} & -\frac{\nu_{yz}}{E_z} & 0 & 0 & 0 \\ -\frac{\nu_{zx}}{E_x} & -\frac{\nu_{zy}}{E_y} & \frac{1}{E_z} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu_{xy}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu_{yz}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\mu_{zx}} \end{bmatrix} \times \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{yz} \\ \sigma_{zx} \end{pmatrix} = \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{xy} \\ \varepsilon_{yz} \\ \varepsilon_{zx} \end{pmatrix}. \quad (4A)$$

In this equation, σ_i and ε_i represent normal stresses and strains in the i direction, while τ_{ij} and γ_{ij} represent shear stresses and strains in the ij plane. Likewise, the shear modulus in the i-j plane, μ_{ij}, indicates the material's resistance to shear deformation in the same plane. Poisson's ratio, ν_{ij}, relates induced lateral strain to applied strain.

Substitution of these Eqs. (1A)–(3A), into Eq. (4A), along with the relation of shear modulus to Poisson's ratio and transverse stiffness (Popov, 1990; Lekhnitskii, 1963) produces the stress/strain relations of Eq. (1).

Alipour, F., Berry, D., and Titze, I. R. (**2000**). "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

ADINA Theory and Modeling Guide Volume 1: Solids and Structures (ADINA R & D, Inc., Watertown, MA, **2005**).

Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.

Horacek, J., and Svec, J. G., (**2002**). "Instability boundaries of a vocal fold modeled as a flexibly supported rigid body vibrating in a channel conveying fluid," *Proceedings of IMECE 2002*, November 17-22, New Orleans, LA, 1–12.

Hunter, E. J., Titze, I. R., and Alipour, F. (**2004**). "A three-dimensional model of vocal fold abduction/adduction," J. Acoust. Soc. Am. **115**, 1747–1759.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1267.

Itskov, M., and Aksel, N., (**2002**). "Elastic constants and their admissible values for incompressible and slightly compressible anisotropic materials," Acta Mech. **157**, 81–96.

Lekhnitskii, S. G. (**1963**). *Theory of Elasticity of an Anisotropic Body* (Holden-Day, San Francisco), p. 25.

Popov, E. P. (**1990**). *Engineering Mechanics of Solids* (Prentice–Hall, Englewood Cliffs, NJ) pp. 150–151.

Tao, C., Jiang, J. J., and Zhang, Y. (**2006**). "Simulation of vocal fold impact pressures with a self-oscillating finite-element model," J. Acoust. Soc. Am. **119**, 3987–3994.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.

Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.

A two-dimensional biomechanical model of vocal fold posturing

Ingo R. Titze[a)]
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242
and National Center for Voice and Speech, Denver Center for the Performing Arts, Denver, Colorado 80204

Eric J. Hunter
National Center for Voice and Speech, Denver Center for the Performing Arts, Denver, Colorado 80204

(Received 3 May 2006; revised 19 January 2007; accepted 24 January 2007)

The forces and torques governing effective two-dimensional (2D) translation and rotation of the laryngeal cartilages (cricoid, thyroid, and arytenoids) are quantified on the basis of more complex three-dimensional movement. The motions between these cartilages define the elongation and adduction (collectively referred to as posturing) of the vocal folds. Activations of the five intrinsic laryngeal muscles, the cricothyroid, thyroarytenoid, lateral cricoarytenoid, posterior cricoarytenoid, and interarytenoid are programmed as inputs, in isolation and in combination, to produce the dynamics of 2D posturing. Parameters for the muscles are maximum active stress, passive stress, activation time, contraction time, and maximum shortening velocity. The model accepts measured electromyographic signals as inputs. A repeated adductory–abductory gesture in the form |*hi-hi-hi-hi-hi*| is modeled with electromyographic inputs. Movement and acoustic outputs are compared between simulation and measurement. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2697573]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 2254–2260

I. INTRODUCTION

Vocal fold dynamics for speaking and singing can be treated in two parts: (1) large and relatively slow deformations occurring when the vocal folds are positioned for voicing, coughing, and breathing by moving laryngeal cartilages with muscle forces, and (2) small and relatively fast deformations occurring when the tissue is driven into self-sustained oscillation by aerodynamic and acoustic pressures. These two parts are referred to, respectively, as vocal fold *posturing* and vocal fold *vibration*. Posturing is further subdivided into adducting (or abducting) the medial surfaces of the vocal folds and elongating (or shortening) of the vocal folds. This posturing occurs in a nonperiodic (but ultimately always cyclic) fashion at frequencies of 1–10 Hz. Vocal fold vibration, on the other hand, occurs at 100–1000 Hz.

Although vocal fold posturing and vocal fold vibration are thus thought to be separate mechanical processes, many parameters of vibration (e.g., fundamental frequency, amplitude of vibration, and voice onset time) are dependent on posturing. For example, adduction of the vocal processes of the arytenoid cartilages has been shown to affect the intensity of the voice (Titze and Sundberg, 1992; Murry *et al.*, 1998), is involved in fundamental frequency regulation (Hirano *et al.*, 1970; Honda, 1983), voice initiation (Cooke *et al.*, 1997), devoicing or vocal offset (Yoshioka, 1981), and ventilation or glottal aspiration (Tomori *et al.*, 1998). Some voice disorders have been attributed to insufficiencies of voice onset and offset (e.g., Peters *et al.*, 1986; Werner-Kukuk and von Leden, 1970; Gallena *et al.*, 2001). Adequacy of vocal fold posturing has been used to quantify vocal improvement as a result of therapy (Boone and McFarlane, 1994; Gallena *et al.*, 2001) and singing training (Miller, 1994).

Speech simulation at the biomechanical level requires a mathematical description of vocal fold posturing. Because many of the empirical data come from electromyography of the laryngeal muscles, the control variables are designed to be simulated muscle activations. The model described in this paper accepts these simulated muscle activations as inputs. The outputs are movement variables of the arytenoid cartilages and (if vocal fold vibration is included) the acoustic pressures and flows in the vocal tract.

Vocal fold adduction, abduction, and elongation are not simple one-dimensional movements because the entire medial surface of the vocal folds is nonuniformly deformed. Anterio-posterior and inferio-superior variations are typical. Also, the three-dimensional (3D) nature of arytenoid cartilage movement (a rocking motion) affects the dynamics of posturing, as described years ago by Frable (1961), Maue and Dickson (1971), von Leden and Moore (1961), and Sellars and Keen (1978). More recently, Selbie *et al.* (2002) studied the cricoarytenoid joint facets in great detail. Earlier it was shown that rocking occurs along the long axis of the cricoid facet, which has curvatures in all directions (Selbie *et al.*, 1998). A matching set of opposite curvatures was found on the arytenoid facet, so that the joint resembled a tongue-in-groove architecture. Selbie *et al.*, (2002) showed that arytenoid cartilage motion was highly constrained and that rotation about a vertical axis perpendicular to the plane of the glottis was not possible, but that such rotation was "apparent" when the 3D motion was projected onto a horizontal plane (the plane of the glottis). Their results suggest that, for the modeling presented in this paper, translational displacements and rotations of the cricoarytenoid joint should be

a)Electronic mail: ingo-titze@dcpa.org

 0001-4966/2007/121(4)/2254/7/$23.00

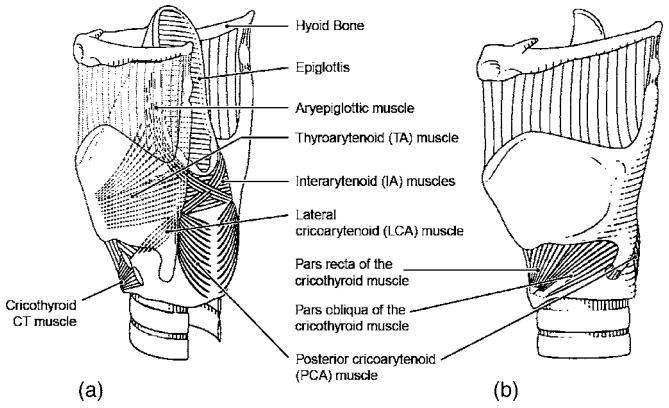


FIG. 1. Framework of the larynx with intrinsic musculature: (a) posterio-lateral view and (b) lateral view.

viewed only as "effective" two-dimensional (2D) displacements, based on direction cosines determined from the actual 3D motions.

A 3D continuum model of vocal fold posturing (Hunter *et al.*, 2004) showed that the range and speed of adduction and abduction were not only dependent on individual laryngeal muscle properties, but also on the rotational and translational mechanics of passive (connective) tissue. Because of the complexity of 3D continuum mechanics, however, only adduction/abduction of the arytenoid cartilages was addressed; elongation from cricothyroid joint motion was not included in the former 3D model, but will be addressed here for a parsimoniously constructed 2D case.

Vocal fold posturing requires at least an x and a y displacement of the vocal processes relative to the glottal midplane. The x displacement (medial-lateral) is for adduction and the y displacement (ventro-dorsal or anterior-posterior) is for elongation. Although the 3D rocking of the arytenoid cartilages over articular surfaces results in x, y, and z motions of the vocal processes, the important motion for our modeling is in the x-y plane (Hunter and Titze, 2004). Therefore, it is possible to capture the main features of vocal fold elongation and adduction in the transverse plane, as will be demonstrated in this paper.

For application to phonetics, the primary question of interest is: Can the biomechanical model predict the time-course of adduction-abduction and vocal fold elongation with muscle activations derived from natural speech?

II. MECHANICS OF TWO-DIMENSIONAL POSTURING

We have begun the mathematical implementation of two-dimensional modeling by first determining the lines of action of all intrinsic laryngeal muscle fibers in a 3D space. The intrinsic muscles are the cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid (LCA), posterior cricoarytenoid (PCA), and interarytenoid (IA). Figure 1 shows a sketch of these muscles. Because the muscles are not uniform in shape, many small muscle fiber bundles needed to be identified and dissected in previous work for a given muscle [Cox *et al.* (1999) for canine and human CT and TA; Mineck *et al.* (2000) for canine LCA, PCA, and IA]. With the use of a 3D digitizer (Microscribe-3D), the point of origin (x_{1j}, y_{1j}, z_{1j}) and the point of insertion (x_{2j}, y_{2j}, z_{2j}) of the jth fiber bundle of each muscle were measured in a Cartesian coordinate system. This allowed the length l_j of each muscle fiber bundle to defined,

$$l_j = [(x_{1j} - x_{2j})^2 + (y_{1j} - y_{2j})^2 + (z_{1j} - z_{2j})^2]^{1/2}, \tag{1}$$

as well as the x and y direction cosines, α_j and β_j,

$$\alpha_j = (x_{1j} - x_{2j})/l_j, \tag{2}$$

$$\beta_j = (y_{1j} - y_{2j})/l_j. \tag{3}$$

From these measures, the composite x and y direction cosines for the entire muscle force were computed as follows:

$$\alpha = \sum_j (A_j a_j) / \sum_j A_j, \tag{4}$$

$$\beta = \sum_j (A_j \beta_j) / \sum_j A_j, \tag{5}$$

where A_j is the cross-sectional area of a given muscle fiber bundle [typically on the order of $1/\text{mm} \times 1/\text{mm}$; Cox *et al.* (1999)]. Figure 2 shows a sketch of the lines of action of the forces, and Table I lists anatomically measured direction cosines α and β for four of the muscles (Mineck *et al.*, 2000; Berry *et al.*, 2003; Titze, 2006). The direction cosines of the CT muscle will be discussed separately because it does not attach to the arytenoid cartilage and is therefore a component of a separate mechanical system. Also in Table I are the directional moment arms γ, to be discussed later.

It must be pointed out that, at this stage of modeling, our data sets represent a mixture of canine and human species. Biomechanical data are mainly from canines and anatomical data are mainly from humans. Hence, our results can only be considered as representative of human biomechanics to the extent that the canine is a reasonable model. Work is continuing to complete data tables for both species (Titze, 2006). Fortunately, human ligament stress-strain curves were included. Hence, the vocal fold length changes (to be discussed) are a fairly good approximation to human length changes.

Figure 3 shows a transverse section through the right vocal fold, arytenoid cartilage, and thyroid cartilage, with the cricoid cartilage sketched underneath (ring-shaped). The 2D lines of action of the muscle forces are shown. TA_1 represents the vocalis portion of the TA muscle, while TA_2 represents the muscularis portion. The overall action of the TA is the vector sum of the two. Also shown are vector forces for the IA, PCA, and LCA.

It has been observed by experimenting with excised larynges (Titze, 2006) that the LCA muscle has the primary function of creating the "effective 2D rotation" of the vocal processes of the arytenoid cartilages (counterclockwise for the right cartilage), which brings the tip of the vocal processes into or near the glottal midplane. In Fig. 3, the symbols ξ_{02} and ψ_{02} describe the position of the vocal process. The IA muscle has the primary function of adducting the arytenoid cartilage near the posterior wall, thereby closing the posterior gap. Abduction is controlled by the PCA

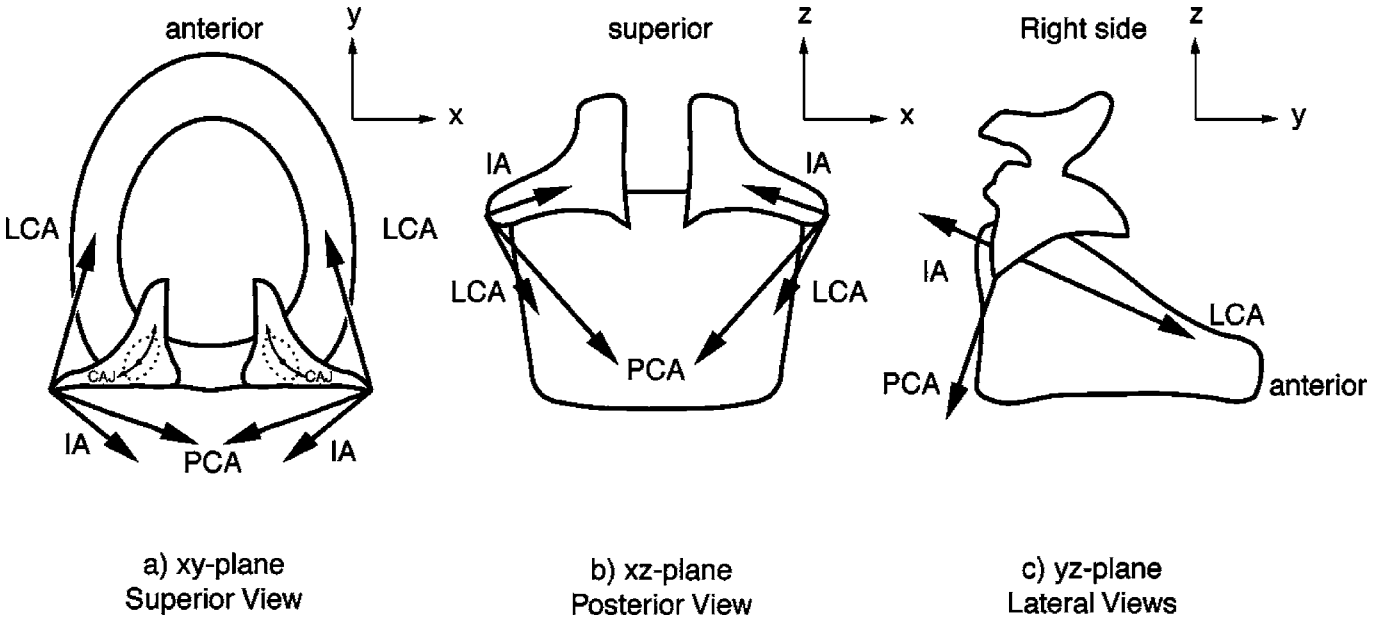


FIG. 2. Lines of action of the forces of the laryngeal muscles.

muscle, which is the antagonist to the LCA for rotation. TA_1 is primarily used for length reduction, but TA_2 is an adductor that creates "effective" counterclockwise rotation of the arytenoid cartilages, similar to the LCA.

A. Equations of motion for cricoarytenoid joint movement

For quantitative modeling, the elastic and viscous forces governing "effective" cricoarytenoid joint (CAJ) translation and rotation need to be known, together with the muscular forces. The x motion of the arytenoid cartilage around the CAJ is

$$\ddot{\xi}_a = \frac{1}{M_a}\left[\sum_{i-1}^{7} \alpha_i F_i - k_y \xi_a - d_x \dot{\xi}_a\right], \tag{6}$$

where α_i is the direction cosines for the ith muscle, ξ_a is the x displacement of the arytenoid cartilage from the center of the CAJ, M_a is the mass of the arytenoid cartilage, k_x is the "effective" translational stiffness, d_x is the damping coefficient, and F_i is the ith tissue force ($i=1$ for the LCA muscle, $i=2$ for the IA muscle, $i=3$ for the PCA muscle, $i=4$ for the CT muscle, $i=5$ for the combined TA muscles, $i=6$ for the vocal ligament, and $i=7$ for the vocal fold mucosa). Since CT is not acting directly on the arytenoid cartilage, $F_4=0$ in Eq. (6), but becomes nonzero for vocal fold elongation to be discussed later.

In a similar fashion, the equations of motion for "effective" y displacements and rotation of the arytenoid cartilage are written as

$$\ddot{\psi}_a = \frac{1}{M_a}\left[\sum_{i=1}^{7} \beta_i F_i - k_y \psi_a - d_y \dot{\psi}_a\right], \tag{7}$$

TABLE I. Direction cosines and directional moment arms for canine CAJ dynamics (averaged over 8 larynges, left and right side). After Cox *et al.* (1999), and Mineck *et al.* (2000). γ is positive for counterclockwise "equivalent rotation" of the right arytenoid cartilage.

Muscle	α	β	γ (mm)
PCA	−0.1	−0.8	−5.49
LCA	−0.198	0.886	3.915
TA	0.015	0.990	0.8
IA	−0.697	−0.644	−3.30

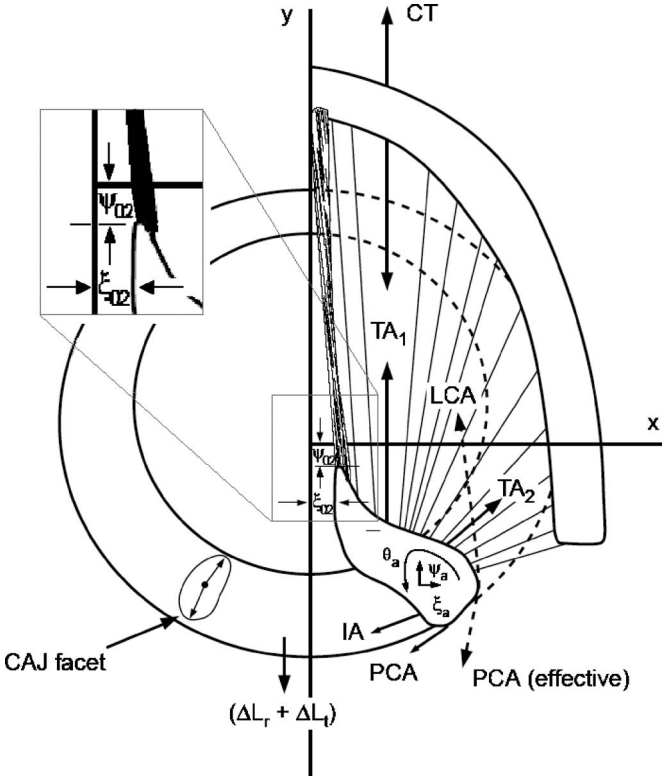


FIG. 3. Forces acting on the arytenoid cartilages for cricoarytenoid movement. Upper left insert shows expanded view of vocal process.

$$\ddot{\theta} = \frac{1}{I_a}\left[\sum_{i=1}^{7} \gamma_i F_i - \kappa\theta_a - d_r\dot{\theta}\right], \tag{8}$$

where ψ_a is the y displacement and θ_a is the rotation around the cricoarytenoid joint, β_i is the ith directional cosine for y displacement (Table I), γ_i is the ith directional moment arm from Table I, k_y and κ are respective translational and rotational stiffnesses, I_a is the moment of inertia of the arytenoid cartilage, and d_r is the rotational damping coefficient. These inertial, elastic, and viscous damping constants of the CAJ are largely nonlinear and have been determined empirically in various excised larynx experiments (Titze, 2006).

B. Equations of motion for cricothyroid joint movement

Dynamical equations for vocal fold elongation have been published previously (Titze *et al.*, 1988; 2006) and are reviewed here only for completeness. For a rotational strain ε_r of the cricothyroid joint (CTJ), which is defined as vocal fold elongation normalized to a resting vocal fold length L_0,

$$\ddot{\varepsilon}_r = \frac{h}{L_0 I_r}\left[wF_4 - hF_5 k_r \frac{L_0}{h}\left(\varepsilon_r + t_r\dot{\varepsilon}_r\right)\right], \tag{9}$$

where I_r is the moment of inertia associated with the CTJ rotation, F_4 and F_5 are the CT and TA forces, respectively, as defined earlier, w is the moment arm for the CT torque, h is the moment arm for the TA torque, and t_r is the rotational time constant for viscous damping, defined as a damping coefficient d_r divided by the rotational stiffness k_r (Titze, 2006). Similarly, the second-order equation for translational strain ε_t around the CTJ joint was found to be

$$\ddot{\varepsilon}_t = \frac{1}{M_t L_0}\left[F_4 \cos\phi - F_5 - k_5 L_0\left(\varepsilon_t + t_t\dot{\varepsilon}\right)\right], \tag{10}$$

where M_t is the mass associated with CTJ translation and t_t is the translational time constant for viscous damping.

All the forces F_i in Eqs. (6)–(8) and forces F_4 and F_5 in Eqs. (9) and (10) are governed by two constitutive first-order equations based on the traditional Kelvin model (Fung, 1981),

$$F_i + t_{si}\dot{F}_i = A_i[\sigma_i + E_i(\varepsilon_t + t_{pi}\dot{\varepsilon}_t)], \tag{11}$$

$$\sigma_i + t_{ai}\dot{\sigma}_i = a_i\sigma_{ai}, \tag{12}$$

where A_i is the cross-sectional area of the muscle, E_i is Young's modulus, ε_i is the strain in the muscle, t_{si} is a series time constant, t_{pi} is a parallel time constant, σ_i is the active internal stress, t_{ai} is the internal activation time constant, a_i is the normalized activity level (ranging between 0.0 and 1.0), and σ_{ai} is the active isometric stress (Titze, 2006).

The total strain for vocal fold elongation (or TA elongation) becomes the sum of an adductory strain ε_a (due to arytenoid cartilage motion on the cricoid cartilage), the rotational strain ε_r, and the translational vocal fold strain ε_t,

$$\varepsilon = \varepsilon_a + \varepsilon_r + \varepsilon_t, \tag{13}$$

where ε_a is the adductory strain yet to be determined.

C. Determination of the adductory strain

With reference to Fig. 3, the x axis is placed at the tip of the vocal process in the cadaveric position, and the y axis is placed at the midsagittal plane in the glottis. If we define x_{CAJ} and y_{CAJ} to be the coordinates of the center of the CAJ (as measured from the origin of our coordinate system), and if we define $\bar{x}_{02}$ to be the cadaveric (resting) x position of the vocal process, then the x displacement of the vocal process tip (ξ_{02}) is geometrically defined as

$$\xi_{02} = x_{\mathrm{CAJ}} - (x_{\mathrm{CAJ}} - \bar{x}_{02})\cos\theta_a + y_{\mathrm{CAJ}}\sin\theta_a + \xi_a \tag{14}$$

and the y displacement of the vocal process tip (ψ_{02}) is

$$\psi_{02} = y_{\mathrm{CAJ}} - y_{\mathrm{CAJ}}\cos\theta_a - (x_{\mathrm{CAJ}} - \bar{x}_{02})\sin\theta_a + \psi_a + L_0(\epsilon_r + \epsilon_t). \tag{15}$$

It is now possible to define the adductory strain that couples the differential equations in the foregoing sections. Recalling that the vocal process rests at $y=0$ for the cadaveric position, the total strain is

$$\varepsilon = -\frac{\psi_{02}}{L_0} = \varepsilon_a + \varepsilon_r + \varepsilon_t. \tag{16}$$

The adductory strain ε_a is now obtained from Eq. (16) by substituting ψ_{02} from Eq. (15),

$$\varepsilon_a = -\frac{1}{L_0}[y_{\mathrm{CAJ}}(1 - \cos\theta_a) - (x_{\mathrm{CAJ}} - \bar{x}_{02})\sin\theta_a + \psi_a]. \tag{17}$$

This adductory strain couples the differential equations of motion for adduction and elongation. Since the stress in every tissue of the vocal fold is affected by the overall vocal fold strain ε, this coupling can be very important in vocal fold posturing.

D. Solution of the combined differential equations

The following is a summary of how the posturing problem is solved:

(1) Define as fourth-order Runge-Kutta dependent variables the forces F_i and active stresses σ_i [Eqs. (11) and (12)] of all five intrinsic muscles, the vocal ligament, and the vocal fold mucosa (12 variables in all; the ligament and mucosa do not have active stresses).
(2) Further define as Runge-Kutta dependent variables ε_r, $\dot{\varepsilon}_r$, ε_t, $\dot{\varepsilon}_t$, ξ_a, $\dot{\xi}_a$, ψ_a, $\dot{\psi}_a$, θ_a, $\dot{\theta}_a$ (10 variables).
(3) Define the derivatives of all of the above-mentioned 22 variables as further independent variables, leading to a total of 44 first-order differential equations to be solved simultaneously.
(4) Define all the mechanical constants as outlined in the text and in Titze (2006).
(5) Define muscle activations as inputs (constant or time-varying).
(6) Inside the Runge-Kutta loop, calculate all strains and strain rates, the active and passive forces of all muscles,

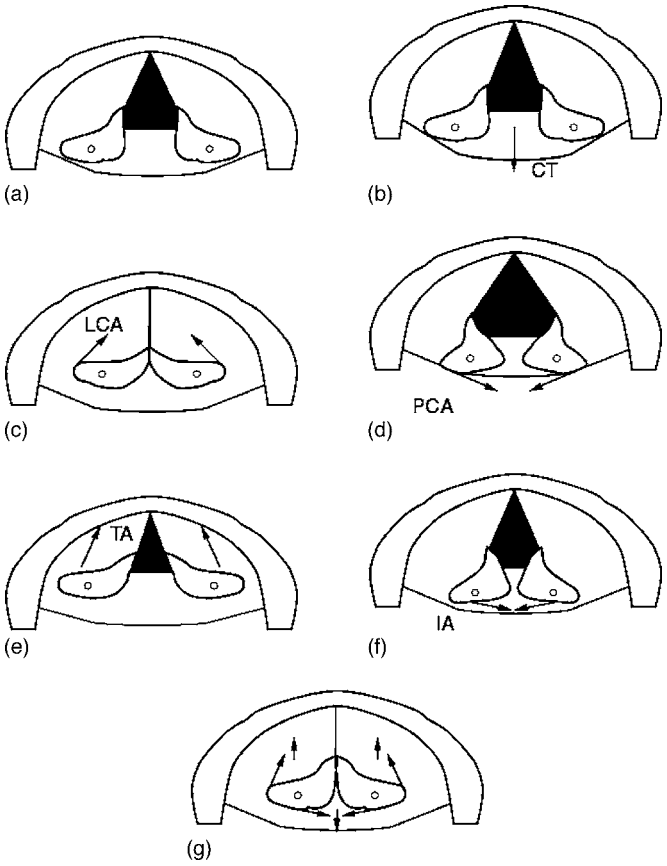


FIG. 4. Two-dimensional arytenoid cartilage posturing with specified muscle activations, (a) cadaveric, (b) 100% CT, (c) 100% LCA, (e) 100% TA, (f) 100% IA, and (g) 20% CT, 50% LCA, 0% PCA, 20% TA, 90% IA.

the passive forces of the ligament and mucosa, the stiffnesses, and the first-order derivatives as defined in step 3.

(7) Plot the time course of the displacement as desired.

III. RESULTS

A pictorial representation of simulated "effective" 2D posturing is shown in Fig. 4. Seven pictures are shown. For each picture, the outside semicircular border is the thyroid cartilage and the paired objects on the inside are the arytenoid cartilages. The black portion is the glottis. The upper left picture is for the cadaveric position (no muscle forces applied). The next five pictures are for 100% contraction of each of the five intrinsic muscles of the larynx in isolation, with the line of action of the muscle force shown by arrows. The bottom picture is for a coordinated vocal fold adduction that involves several muscles simultaneously. Note the CT action pulls the arytenoid cartilages backwards, LCA action effectively rotates the tips of the vocal processes toward the glottal midline, PCA action creates the opposite rotation, TA action shortens and adducts the vocal folds, and IA action adducts the posterior glottis.

Corresponding time variations for the vocal process of the arytenoid cartilage are shown in Fig. 5. Figure 5(a) shows the variation of ξ_{02} and Fig. 5(b) shows the variation of ψ_{02}. Recall that these are the x and y displacements, respectively, of the tips of the vocal processes of the arytenoid

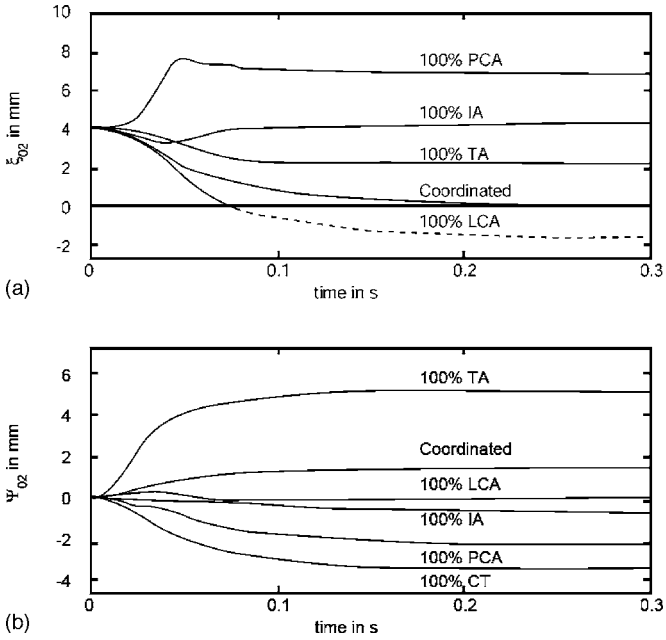


FIG. 5. Time course of posturing variables, (a) adduction variable ξ_{02} and (b) elongation variable ψ_{02}.

cartilages. Response time from cadaveric position to maximum glottal opening with 100% PCA is 50 ms. Response time from cadaveric position to glottal closure ($\xi_{02}=0$) with 100% LCA is 75 ms.

Figure 6 shows measured EMG data from Poletto *et al.* (2004) for the utterance $|hi\text{-}hi\text{-}hi\text{-}hi\text{-}hi|$. In the top trace we see the microphone signal, followed by the right PCA activity (normalized to 1.0), the right CT activity, the left TA activity, the left LCA activity, and the approximation angle (also normalized to 1.0). Note that the peak of the PCA activity occurs at the beginning of voice offset and that the peaks of TA and LCA activities occur just prior to voice onset. Thus, the adductors TA and LCA toggle with the abductor PCA to produce voicing and devoicing, roughly in equal intervals. For this subject, variation in LCA was less significant than variation in TA. CT activity correlated roughly with PCA activity, suggesting that CT is a mild abductor for devoicing.

The four EMG signals were introduced as inputs to the 2D posturing model [the a_i values in Eq. (12)] in an attempt to simulate the utterance $|hi\text{-}hi\text{-}hi\text{-}hi\text{-}hi|$. Computer outputs are shown in Fig. 7. On the left panel, from top to bottom, we show the vocal tract configuration for the vowel $|i|$, followed by vocal fold contact area *ca*, glottal area *ga*, glottal flow *ug*, and glottal flow derivative *dug*. On the right panel, again from top to bottom, we see oral output pressure *Po* (the simulated microphone signal corresponding to the top trace of Fig. 6), pressure in the mouth *Pm* (just behind the lips), pressure input to the vocal tract at the epilarynx tube *Pe*, intraglottal pressure *Pg*, and subglottal pressure *Ps*. In the Poletto *et al.* (2004) data set, no subglottal pressure was reported. We assumed a lung pressure that varied from 1.0 kPa at the beginning of the utterance to 1.5 kPa at the end. This was necessary to maintain voicing in light of a slightly overall rising of PCA from beginning to end and a

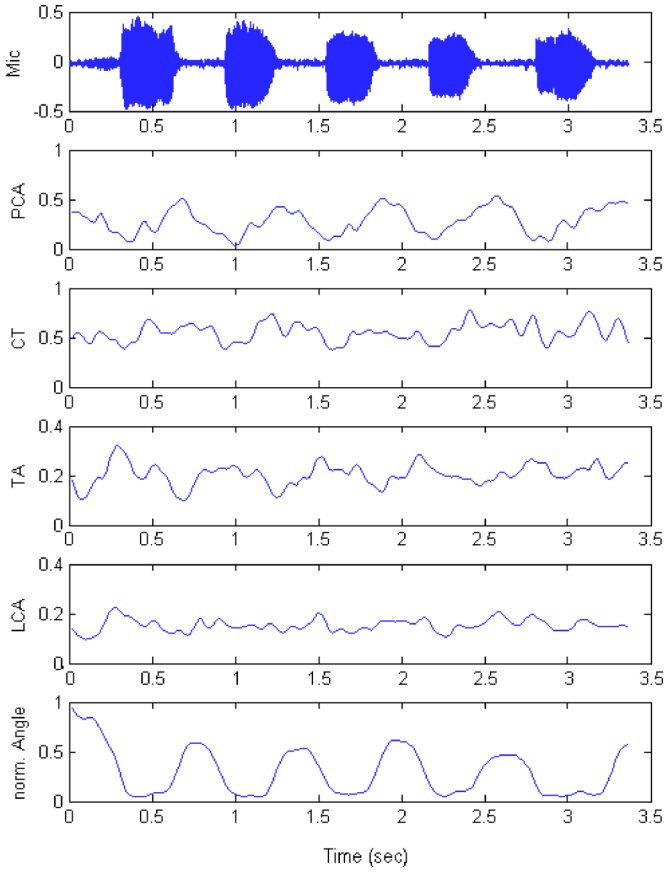


FIG. 6. (Color online) (Following Poletto *et al.* 2004). Data from production of |*hi-hi-hi-hi-hi*|. The head mounted microphone signal is shown in the top row. The following four rows are the rectified and smoothed EMG signals for the four muscles: PCA, CT, TA, and LCA. Normalized glottal angle is shown in the bottom row.

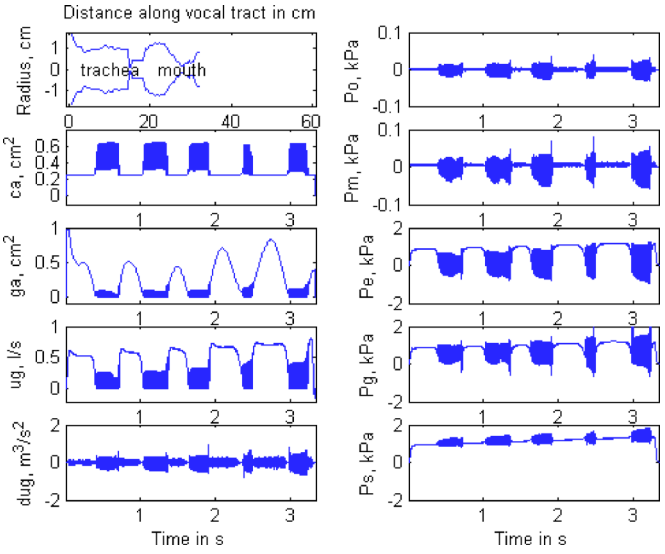


FIG. 7. (Color online) Wave form outputs of the simulator. Left column: Vocal tract outline, contact area, glottal area *ga*, flow *ug*, flow derivative dug. Right column: Pressure out *Po*, pressure behind lips *Pm*, pressure at epiglottis *Pe*, pressure at glottis *Pg*, and subglottal pressure *Ps*. The upper right is the pressure that can be played back as simulated voice and compared to row 1 of Fig. 6.

slight falling of TA from beginning to end. Also, no IA activity was reported by Poletto *et al.* (2004). We held IA activity constant at 0.35 (35% of maximum) to close the posterior gap.

Note that the simulated glottal area *ga* in Fig. 7 (left panel, in the middle) has a peak value ranging from 0.5 to 0.8 cm^2 across the four devoicing periods. Knowing that in our simulation the glottis was triangular and that the glottal length was 1.0 cm, the mean glottal angle at maximum opening ranged from 60° to 80°, comparable to the 60° angle measured by Scherer (1995). Using high-speed endoscopy, Hunter (2001) showed the maximum glottal angle in a repeated *sniff*-/i/ task to be 51±20° for three subjects. In three studies of the range of motion of the arytenoid [reviewed in Hunter and Titze (2005)], maximum glottal angle (using a 1 cm glottis) can be calculated to be 56±2°. Also in Hunter and Titze, two other models of vocal fold posturing were discussed to underestimate arytenoid motion (a maximum glottal angle of about 33°). Thus, the data presented in this paper are within the range described in the literature, although on the high end.

IV. CONCLUSION

To our knowledge, this study was the first successful attempt to simulate a speech-like utterance with EMG signals as the sole inputs to a biomechanical vocal fold model. The simulation reproduced the dynamic posturing gestures for voicing and devoicing in the utterance |*hi-hi-hi-hi-hi*|. Both the maximum excursion of the vocal process and the time course of adduction-abduction agreed reasonably well with measured data. The internal mechanical properties of laryngeal muscles (contraction times, maximum active stress, shortening velocity, etc.) were solely responsible for this time course and extent of movement. But it must be reiterated that biomechanical data sets from both human and canine species were used, making the results applicable to humans only to the extent that the canine model is accurate.

We conclude that a 2D model of vocal fold posturing, based entirely on the mechanics of "effective rotation and translation" of cricothyroid and cricoarytenoid joints and contraction of five intrinsic laryngeal muscles, can be used to simulate the dynamics of adduction and abduction in speech. But it requires careful resolution of both force and displacement vectors from 3D to 2D with the use of direction cosines, given that the true motion is a rocking rather than gliding and rotating. Also, the "effective 2D viscoelastic parameters," such as spring constants and damping ratios, must be viewed as being unique to the 2D geometry. Further refinements of effective movement arms of the forces, joint stiffnesses, and maximum active forces in the muscles would be helpful to establish more precise ranges of motion and the exact response times for the posturing gestures. More important, future 3D investigations such as the one by Hunter *et al.* (2004) should address the simultaneity of the dynamics of length change (and corresponding F_0 variations) in conjunction with adductory-abductory posturing. This would ultimately make 2D approaches obsolete and make 3D models useful in speech simulation.

ACKNOWLEDGMENT

This work was supported by Grant No. RO1 DC04347 from the National Institute on Deafness and Other Communication Disorders.

Berry, D., Montequin, D., Titze, I., and Hoffman, H. (**2003**). "An investigation of cricoarytenoid joint mechanics using simulated muscle forces," J. Voice **17**, 47–62.

Cox, K., Alipour, F., and Titze, I. (**1999**). "Geometric structure of the human and canine cricothyroid and thyroarytenoid muscles for biomechanical applications," Ann. Otol. Rhinol. Laryngol. **108**, 1151–1158.

Boone, D., and McFarlane, S. C. (**1994**). *The Voice and Voice Therapy*, 5th ed. (Prentice Hall, Englewood Cliffs, NJ), Chap. 66.

Cooke, A., Ludlow, C. L., Hallett, N., and Selbie, W. S. (**1997**). "Characteristics of vocal fold adduction related to voice onsets," J. Voice **11**, 12–22.

Frable, M. A. (**1961**). "Computation of motion at the cricoarytenoid joint," Arch. Otolaryngol. **73**, 73–78.

Fung, Y. C. (**1981**). *Biomechanics: Mechanical Properties of Living Tissues* (Springer, New York).

Gallena, S., Smith, P. J., Zeffiro, T., and Ludlow, C. L. (**2001**). "Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease," J. Speech Lang. Hear. Res. **44**, 1284–1299.

Hirano, M., Vennard, W., and Ohala, J. (**1970**). "Regulation of register, pitch and intensity of voice: An electromyographic investigation of intrinsic laryngeal muscles," Folia Phoniatr (Basel) **22**, 1–20.

Honda, K. (**1983**). "Variability analysis of laryngeal muscle activities," in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (The Denver Center for the Performing Arts, Denver, CO), pp. 127–137.

Hunter, E. J. (**2001**). "Three-dimensional biomechanical model of vocal fold posturing," Ph.D. dissertation, University of Iowa.

Hunter, E. J., Titze, I. R., and Alipour, F. (**2004**). "A three-dimensional model of vocal fold adduction/abduction," J. Acoust. Soc. Am. **115**, 1747–1759.

Hunter, E. J., and Titze, I. R. (**2005**). "Review of range of arytenoid cartilage motion," ARLO **6**, 112–117.

Maue, W. M., and Dickson, D. R. (**1971**). "Cartilages and ligaments of the adult human larynx,"Arch. Otolaryngol. **94**, 432–439.

Miller, R. (**1994**). "The mechanics of singing: Coordinating physiology and acoustics in singing," in *Vocal Arts Medicine: The Care and Prevention of Professional Voice Disorders*, edited by M. Benninger, B. Jacobson, and A. Johnson (Thieme, New York), Chap. 62.

Mineck, C. W., Tayama, N., Chan, R., and Titze, I. R. (**2000**). "Three-dimensional anatomic characterization of the canine laryngeal abductor and adductor musculature," Ann. Otol. Rhinol. Laryngol. **109**, 505–513.

Murry, T., Xu, J. J., and Woodson, G. E. (**1998**). "Glottal configuration associated with fundamental frequency and vocal register," J. Voice **12**, 44–49.

Peters, H. F., Boves, L., and van Dielen, I. C. (**1986**). "Perceptual judgment of abruptness of voice onset in vowels as a function of the amplitude envelope," J. Speech Hear Disord. **51**, 299–308.

Poletto, C. J., Verdun, L. P., Strominger, R., and Ludlow, C. L. (**2004**). "Correspondence between laryngeal vocal fold movement and muscle activity during speech and nonspeech gestures," J. Appl. Physiol. **97**, 858–866.

Selbie, W. S., Zhang, L., Levine, W. S., and Ludlow, C. L. (**1998**). "Using joint geometry to determine the motion of the cricoarytenoid joint," J. Acoust. Soc. Am. **103**, 1115–1127.

Selbie, W. S., Gewalt, S. L., and Ludlow, C. L. (**2002**). "Developing an anatomical model of the human laryngeal cartilages from magnetic resonance imaging," J. Acoust. Soc. Am. **112**, 1077–1090.

Sellars, I. E., and Keen, E. N. (**1978**). "The anatomy and movements of the cricoarytenoid joint," Laryngoscope **88**, 667–674.

Scherer, R. (**1995**). "Laryngeal function during phonation," in *Diagnosis and Treatment of Voice Disorders*, edited by J. Rubin, R. Sataloff, G. Korovin, and W. Gould (IGAKU-SHOIN, New York), pp. 93–104.

Titze, I. R., Jiang, J., and Druker, D. G. (**1988**). "Preliminaries to the body-cover model of pitch control," J. Voice **1**, 314–319.

Titze, I. R., and Sundberg, J. (**1992**). "Vocal intensity in speakers and singers," J. Acoust. Soc. Am. **91**, 2936–2946.

Titze, I. R. (**2006**). *The Myoelastic-Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Denver, CO).

Tomori, Z., Benacka, R., and Donic, V. (**1998**). "Mechanisms and clinico-physiological implications of the sniff- and gasp-like aspiration reflex," Respir. Physiol. **114**, 83–98.

Von Leden, H., and Moore, P. (**1961**). "The mechanics of the cricoarytenoid joint," Arch. Otolaryngol. **73**, 63–72.

Werner-Kukuk, E., and von Leden, H. (**1970**). "Vocal initiation: High speed cinematographic studies on normal subjects," Folia Phoniatr (Basel) **22**, 107–116.

Yoshioka, H. (**1981**). "Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese," Phonetica **38**, pp. 236–251.

Morphological predictability and acoustic duration of interfixes in Dutch compounds

Victor Kuperman[a] and Mark Pluymaekers
Radboud University Nijmegen, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

Mirjam Ernestus and Harald Baayen
Radboud University Nijmegen and Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

(Received 7 July 2006; revised 9 January 2007; accepted 16 January 2007)

This study explores the effects of informational redundancy, as carried by a word's morphological paradigmatic structure, on acoustic duration in read aloud speech. The hypothesis that the more predictable a linguistic unit is, the less salient its realization, was tested on the basis of the acoustic duration of interfixes in Dutch compounds in two datasets: One for the interfix *-s-* (1155 tokens) and one for the interfix *-e(n)-* (742 tokens). Both datasets show that the more probable the interfix is, given the compound and its constituents, the *longer* it is realized. These findings run counter to the predictions of information-theoretical approaches and can be resolved by the Paradigmatic Signal Enhancement Hypothesis. This hypothesis argues that whenever selection of an element from alternatives is probabilistic, the element's duration is predicted by the amount of paradigmatic support for the element: The most likely alternative in the paradigm of selection is realized longer. [DOI: 10.1121/1.2537393]

PACS number(s): 43.70.Bk, 43.70.Fq [AL] Pages: 2261–2271

I. INTRODUCTION

One of the organizing principles of speech production is the trade-off between economy of articulatory effort and discriminability of the speech signal (Lindblom, 1990). Speech communication often takes place in noisy conditions. In order to ensure robust recognition of their acoustic output, speakers need to invest effort in articulation. Yet clear and careful articulation is costly and hence tends to be dispensed efficiently (cf., Aylett and Turk, 2004; Hunnicutt, 1985). As a consequence, elements with low information load (or high predictability) have shorter or otherwise less salient realizations than relatively more informative elements of an utterance.

The informational redundancy of speech elements is often operationalized in terms of the probability (relative frequency of occurrence) of a linguistic unit (e.g., phoneme, syllable, word, or phrase) in its context. High probability has been observed to correlate with acoustic reduction in a large variety of language domains: Syntactic, discourse-related, phonological and prosodic, and lexical (e.g., Aylett and Turk, 2004; Bard *et al.*, 2000; Fowler and Housum, 1987; Jurafsky *et al.*, 2001; Lieberman, 1963; McAllister *et al.*, 1994; Pluymaekers, Ernestus and Baayen, 2005a; Pluymaekers, Ernestus and Baayen, 2005b; Samuel and Troicki, 1998; Scarborough, 2004; Van Son and Pols, 2003; Van Son and Van Santen, 2005). The attested types of reduction include—apart from widely reported durational shortening of syllables and individual phonemes—deletion of phonemes and complete syllables (e.g., Ernestus, 2000; Johnson, 2004), decrease in spectral center of gravity (Van Son and Pols, 2003), decrease in mean amplitude (Shields and Balota, 1991), higher degree of centralization of vowels (Munson and Solomon, 2004), and lower degree of coarticulation (Scarborough, 2004). The informational redundancy associated with a particular unit is a juxtaposition of the unit's probabilities given all relevant contexts. For instance, a word can be predictable because it has a high frequency, but also because it is frequently used with the word that precedes it. Both factors diminish the word's informativeness and both are expected to correlate with durational shortening.

The information-theoretical framework developed by Shannon (1948) has been used to explain the association between acoustic salience and informational redundancy. The efficiency of information transmission is optimal if the information in the signal is distributed equally, or smoothly, per time unit (e.g., Aylett and Turk, 2004; Aylett and Turk, 2006). When an important element is transmitted for a longer time, the probability of losing this element to noise decreases and the probability of the element being recognized correctly increases. This theoretical paradigm views acoustic duration as a means of smoothing the amount of information in the signal over time.

The present paper shows how the information carried by morphological paradigmatic structure modulates acoustic duration. Previous research (cf., Hay, 2003; Losiewicz, 1992) reported morphological effects on the acoustic duration of affixes in complex words. A related line of research demonstrated the influence of lexical neighborhood density on durational characteristics and coarticulation in speech production (e.g., Munson and Solomon, 2004; Scarborough, 2004; Vitevitch, 2002). The morphological objects that are central in the present study are interfixes in Dutch noun-noun com-

[a]Author to whom correspondence should be addressed. Electronic mail: victor.kuperman@mpi.nl

pounds. We will show that the acoustic duration of these interfixes creates an apparent paradox for the proposed information-theoretical principle of "less information, more reduction", which underlies the Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004), the Probabilistic Redundancy Hypothesis (Jurafsky *et al.*, 2001), and research on speech efficiency (e.g., Van Son and Pols, 2003). In our data, the more predictable the interfix is, the *longer* its articulation.

The distributional characteristics of the interfixes in Dutch compounds provide a clear-cut example of probabilistic, noncategorical morphological structure. Compounding is very productive in Dutch and is defined as the combination of two or more lexemes (or constituents) into a new lexeme (cf. Booij, 2002). In this paper we based our decisions of whether a given word is a compound and what its constituents are on the morphological parsing provided in the CELEX lexical database (Baayen, Piepenbrock and Gulikers, 1995). Compounds in Dutch can be realized with the interfix *-s-* (e.g., *oorlog-s-verklaring*, "announcement of war"), or with the interfix *-en-* (or its variant *-e-*) (e.g., *dier-en-arts* "veterinary"). Most compounds in Dutch, however, have no interfix (e.g., *oog-arts* "ophthalmologist"): For ease of exposition, we will henceforth refer to these latter words as compounds with the zero-interfix, or -∅-. In the frameworks that adopt deterministic rules, the distribution of interfixes in Dutch is enigmatic and inexplicable. Krott, Baayen and Schreuder (2001), however, have shown that the distribution of interfixes follows probabilistic principles defined over constituent families. The left (or right) constituent family of a compound is the set of all compounds which share the left (or right) constituent with this compound. For instance, the left constituent family of the compound *banknote* includes *bankbill, bankbook, bank-draft, bank-rate*, and *bankroll*. Krott, Baayen and Schreuder (2001), Krott *et al.* (2002) and Krott, Schreuder and Baayen (2002) show that the selection of the interfix is biased towards the interfix that is most commonly used with the given left constituent and, to a lesser extent, with the right constituent. Thus, besides having their own probability of occurrence, interfixes exhibit dependencies on larger morphological units both to the left and to the right. For this reason, interfixes serve as an appealing testing ground for studying the consequences of morphological predictability for acoustic realization.

The primary focus of the present study is the relationship between the predictability of the interfix given the morphological constituents of the compound, and its duration. We study the information-theoretical approach for two datasets with interfixed compounds and against the backdrop of multiple sources of redundancy, ranging from morphological to phonological and lexical information. Along the way, we replicate findings of laboratory studies of durational reduction for lively read-aloud speech.

II. METHODOLOGY

A. Materials

Acoustic materials were obtained from the Read Speech (or the "Library for the Blind") component of the Spoken Dutch Corpus (Oostdijk, 2000). Within this corpus of approximately 800 h of recorded speech, the Read Speech component comprises 100 h of recordings of written texts read aloud by speakers of Northern Dutch from the Netherlands and Southern Dutch from the Flanders area of Belgium. In the preparation of the recordings, speakers were pre-screened for the quality of their voice and clarity of pronunciation, and texts were made available to the speakers beforehand for preparatory reading. We chose to concentrate on read speech primarily because of the low level of background noise of the recordings. Quality was essential, since Automatic Speech Recognition (henceforth, ASR) was used for obtaining the segmental durations (see below). It should be noted that since these texts of fiction were read for the collection of the Library for the Blind, the reading style was a lively, rather than monotonous recitation, especially in the dialogs, where readers often mimicked casual speech.

Two datasets of Dutch noun-noun compounds were compiled: One with tokens containing the interfix *-s-* and one with compounds containing the interfix *-e(n)-*. Tokens in which the interfix *-s-* was either preceded or followed by the phonemes [s], [z] or [ʃ] were excluded from the dataset, since such an environment makes it difficult to reliably segment the interfix from its neighboring segments. The final dataset for the interfix *-s-* consisted of 1155 tokens. Similarly, tokens in which the second constituent begins with the segments [n] or [m] were taken out off the dataset of *-e(n)-* interfixes, resulting in a dataset of 742 tokens.

B. Measurements

Acoustic analysis of the selected tokens was performed using ASR technology. This was done for several reasons. First of all, the ASR technology allows to process a large volume of data in a relatively short time, which was important given the size of datasets used in this study. Moreover, it is possible to train an ASR device that bases its decisions purely on the characteristics of the acoustic signal, without reference to general linguistic knowledge. This is very difficult for human transcribers, who are bound to be influenced by expectations based on their knowledge of spelling, phonotactics, and so on (Cucchiarini, 1993). Second, ASR devices are perfectly consistent: Multiple analyses of the same acoustic signal always yield exactly the same result. Finally, the reliability of segmentations generated by an ASR system is equal to that of segmentations made by human transcribers (Vorstermans, Martens and Van Coile, 1996), provided that a phonemic transcription of the signal is available to the ASR algorithm.

For the present analysis, we utilized a Hidden Markov Model speech recognizer. This recognizer was trained using the software package HTK (Young *et al.*, 2002), which comprises 37 phone models representing the 36 phonemes of Dutch and silence, and uses for each model three-state HMMs with 32 Gaussians per state (Kessens and Strik, 2001). The HTK recognizer operates in two modes: If it is provided with the transcription of the speech recording, it determines segmental temporal boundaries; if no such transcription is provided, it identifies both the phonemes and the positions of their temporal boundaries. The accuracy of seg-

mentation is higher in the transcription-based mode. The sample rate of the HTK is 10 ms. The reliability of the ASR's segmentation with predefined transcriptions was established in a test in which the positions of phoneme boundaries placed by the ASR were compared to the positions of the same boundaries placed by a trained phonetician. The materials used for this test consisted of 189 words spoken in isolation. Comparison between the ASR-generated and manual segmentations revealed that, after postprocessing, 81% of the automatic boundaries were placed within 20 ms of the corresponding hand-coded boundaries. This level of accuracy is in accordance with international standards (Vorstermans *et al.*, 1996), and we considered it sufficient for present purposes.

Acoustic analysis proceeded as follows. First, the speech signal corresponding to the target compound was manually excised from its utterance context and parametrized using Mel Frequency Cepstral Coefficients. The parametrized signal was then supplied to a Viterbi segmentation algorithm, along with a phonemic transcription of the word. This transcription was taken from the CELEX lexical database. However, for words with the interfix *-e(n)-*, a cursory inspection of sound files established that many instances of this interfix were not realized as [ə] (the canonical pronunciation in CELEX), but rather as [ən]. An inspection of the sound files from the dataset with the interfix *-s-* revealed cases where the interfix was realized as [s] instead of the CELEX transcription [z] due to the regressive voice assimilation. Therefore, two trained phoneticians independently transcribed the realization of interfixes in both datasets. Initially, they disagreed on 10% of tokens from the *en* dataset and 13% of tokens from the *s* dataset. In both cases, they subsequently carried out a joint examination of the problematic tokens and came up with consensus transcriptions. The resulting transcriptions were provided to the segmentation algorithm, which estimated the boundaries of the phonemes in the acoustic signal. In this way, we obtained information about the durations of all segments for all words.

The acoustic duration of the whole interfix (henceforth, *InterfixDuration*) was taken as the main dependent variable in this study.

III. MORPHOLOGICAL VARIABLES

As shown in Krott *et al.* (2001), the more frequent an interfix is for the left constituent family of a compound, the more biased speakers are to use this interfix in that compound. The measures for this morphologically based bias will be at the center of our interest. They are defined as the ratio of the number of compounds where the left constituent is followed by *-s-*, *-e(n)-*, or *-∅-* respectively, and the total number of compounds with the given left constituent (henceforth, the left family size). To give an example, the Dutch noun *kandidaat* "candidate" appears as the left constituent in one compound with the interfix *-s-*, *kandidaat-s-examen* "bachelor's examination," in one compound with the interfix *-en-*, *kandidat-en-lijst* "list of candidates," and in one compound without an interfix, *kandidaat-stelling* "nomination." The type-based bias of this left constituent family towards the interfix *-s-* is $1/(1+2)=0.33$. The bias of the interfix *-e(n)-* has the value of $1/(1+2)=0.33$ as well, and so does the bias of the zero interfix. The measures of bias are labeled *TypeSBias*, *TypeEnBias* and *TypeZeroBias*.

Alternative, token-based, estimates of the bias are defined in terms of the frequencies of occurrence, rather than the type count of the compounds. The performance of token-based measures is consistently worse in our models than that of the type-based ones. Therefore, the token-based measures are not reported here. Furthermore, we only consider left constituent families, since the effect of the right bias is reported as either weak or absent (Krott, Schreuder and Baayen, 2002; Krott *et al.*, 2004).

The predictivity of constituent families for the duration of the interfix may extend beyond the bias measures, which only estimate the ratio of variants in the constituent family, without taking the magnitude (size, frequency, or information load) of the constituent family into account. However, these magnitudes are expected to exhibit effects in our analysis, since they repeatedly emerged as significant predictors in both the comprehension and production of Dutch compounds (e.g., Bien *et al.*, 2005; De Jong *et al.*, 2002; Krott *et al.*, 2004). To estimate the magnitude of constituent families, we incorporate in our study position-specific measures of entropy proposed by Moscoso del Prado Martín, Kostić and Baayen (2004). These measures employ the concept of Shannon's entropy (Shannon, 1948), which estimates the average amount of information in a system on the basis of the probability distribution of the members of that system. The probability of each member (p_{sys}) is approximated as the frequency of that member divided by the sum of the frequencies of all members. The entropy of a system with n members is then the negative weighted sum of log-transformed (base 2) probabilities of individual members:

$$H = -\sum_{i=1}^{n} p_{sys} * \log_2 p_{sys}.$$

Note that the entropy increases when the number of paradigm members is high (i.e., family size is large) and/or when the members are equiprobable.

Let us consider the positional entropy measure of the left constituent family of the Dutch noun *kandidaatstelling*. This family consists of three members: *kandidaatsexamen* has a lemma frequency of 22, *kandidaatstelling* has a lemma frequency of 15, and *kandidatenlijst* has a lemma frequency of 19 in the CELEX lexical database, which is based on a corpus of 42 million word forms. The cumulative frequency of this family is $22+15+19=56$, and the relative frequencies of these three family members are $22/56=0.39$ for *kandidaatsexamen*, $15/56=0.27$ for *kandidaatstelling*, and $19/56=0.34$ for *kandidatenlijst*. The left positional entropy of this constituent family therefore equals $-(0.39*\log_2 0.39+0.27*\log_2 0.27+0.34*\log_2 0.34)=1.57$ bit.

We consider the positional entropy measures for both the left and the right constituent families, henceforth *LeftPositionalEntropy* and *RightPositionalEntropy*, as potential predictors of the acoustic duration of the interfix. The informativeness of the right constituent family is meaningful as a

measure of the cost of planning the right constituent: Planning upcoming elements with a low information load has been shown to predict reduction in the fine phonetic detail of the currently produced elements (Pluymaekers *et al.*, 2005a).

IV. OTHER VARIABLES

Since acoustic duration is known to depend on a wide range of factors, we used stepwise multiple regression to bring these factors under statistical control. Two sets of factors were considered: Lexical frequency-based probabilities, and phonetic, phonological and sociolinguistic variables.

A. Probabilistic factors

Phrasal level: A higher likelihood of a word given its neighboring words has been shown to correlate with vowel reduction, segmental deletion, and durational shortening (Bell *et al.*, 2003; Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a). To quantify this likelihood, for each compound token in our data we calculated its mutual information with the preceding and the following word (*BackMutualInfo*, *FwdMutualInfo*) by using the following equation (X and Y either denote the previous word and the compound, or they denote the compound and the following word; XY denotes the combination of the two words):

$$MI(X;Y) = -\log \frac{\text{Frequency}(XY)}{\text{Frequency}(X) * \text{Frequency}(Y)}.$$

The measures were computed on the basis of the Spoken Dutch Corpus, which contains 9 million word tokens. All frequency measures were (natural) log transformed. Obviously, the values could not be computed for the instances where the target word was utterance-initial or utterance-final, respectively.

For those words for which mutual information with the preceding or the following word could be computed, we checked whether it was a significant predictor of the duration of the interfix over and beyond other factors. Neither *BackMutualInfo* nor *FwdMutualInfo* reached significance in our datasets. This result may originate in the properties of the datasets which comprise relatively low-frequency compounds. Obviously, these low-frequency compounds have even lower frequencies of co-occurrence with their neighboring words. For instance, for the *s* dataset the average frequency of co-occurrence of the compounds with the preceding word is a mere 1.63(SD=0.77), and with the following word a mere 1.20(SD=0.30). Another explanation may be that effects of contextual predictability do not extend to phonemes in the middle of long compounds. They may only emerge for segments at word boundaries (e.g., Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a).

Word level: The lexical frequency of a word is known to codetermine articulation and comprehension (e.g., Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a; Scarborough, Cortese and Scarborough, 1977; Zipf, 1929). Moreover, previous research has shown that whole word frequency robustly affects production and comprehension of compounds even in the low-frequency range (cf. e.g., Bertram and Hyönä, 2003; Bien *et al.*, 2005). Therefore we include the natural log-transformed compound frequency (*WordFrequency*) as a control variable in the analyses. Together with the measure of the bias and the left positional entropy, this variable forms a cluster of predictors that capture different aspects of the same phenomenon. The measure of the bias estimates the *proportion* of the positional family of compounds that supports the interfix. The corresponding entropy estimates the number and average information load of the members in this family, i.e., it gauges the reliability of the knowledge base for the bias. Finally, a high compound frequency quantifies the evidence for the co-occurrence of the left and right constituents with the interfix. We expect these variables to behave similarly in predicting the durational characteristics of the interfix.

Segmental level: Another dimension of predictability for segmental duration is the amount of lexical information in the individual segment given the preceding fragment of the word (i.e., given the "word onset"). Following Van Son and Pols (2003), we define an information-theoretical measure that quantifies segmental lexical information (*TokenSegmentalInfo*):

$$I_L = -\log_2 \frac{\text{Frequency}([\text{word onset}] + \text{target segment})}{\text{Frequency}([\text{word onset}] + \text{any segment})}$$

Van Son and Pols (2003) interpret this measure as estimating the segment's incremental contribution to word recognition. The occurrence of a segment that is improbable given the preceding fragment of the word limits the cohort of matching words substantially and thus facilitates recognition. To give an example, the amount of lexical information of the segment [s] given the preceding English word fragment [k a ʊ] is calculated as the negative log-transformed ratio of the cumulative frequency of words that begin with the string [k a ʊ s] (e.g., *cows, cowskin, cowslip, cowslips*) and the cumulative frequency of the words that begin with the string [k a ʊ] plus any segment (e.g., *cows, cowpat, cowshed, cowskin, cowslip, cowslip, etc.*) In the present study, segmental lexical information measures are based on the frequencies of single words, such as made available in CELEX, and do not account for combinations of words, even if those may acoustically be valid matches for the phonetic string. For instance, the combination *cow stopped* is not included in the calculation of the lexical information for the segment [s] in the string [k a ʊ s].

A positive correlation of this token-based segmental lexical information and segmental duration was reported in Van Son and Pols (2003) for different classes of phonemes grouped by manner of articulation: For read speech, the r values of correlations that reached significance ranged between 0.11 and 0.18 (55,811 df). If segmental lexical information indeed modulates fine phonetic detail, it is a potential predictor of the duration of the interfix.

To this token-based measure of segmental lexical information (*TokenSegmentalInfo*), we add a type-based measure, *TypeSegmentalInfo*, which is based on the *number* of words matching the relevant strings, rather than their cumulated frequencies:

$$S_L = -\log_2 \frac{\text{Number}([\text{word onset}] + \text{target segment})}{\text{Number}([\text{word onset}] + \text{any segment})}$$

We validated both the token-based and the type-based measures of segmental lexical information against our own dataset to establish how the performance of the type-based estimate S_L compares with that of the token-based measure I_L. Our approach differs from that of Van Son and Pols (2003) in that it considers the divergence of phonemes from their mean durations, rather than the raw durations of these phonemes. Different phonemes, even those that share manner of articulation, intrinsically differ in their durations. Therefore, pooling the durations of large classes of phonemes introduces unnecessary noise in the correlation analyses. We gauged the divergence of each instantiation of every phoneme from the mean duration of this phoneme and tested whether this divergence can be explained by the amount of lexical information carried by the phoneme. Our survey is based on *all* segments in the *s* dataset and in the compounds of the *en* dataset in which the interfix is realized as [ə].

We collected the data on mean durations from the Read Text component of the IFA corpus, a hand-aligned phonemically segmented speech database of Dutch (Van Son, Binnenpoorte, van den Heuvel and Pols, 2001). We log transformed the individual durations and computed the means and standard deviations of all tokens of each phoneme. Then, moving phoneme by phoneme through our compound dataset we calculated the z score for each phoneme, that is, the difference between its actual log-transformed duration and its mean log duration, in units of standard deviation from the mean. The correlation between the observed durational difference and the corresponding amount of type-based segmental lexical information yields an r value of $0.06(t(17{,}694)=7.41, p<0.0001)$. This order of magnitude is comparable with the results that Van Son and Pols (2003) obtained for the token-based measure of lexical information. The observed correlation is a rough estimate of the baseline effect that segmental lexical information may have on acoustic duration. The correlation is highly significant but the correlation coefficient is quite small. This is expected, given the multitude of phonetic, phonological, sociolinguistic and probabilistic factors that determine acoustic duration in speech production that are not taken into account here. As the type-based measure is predictive for durations of segments across the dataset, we decided to include it in our analyses of the interfix durations. Thus, we take as control variable the value of *TypeSegmentalInfo* for the (first) segment of the interfix.

Importantly, the durations show a weaker correlation with the token-based segmental lexical information, proposed by Van Son and Pols (2003) ($r=0.03, t(17{,}694)=4.25, p<0.0001$), than for its type-based counterpart ($r=0.06$). This measure also performs worse in the models reported below. Since the token- and type-based measures are highly correlated, we incorporated only *TypeSegmentalInfo* in our analysis.

B. Phonetic, phonological and sociolinguistic variables

Speech rate is an obvious predictor of acoustic duration (e.g., Crystal and House, 1990; Fosler-Lussier and Morgan, 1999; Pluymaekers *et al.*, 2005a). Two different measures estimating speech rate were included as control variables. First, we defined an utterance-based rate of speech, *SpeechRate*, as the number of syllables in the utterance divided by the acoustic duration of the utterance. Utterance is defined here as the longest stretch of speech containing the compound and not containing an audible pause.

Second, we defined a more local speech rate for the interfix *-s-*. In the *s* dataset, the interfix *-s-* always belongs to the coda of the preceding syllable. We measured the average segmental duration in the interfix-carrying syllable minus the *-s-* interfix, and considered it as an estimate of the local speed of articulation in the part of the syllable that precedes the interfix *-s-*, henceforth *SyllableSpeed*. The syllable from which the final segment [s] was subtracted is structurally complete, with an onset, a vowel, and (in 83% of tokens) a coda of one or more consonants. Note that for words with the interfix *-e(n)-* this measure of local speech rate is not meaningful. It would subtract the complete rhyme of the relevant syllable, leaving only the onset, the duration of which is above all determined by the number and types of its consonants.

Nooteboom (1972) observed that segments are shorter the greater the number of syllables or segments in the word. We therefore considered the total number of segments in the word, *NumberSegments*, and the number of segments following the interfix, *AfterSegments*.

We also took into account the sex, age, and language variety of the speaker (cf., Keune, Ernestus, Van Hout and Baayen, 2005). The binary variable *SpeakerLanguage* encodes the speaker's variant as Southern Dutch or Northern Dutch. If the information about age was missing, we filled in the average age of our speakers' population.

Prosody may affect the duration of segments as well. For instance, words at the beginning and the end of utterances show articulatory strengthening (e.g., Bell *et al.*, 2003; Cambier-Langeveld, 2000; Fougeron and Keating, 1997). To control for the word's position in the utterance, we coded each token with two binary variables *UtteranceInitial* and *UtteranceFinal*.

Furthermore, stressed syllables are pronounced longer than unstressed ones (e.g., Ladefoged, 1982). We coded each compound with the interfix *-s-* for whether its interfix-containing syllable carries a (primary or secondary) stress (the binary variable *Stressed*).

The interfix *-e(n)-* is never stressed. The common stress pattern for compounds with the interfix *-e(n)-* is for the primary stress to fall on the syllable immediately preceding the interfix-containing syllable, and the secondary stress on the syllable immediately following the interfix-containing syllable: The insertion of *-e(n)-* prevents a stress clash between the two constituents. The rhythmic structure of compounds has been proposed as a factor codetermining the selection of the interfix, in addition to lexical constituent families and

several other factors (Neijt *et al.*, 2002). To test the acoustic consequences of the rhythmic pattern, we coded each compound in the *en* dataset as to whether the interfix syllable intervenes between two immediately adjacent stressed syllables (binary variable *Clash*).

Compounds with the interfix *-e(n)-* were coded for the presence or absence of [n] in the acoustic realization of the interfix (*NPresent*), as established by two phoneticians (see Sec. II). Similarly, compounds with the interfix *-s-* were coded for whether the interfix was realized as [z], variable *PhonemeZ.*

Finally, the immediate phonetic environment can make a segment more or less prone to reduction. Unstressed vowels in Dutch tend to lengthen before oral stops (cf., Waals, 1999). Therefore, each compound in the dataset with the *-e(n)-* interfix was coded for the manner of articulation of the following segment (binary variable *FollowedbyStop*).

V. RESULTS

A. The interfix *-s-*

The dataset for the interfix *-s-* included 1155 tokens. The number of different word types was 680, and their token frequencies followed a Zipfian distribution ranging from 1 to 19. We fitted a stepwise multiple regression model with the acoustic duration of the interfix as the dependent variable. The values of this variable were (natural) log transformed to remove skewness of the distribution. The resulting variable *InterfixDuration* has a mean of 4.37 of log units of duration ($SD=0.35$). The log transformation in this model and the models reported below was applied purely for statistical reasons, such as reducing the likelihood that the estimates of the coefficients are distorted by atypically influential outliers. The coefficients of the regression models that are presented here in log units of duration can easily be converted back into milliseconds by applying the exponential function e^F to the fitted values (F) of the model.

We identified 21 data points that fell outside the range of −2.5 to 2.5 units of *SD* of the residual error, or had Cook's distances exceeding 0.2. These outliers were removed from the dataset and the model was refitted. Below we only report variables that reached significance in the final model.

The strength of the bias for the *-s-* interfix, *TypeSBias*, emerged as a main effect with a positive slope: Surprisingly, the duration of *-s-* was longer for compounds with a greater bias for this interfix [$\hat{\beta}=0.35, t(1125)=5.20, p<0.0001$]. A positive correlation with duration was present for the predictor *RightPositionalEntropy* as well [$\hat{\beta}=0.07, t(1125)=4.10, p<0.0001$], indicating that the duration of the interfix increases with the informational complexity of the right constituent. These main effects were modulated by an interaction between *TypeSBias* and *RightPositionalEntropy* [$\hat{\beta}=-0.07, t(112.5)=-3.67, p=0.0003$]. Inspection of conditioning plots revealed that the influence of the bias measure was greater when the value of the right positional entropy was low. In addition, *WordFrequency* had an unexpected positive slope that just failed to reach significance: [$\hat{\beta}=0.01, t(1125)=1.95, p=0.0510$]. We found no effect of the *LeftPositionalEntropy*.

Importantly, the lexical segmental information of the interfix was predictive in the expected direction: Segments conveying more information tended to be longer [*TypeSegmentalInfo*: $\hat{\beta}=0.12, t(1125)=3.86, p<0.0001$].

Among the phonological and phonetic variables, the measure of the speech rate also demonstrated the expected behavior. The greater the local speed of articulation, the shorter the realization of this interfix [*SyllableSpeed*: $\hat{\beta}=-0.51, t(1125)=-5.27, p<0.0001$]. Whether the interfix-carrying syllable was stressed was a significant predictor as well, with stress predicting durational shortening of the interfix [*Stressed*: $\hat{\beta}=-0.09, t(1125)=-3.96, p<0.0001$]. Finally, interfixes realized as [z] were shorter than those realized as [s], as expected given the findings by, for instance, Slis and Cohen (1969) [*PhonemeZ*: $\hat{\beta}=-0.16$, $t(1125)=-3.17, p=0.0016$].

All significant predictors were tested for possible nonlinearities; none reached significance. The bootstrap validated R^2 of the model was 0.104. The unique contribution of the morpholexical factors *TypeSBias, RightPositionalEntropyRight*, and *WordFrequency* to the explained variance over and above the other predictors was 2.0%, as indicated by the drop in R^2 when these variables were removed from the model.

B. Discussion

Three related morpholexical variables emerge as significant predictors of the duration of the interfix: *TypeSBias, RightPositionalEntropy*, and (marginally) *WordFrequency*. The positive correlations of *TypeSBias* and *WordFrequency* with the duration of the interfix lead to the paradoxical conclusion that a greater likelihood for a linguistic unit may lead to a longer acoustic realization of that unit, contradicting the information-theoretical approach to the distribution of acoustic duration. We will address this issue in Sec. VI. General Discussion.

The interaction of the right positional entropy with the bias hints at planning processes at work. According to Pluymaekers *et al.* (2005b), the planning of upcoming linguistic elements may interfere with the planning and production of preceding elements. We interpret the right positional entropy measure as tapping into the costs of planning the right constituent. The observed interaction indicates that the bias allows greater durational lengthening of the interfix when planning the next constituent is easy.

In accordance with previous reports (e.g., Van Son and Pols, 2003), a high amount of lexical information carried by an individual segment (*TypeSegmentalInfo*) predicts the acoustic lengthening of this segment. In other words, segments with a larger contribution to the word's discriminability are produced with increased articulatory effort, and hence prolonged duration. This highlights the paradox with which we are confronted: Conventional measures, such as the segmental lexical information, behave as expected, while measures for the likelihood of the interfix exhibit exceptional behavior.

The effects of *TypeSegmentalInfo* and of *TypeSBias* may

appear to contradict each other: For the same segment [s], the former variable predicts acoustic reduction, while the higher bias correlates with acoustic lengthening. Yet the two variables operate independently on different levels: The level of morphological word structure for the bias, and the segmental level for the lexical information. In the model, their (opposite) effects are simply additive.

The position of the compound in the utterance did not affect the durational characteristics of the interfix significantly, which is in line with observations by Cambier-Langeveld (2000). Cambier-Langeveld argues that final lengthening in Dutch only applies to the last syllable in the word or, if the vowel in this last syllable is [ə], to the penultimate syllable. Thus, the interfix lies beyond the scope of this effect. Similarly, the interfix emerges as outside the domain of influence of initial lengthening.

Segments are typically longer in a stressed syllable. This may have gone hand in hand with compensatory shortening of the duration of the following *-s-*. Compensatory reduction of the *-s-* in the coda of a stressed syllable may therefore provide an explanation for the observed effect of *Stressed*. Alternatively, acoustic reduction of the interfix may have arisen from the fact that stress on the syllable preceding the interfix *-s-* correlates with a higher local speech rate, which we calculated as the number of segments in the syllable (minus *-s-*) divided by the total duration of the syllable (minus *-s-*). This finding may appear counterintuitive, but it derives from the following observation. It is true that stressed syllables in our dataset have longer realizations than unstressed ones [two-tailed *t*-test: $t(1097)=30.0$, $p<0.0001$], but more importantly, they consist of more segments [two-tailed *t*-test: $t(1146)=22$, $p<0.0001$]. The net effect is the greater speech rate at stressed syllables. To test whether the latter finding is idiosyncratic to our dataset, we computed the number of segments for each syllable in Dutch monomorphemic words using CELEX phonological transcriptions. Again, we found that stressed syllables contained more segments than unstressed ones (2.76 vs 2.17 segments per syllable, two-tailed *t*-test: $t(192{,}546)=208.8$, $p<0.0001$). This difference retained significance when the counts were corrected for ambisyllabicity. We conclude that a higher local speech rate may have contributed to the shortening of *-s-*interfixes that follow stressed syllables.

C. The interfix -*e*(*n*)-

The *en* dataset contained 742 tokens of compounds. The number of different word types equalled 305, and the Zipfian distribution of tokens per type ranged from 1 to 74. We log transformed the acoustic durations of the interfixes, which then had a mean of 4.065 log units of duration ($SD=0.420$). We fitted a stepwise multiple regression model to these durations. This time, 19 data points fell outside the range of −2.5 to 2.5 units of *SD* of the residual error or had Cook's distances exceeding 0.2. These outliers were removed from the dataset, and the model was refitted. Only predictors that reached significance are reported.

The morpholexical predictors performed as follows: A higher bias for the interfix *-e*(*n*)-, *TypeEnBias*, correlated with longer interfixes: $[\hat{\beta}=0.14, t(716)=5.39, p<0.0001]$. The positional entropy of the right constituent family also had a positive main effect $[\hat{\beta}=0.08, t(716)=4.56, p<0.0001]$. The interaction of these two variables was not significant ($p>0.4$). *LeftPositionalEntropy* and *WordFrequency* did not reach significance either ($p>0.1$).

As in the model for the interfix *-s-*, a higher amount of lexical information, as attested by *TypeSegmentalInfo* for the first segment of the interfix, correlated with longer articulation $[\hat{\beta}=0.07, t(716)=3.09, p=0.002]$. This effect is again in line with predictions of the information-theoretical approach.

The interfixes of 226 tokens (29%) in the dataset were realized as [ən], while 516 tokens were pronounced with [ə]. As expected, the presence of [n] in the interfix implied a substantial increase in the total duration of the interfix. The factor *NPresent* was the most influential predictor $[\hat{\beta}=0.71, t(716)=37.80, p<0.0001]$, and its unique contribution to the explained variance of this duration was 55%.

Two phonetic factors contributed to the duration of the interfix. Unsurprisingly, the interfix was shorter when the utterance-based speech rate was higher [*SpeechRate*: $\hat{\beta}=-0.04, t(716)=-4.17, p<0.0001$]. Factor *FollowedbyStop* also had an effect $[\hat{\beta}=0.23, t(716)=13.10, p<0.0001]$, which supports the observation by Waals (1999) that an unstressed vowel is pronounced longer before oral stops. It is worth noting that Waals' observation, which was made under thoroughly controlled laboratory conditions, is replicated here in more natural read aloud speech.

All significant predictors in the model were checked for nonlinearities, none of which reached significance. The bootstrap validated R^2 value for the model was 0.72. The unique contribution of the morphological predictors *TypeEnBias* and *RightPositionalEntropy* to the variance explained by the model was 2.3%, as indicated by the drop in R^2 after the removal of these variables from the model. This contribution is close to that provided by the morpholexical predictors in the *s* dataset (2.0%).

D. Discussion

The analysis of the *en* dataset replicates the unexpected direction of the influence of the morphologically determined redundancy that we reported for the dataset with the interfix *-s-*: We found again that higher values for the bias estimates correlate with a longer duration of the interfix. We will return to this role of the bias in Sec. VI.

The positive simple main effect of the right positional entropy supports the hypothesis of continuous planning of articulation, according to which the planning complexity of upcoming elements may modulate acoustic characteristics of preceding elements.

Given the dominant contribution of the variable *NPresent* to the explained variance, we set out to establish what factors affected the selection of the variant [ən] vs [ə]. The interfix *-e*(*n*)- is spelled as either *-e-* or *-en-*, depending on orthographic rules. Compounds spelled just with *-e-* are unlikely to be pronounced with [ən]. The subset of compounds spelled with *-en-* contains 653 tokens. We fitted a logistic

regression model that predicted the log odds of the selection of [ən] vs [ə] in this subset. The model uses the binomial link function and considers the presence of [n] in the realization of the interfix as a success, and its absence as a failure. The results demonstrate no effect of *TypeEnBias* on the selection of the phonetic variant ($p>0.5$). Apparently the realization of an extra phoneme in the interfix is independent of the morphological likelihood of the interfix. The presence of [n] was more likely when *WordFrequency* was high [$\hat{\beta}=0.63, p<0.0001$], *RightPositionalEntropy* was high [$\hat{\beta}=2.11, p<0.0001$], the speaker's language was Southern Dutch [$\hat{\beta}=1.37, p<0.0001$], the number of segments after the interfix, *AfterSegments*, was high [$\hat{\beta}=2.06, p<0.0001$], and a stress clash was attenuated [$\hat{\beta}=4.19, p<0.001$]. The likelihood of [n] was lower when *LeftPositionalEntropy* was high [$\hat{\beta}=-0.60, p<0.0001$].

In a second supplementary analysis, we investigated whether morpholexical factors are better predictors for acoustic duration if we consider the duration of [ə] as the dependent variable, rather than the duration of the whole interfix. In such a model, we expect the presence of [n] to exercise less influence and the morpholexical predictors to have greater explanatory value than in the model for the duration of the interfix as a whole. We fitted a stepwise multiple regression model to the data with the (natural) log-transformed acoustic duration of the phoneme [ə] in the interfix as the dependent variable. After removal of 25 outliers, the model was refitted against the remaining 717 datapoints.

In line with our expectations, we observe a decrease in the predictive power of *NPresent* to only 15% of the explained variance, while the share of morphological variables *TypeEnBias* and *RightPositionalEntropy*, which retain significance as predictors of acoustic lengthening, increases to 4.3% of the explained variance. We conclude that morphological structure code termines the acoustic characteristics of the interfix -*e*(*n*)- over and beyond major phonological and phonetic predictors.[1]

VI. GENERAL DISCUSSION

According to the information-theoretical approach to acoustic salience developed in the last decade, a higher likelihood of a linguistic unit is correlated with more acoustic reduction. The main finding of the present study is that the effect of morphologically determined probability on the duration of interfixes in Dutch compounds runs counter to this prediction. This pattern of results is especially puzzling, since our data also provide evidence *in favor* of the information-theoretical approach in the form of an effect of segmental lexical information. Thus, we do find that a higher probability of a segment given the preceding word fragment leads to more acoustic reduction.

The speakers in the Spoken Dutch Corpus read the compounds and thus received unambiguous visual information about the correct interfix. It is therefore remarkable that we nevertheless observed effects of morpholexical factors on the planning and implementation of speech production. We note, however, that the bias of the interfix as determined by the left constituent family is known to predict the speed of reading comprehension of novel and existing compounds (Krott, Hagoort and Baayen, 2004). We therefore expect the acoustic consequences of the bias to have a larger scope when visual cues to the appropriate morphemes are absent, as in spontaneous speech genres.

What may be the solution for the problem that the present data appear to pose for the information-theoretical framework? One explanation might be that morphological information has a fundamentally different status from other types of linguistic information, and is typically associated with careful articulation. However, this line of reasoning is refuted by research on prefixes and suffixes in English (e.g., Hay, 2003) and Dutch (e.g., Pluymaekers *et al.*, 2005a, Pluymaekers *et al.*, 2005b).

Another solution might refer to the fact that interfixes are homophonous with plural markers in Dutch (cf., *boek-en* "books" and the compound *boek-en-kast* "bookshelf"). The frequency of the plural word forms might codetermine the duration of the interfix and be confounded with the bias. This explanation, however, can be discarded on the following grounds. First, there was no consistency in the correlation between the frequency of plural nouns and the bias of the interfix across datasets. For the -*s*-dataset the correlation was positive [$r=0.12, t(1154)=4.24, p<0.0001$], while for the -*en*-dataset it was negative [$r=-0.28, t(740)=-8.15, p<0.0001$]. Second, the frequency of the plural homophonous forms did not reach significance when included as a covariate in the regression models for both datasets. Finally, previous work on German compounds by Koester, Gunter, Wagner and Friederici (2004) has shown that plural suffixes and interfixes may not be perfectly homophonous in terms of systematic fine phonetic detail: Compound constituents followed by an interfix are shorter and have a higher pitch than their stand-alone plural counterparts.

The hypothesis that we would like to offer as a solution for the present paradox is that fine phonetic detail in speech is governed by two orthogonal dimensions, a syntagmatic dimension and a paradigmatic dimension. The information-theoretical approach that underlies the Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004) and the Probabilistic Reduction Hypothesis (Jurafsky *et al.*, 2001), as well as research on speech efficiency (Van Son and Pols, 2003; Van Son and Van Santen, 2005), views information from the syntagmatic perspective by considering the probability of a linguistic unit in its phonetic, lexical, or syntactic context. These syntagmatic relationships are inherently sequential and govern the temporal distribution of information in the speech stream. For instance, the extent to which a segment contributes to the identification of the word *given the preceding word fragment* (Van Son and Pols, 2003) is a syntagmatic measure that is positively correlated with duration: The greater the contribution of the segment, the longer its acoustic implementation.

The syntagmatic measures proceed upon the premise that there is no (probabilistic) variation in the elements forming the word or the syntactic clause to be realized by the speaker. When the speaker wants to express the concept

book, there is no doubt that the element following [bʊ] is [k].

However, the identity of the elements is not always known with such certainty: The interfix in Dutch compounds is one such example. We label such elements "pockets of indeterminacy." Paradigmatic relations, here defined over constituent families, provide the probabilistic basis for resolving this indeterminacy. The bias measures quantify the extent of support provided by paradigmatics for the different interfixes available for selection: A greater support increases the likelihood of a given interfix. Our experimental results indicate that such a greater likelihood is paired with a longer acoustic realization. Moreover, we have shown that a higher frequency of a compound correlates with an increased chance of a more salient realization of the interfix -*e*(*n*)- as [ən], rather than [ə].

Whereas the syntagmatic dynamics of lexical disambiguation are intrinsically temporal, paradigmatic inference is a-temporal in nature. In the a-temporal domain of paradigmatic inference for positions of choice, a greater probability implies a broader empirical basis for selection of a given alternative, and comes with increased acoustic duration.

Importantly, paradigms as a source of support for alternatives for selection are not restricted to morphological structure: We consider paradigms in a general Saussurean sense, as sets of linguistic elements over which the operation of selection is defined (De Saussure, 1966).

The amount of evidence for the alternatives apparently determines the confidence with which an interfix is selected. That a lack of confidence may lead to a decrease in acoustic duration may be illustrated by an analogy: When producing case endings of German nouns, non-native speakers of German may hush up their realizations if they have doubts about the appropriate morpheme, but articulate the endings carefully and clearly if they are certain about which ending to choose. This example serves as an analogy only, and there is no implication that speakers make deliberate, conscious choices based on the morphological bias. The support measured as the bias is rather an estimate of the "naturalness" of the association between the available interfixes and the constituents of the compound.

Our hypothesis that paradigmatic inference for pockets of indeterminacy leads to longer (or otherwise more salient) realizations, henceforth the Paradigmatic Signal Enhancement Hypothesis, offers straightforward, testable predictions at various levels of linguistic structure. First consider the level of morphology. It is well known that English irregular verbs cluster into sets according to the kind of vocalic alternation that they exhibit in the past tense form (*keep/kept*, *run/ran*). The Paradigmatic Signal Enhancement Hypothesis predicts that a past-tense vowel—a pocket of indeterminacy—is realized with increased acoustic salience when the vocalic alternation is supported by a larger set of irregular verbs. Effects of paradigmatic gangs might even be found for the vowels of regular verbs (Albright and Hayes, 2003).

At the interface of morphology and phonology, we call attention to the phenomenon of final devoicing. In German and Dutch, a stem-final obstruent may alternate between voiced and voiceless, compare Dutch [hɔnt] *hond* ("dog") with [hɔndə] *honden* ("dogs"). Ernestus and Baayen (2003, 2004) have shown that this alternation, traditionally regarded as idiosyncratic, is affected by paradigmatic structures driven by the rhyme of the final syllable. In addition, they have shown that devoiced obstruents (e.g., the [t] of [hɔnt]) may carry residual traces of voicing, and that listeners are sensitive to these residual traces (Ernestus and Baayen, 2006). The Paradigmatic Signal Enhancement Hypothesis builds on these findings by predicting that greater paradigmatic support for voicing will correlate with enhanced acoustic salience of residual voicing in the devoiced obstruent.

Additional evidence for the Paradigmatic Signal Enhancement Hypothesis emerges from research on intrusive /r/ in New Zealand English (Hay and Maclagan, in press): The more likely speakers are to produce intrusive /r/ given a range of linguistic and sociolinguistic factors, the more salient its realization (as reflected in the degree of constriction).

Finally, the probabilistic dependencies between morphemes, such as exist between the interfix, the compound's left and right constituents, and the whole compound, challenge the fully decompositional theory of morphological encoding in speech production, developed by Levelt, Roelofs and Meyer (1999). According to this model, an abstract lemma representation provides access to a word's individual constituents. The planning for articulation of these individual constituents is fully encapsulated from all other morphemes and their paradigmatic relations. This model is challenged not only by the present findings, but also by those of Van Son and Pols (2003), Pluymaekers *et al.* (2005a), Pluymaekers *et al.* (2005b), Hay (2003), and Ernestus *et al.* (2006). What the present paper adds to this literature is the surprising observation that fine phonetic detail is not only determined by the properties of the word itself and its nearest phonological neighbors, but also by its morphological paradigmatic structure.

ACKNOWLEDGMENTS

This research was supported by the Netherlands Organization for Scientific Research (NWO) Grant No. 360-70-130 to the second, third and fourth author. We thank Anders Löfqvist and an anonymous reviewer for their helpful comments on an earlier version of this paper.

[1]If a compound is spelled with -*e*(*n*)-, it can be realized as [ən] or [ə] in speech. We have shown that a higher word frequency favors the presence of [n] in the realization of the interfix. Might it be the case that the realization of the interfix as [ə] is longer in a compound that is more often realized with [ən]? To check this possibility, we computed the percentage of tokens realized as [ən] for each -*e*(*n*)-compound. This percentage was not a significant predictor of acoustic duration of [ə] ($p > 0.05$). Thus we rule out an impact of the relative frequency of [ən] realization (more probable in read speech) on [ə]-realization (more probable in spontaneous speech).

Albright, A., and Hayes, B. (**2003**). "Rules vs. analogy in English past tenses: A computational/experimental study," Cognition **90**, 119–161.

Aylett, M., and Turk, A. (**2006**). "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang Speech **47**, 31–56.

Aylett, M., and Turk, A. (**2006**). "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllabic nuclei," J.

Acoust. Soc. Am. **119**, 3048–3058.
Baayen, R. H., Piepenbrock, R., and Gulikers, L. (**1995**). *The CELEX lexical database (CD-ROM)*, Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.
Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (**2000**). "Controlling the intelligibility of referring expressions in dialogue," J. Mem. Lang. **42**, 1–22.
Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., and Gildea, D. (**2003**). "Effects of disfluencies, predictability, and utterance position on word form variation in English conversation," J. Acoust. Soc. Am. **113**, 1001–1024.
Bertram, R., and Hyönä, Y. (**2003**). "The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds," J. Mem. Lang. **48**, 615–634.
Bien, H., Levelt, W., and Baayen, R. (**2005**). "Frequency effects in compound production," Proc. Natl. Acad. Sci. U.S.A. **102**, 17876–17881.
Booij, G. (**2002**). *The Morphology of Dutch* (Oxford University Press, Oxford).
Cambier-Langeveld, T. (**2000**). *Temporal Marking of Accents and Boundaries* (LOT, Amsterdam).
Crystal, T., and House, A. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech," J. Acoust. Soc. Am. **88**, 101–112.
Cucchiarini, C. (**1993**). *Phonetic Transcription: A Methodological and Empirical Study* (University of Nijmegen, Nijmegen).
De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and Baayen, R. H. (**2002**). "The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects," Brain Lang **81**, 555–567.
De Saussure, F. (**1966**). *Course in General Linguistics* (McGraw-Hill, New York).
Ernestus, M. (**2000**). *Voice Assimilation and Segment Reduction in Casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface* (LOT, Utrecht).
Ernestus, M., and Baayen, R. H. (**2003**). "Predicting the unpredictable: Interpreting neutralized segments in Dutch," Lang **79**, 5–38.
Ernestus, M., and Baayen, R. H. (**2004**). "Analogical effects in regular past tense production in Dutch," Linguistics **42**, 873–903.
Ernestus, M., and Baayen, R. H. (**2006**). "The functionality of incomplete neutralization in Dutch: The case of past-tense formation," in *Lab Phon 8*, edited by L. Goldstein, D. Whalen, and C. Best, 27–49 (Mouton de Gruyter, Berlin).
Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (**2006**). "Lexical frequency and voice assimilation," J. Acoust. Soc. Am. **120**, 1040–1051.
Fosler-Lussier, E., and Morgan, N. (**1999**). "Effects of speaking rate and word frequency on pronunciations in conversational speech," Speech Commun. **29**, 137–158.
Fougeron, C., and Keating, P. (**1997**). "Articulatory strengthening at the edges of prosodic domains," J. Acoust. Soc. Am. **101**, 3728–3740.
Fowler, C., and Housum, J. (**1987**). "Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction," J. Mem. Lang. **26**, 489–504.
Hay, J. (**2003**). *Causes and Consequences of Word Structure* (Routledge, New York).
Hay, J. and Maclagan, M. (in press). "Social and phonetic conditioners on the frequency and degree of "intrusive /r/" in New Zealand English," *Methods in Sociophonetics*, edited by D. Preston and N. Niedzielski.
Hunnicutt, S. (**1985**). "Intelligibility versus redundancy - conditions of dependency," Lang Speech **28**, 47–56.
Johnson, K. (**2004**). "Massive reduction in conversational American English," in *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th International symposium*, 29–54 (The National International Institute for Japanese Language, Tokyo, Japan).
Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper, 229–254 (Benjamins, Amsterdam).
Kessens, J., and Strik, H. (**2001**). "Lower WERs do not guarantee better transcriptions," in *Proc. Eurospeech-2001*, 1721–1724 (Aalborg, Denmark).
Keune, K., Ernestus, M., Van Hout, R., and Baayen, R. H. (**2005**). "Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok'," Corpus Ling. Ling. Theory **1**, 183–223.
Koester, D., Gunter, T. C., Wagner, S., and Friederici, A. D. (**2004**). "Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds," J. Cogn Neurosci. **16**, 1647–1668.
Krott, A., Baayen, R. H., and Schreuder, R. (**2001**). "Analogy in morphology: Modeling the choice of linking morphemes in Dutch," Linguistics, **39**, 51–93.
Krott, A., Hagoort, P., and Baayen, R. H. (**2004**). "Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds," Lang. Cognit. Processes **19**, 453–471.
Krott, A., Krebbers, L., Schreuder, R., and Baayen, R. H. (**2002**). "Semantic influence on linkers in Dutch noun-noun compounds," Folia Linguis. **36**, 7–22.
Krott, A., Schreuder, R., and Baayen, R. H. (**2002**). "Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies," Brain Lang **81**, 708–722.
Ladefoged, P. (**1982**). *A Course in Phonetics*, 2nd ed. (Hartcourt, Brace, Jovanovich, New York).
Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (**1999**). "A theory of lexical access in speech production," Behav. Brain Sci. **22**, 1–38.
Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal, 403–440 (Kluwer, Dordrecht).
Losiewicz, B. (**1992**). "The effect of frequency on linguistic morphology," Ph.D. thesis, University of Texas.
McAllister, J., Potts, A., Mason, K., and Marchant, G. (**1994**). "Word duration in monologue and dialog speech," Lang Speech **37**, 393–405.
Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (**2004**). "Putting the bits together: An information theoretical perspective on morphological processing," Cognition **94**, 1–18.
Munson, B., and Solomon, N. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.
Neijt, A., Krebbers, R., and Fikkert, P. (**2002**). "Rhythm and semantics in the selection of linking elements," in *Linguistics in the Netherlands 2002*, edited by H. Broekhuis and P. Fikkert, 117–127 (Benjamins, Amsterdam).
Nooteboom, S. G. (**1972**). *Production and Perception of Vowel Duration: A Study of the Durational Properties of Vowels in Dutch* (University of Utrecht, Utrecht).
Oostdijk, N. (**2000**). "The Spoken Dutch Corpus Project," The ELRA Newsletter **5**, 4–8.
Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005a**). "Articulatory planning is continuous and sensitive to informational redundancy," Phonetica **62**, 146–159.
Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005b**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **118**, 2561–2569.
Samuel, A., and Troicki, M. (**1998**). "Articulation quality is inversely related to redundancy when children or adults have verbal control," J. Mem. Lang. **39**, 175–194.
Scarborough, D. L., Cortese, C., and Scarborough, H. S. (**1977**). "Frequency and repetition effects in lexical memory," J. Exp. Psychol. Hum. Percept. Perform. **3**, 1–17.
Scarborough, R. (**2004**). "Degree of Coarticulation and Lexical Confusability," in *Proceedings of the 29th Meeting of the Berkeley Linguistic Society, February 14–17, 2003*.
Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.
Shields, L., and Balota, D. (**1991**). "Repetition and associative context effects in speech productior," Lang Speech **34**, 47–55.
Slis, I. H., and Cohen, A. (**1969**). "On the complex regulating the voiced-voiceless distinction. II," Lang Speech **12**, 137–155.
Van Son, R., Binnenpoorte, D., van den Heuvel, H., and Pols, L. (**2001**). "The IFA Corpus: a phonemically segmented Dutch Open Source speech database," in *Proc. Eurospeech-2001* (Aalborg, Denmark).
Van Son, R., and Pols, L. (**2003**). "Information structure and efficiency in speech production," in *Proc. Eurospeech-2003* (Geneva, Switzerland).
Van Son, R., and Van Santen, J. (**2005**). "Duration and spectral balance of intervocalic consonants: A case for efficient communication," Speech Commun. **47**, 100–123.
Vitevitch, M. S. (**2002**). "The influence of phonological similarity neighborhoods on speech production," J. Exp. Psychol. Learn. Mem. Cogn. **28**, 735–747.

Vorstermans, A., Martens, J., and Van Coile, B. (**1996**). "Automatic segmentation and labeling of multi-lingual speech data," Speech Commun. **19**, 271–293.
Waals, J. (**1999**). *An Experimental View of the Dutch Syllable* (Holland Academic Graphics, The Hague).
Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2002**). *The HTK Book 3.2* (Entropic, Cambridge).
Zipf, G. K. (**1929**). "Relative frequency as a determinant of phonetic change," Harv. Studies Classical Philol. **15**, 1–95.

Longitudinal developmental changes in spectral peaks of vowels produced by Japanese infants[a)]

Kentaro Ishizuka,[b)] Ryoko Mugitani, Hiroko Kato, and Shigeaki Amano
NTT Communication Science Laboratories, NTT Corporation, Hikaridai 2-4, Seikacho, Sourakugun, Kyoto, 619-0237, Japan.

(Received 25 January 2006; revised 8 January 2007; accepted 10 January 2007)

This paper describes a longitudinal analysis of the vowel development of two Japanese infants in terms of spectral resonant peaks. This study aims to investigate when and how the two infants become able to produce categorically separated vowels, and covers the ages of 4 to 60 months in order to provide detailed findings on the developmental process of speech production. The two lower spectral peaks were estimated from vowels extracted from natural spontaneous speech produced by the infants. Phoneme labeled and transcription-independent unlabeled data analyses were conducted. The labeled data analysis revealed longitudinal trends in the developmental change, which correspond to the articulation positions of the tongue and the rapid enlargement of the articulatory organs. In addition, the distribution of the two spectral peaks demonstrates the vowel space expansion that occurs with age. An unlabeled data analysis technique derived from the linear discriminant analysis method was introduced to measure the vowel space expansion quantitatively. It revealed that the infant's vowel space becomes similar to that of an adult in the early stages. In terms of both labeled and unlabeled properties, these results suggested that infants become capable of producing categorically separated vowels by 24 months.
 [DOI: 10.1121/1.2535806]

PACS number(s): 43.70.Ep [BHS] Pages: 2272–2282

I. INTRODUCTION

Children's speech has long been analyzed acoustically to reveal the speech development process. In particular, the developmental change in spectral resonant peaks such as formants has been intensively investigated, because these acoustic features reflect the acquisition of manners of articulation and the maturity of articulatory organs such as the length of the vocal tract, the descent of the larynx, and the size of the pharynx. To date, most systematic studies that have included acoustic analyses of children's vowels covered 3 to 18 years of age (e.g., Eguchi and Hirsh, 1969; Bennett, 1981; Hillenbrand *et al.*, 1995; Busby and Plant, 1995; Lee *et al.*, 1999; Whiteside, 2001). The results of these studies confirmed that the formant frequencies decrease linearly with age, presumably due to the maturation of the vocal tract length after 3 years.

However, from an anatomical point of view, while articulatory organs continue to mature gradually after 3 years of age, their maturation is more rapid before 3 years of age. Fitch and Giedd (1999) measured the vocal tract lengths of people aged 2–25 years using magnetic resonance imaging (MRI) and revealed a global tendency for there to be a linear increment in the vocal tract length with age. Vorperian *et al.* (2005) then investigated the anatomical restructuring of the vocal tract in more detail using MRI undertaken between the ages of 2 weeks and 6 years 9 months. Their analyses used a linear autoregressive model with a breakpoint, revealing that the vocal tract structures exhibit accelerated growth especially between birth and 18 months.

This acceleration in anatomical growth means that we can expect the acoustic features of speech to change greatly in the first few years. As regards the frequency spectra of infants less than 3 years old, several cross-sectional studies have been conducted that reveal the early development of formant frequencies and vowel space formation. For example, Gilbert (1973) revealed that there is a tendency for formant frequencies to decrease with increases in age from 14 to 84 months, whereas Gilbert *et al.* (1997) indicated that the formant values remain unchanged from 15 to 24 months of age, and decrease significantly from 24 to 36 months. In terms of vowel space expansion, Kent and Murray (1982) analyzed the resonant frequencies of infant vocalic utterances at 3, 6, and 9 months and showed that the ranges of the $F1$ and $F2$ values expand with age from 3 to 9 months without introducing phonemic aspects into the analysis. In terms of the categorical separation of vowels, Kuhl and Meltzoff (1996) indicated that the vowel categories become more separated in the vowel space from 12 to 20 weeks.

Although these studies shed light on the global process of the formant frequency decrement and vowel space expansion within a limited age range, speech development in the early stages has a wide interindividual variability (e.g., Ferguson, 1979; Ferguson and Farwell, 1975). This fact restricts any statistical analysis of cross-sectional data as described above in terms of dealing with the development process in detail, and stimulated the need for longitudinal data collec-

[a)]Preliminary results of this work were presented at the 9th European Conference on Speech Communication and Technology (Interspeech 2005 "Eurospeech") held in September 2005 in Lisbon, Portugal under the title "A longitudinal analysis of the spectral peaks of vowels for a Japanese infant."

[b)]Electronic mail: ishizuka@cslab.kecl.ntt.co.jp

tion and analysis. As regards longitudinal studies, Lieberman (1980) collected speech data produced by five infants aged from 4 to 60 months, and his analysis of formant frequency plots over time revealed the emergence of a well-developed vowel triangle. Buhr (1980) also provided a detailed analysis for 4 to 16 months. In addition, Bond *et al.* (1982) conducted a longitudinal analysis of a child's speech from 17 to 29 months and showed that vowel formants shift over time towards a precisely defined vowel space. In terms of gestural coordination, Goodell and Studdert-Kennedy (1993) also conducted a longitudinal study between 22 and 32 months.

These longitudinal studies helped to depict the long-term trend of the organization process in formant frequencies with detailed observations. However, these studies also have certain limitations either as regards the amount of speech data (Lieberman, 1980; Buhr, 1980; Bond *et al.*, 1982) or in the number of observation points (Goodell and Studdert-Kennedy, 1993). Lieberman (1980) and Buhr (1980) analyzed identical data, which allow the analysis of 33.75 vowels per week on the average. In addition, the periods of recorded data reported in Lieberman (1980) were different for the five infant subjects (i.e., 16–64 weeks, 38–69 weeks, 66–147 weeks, 69–162 weeks, and 125–169 weeks). The amount of data reported by Bond *et al.* (1982) was an average of 39.6 vowels per month, and they analyzed the data at 17, 19, 22, 26, and 29 months. Goodell and Studdert-Kennedy (1993) only analyzed the data at two observation points, i.e., 22 and 32 months. Such quantitative deficiencies confined the discussions to qualitative rather than quantitative aspects of speech development.

The application of labeled data to infants' speech production leads to another problem. Labeled data analyses are indeed effective for revealing how infants acquire the ability to produce vowels from a listener's point of view. However, the transcriber's subjectivity cannot be eliminated since transcribers must label even babbling and gibberish in the early stages. Lieberman (1980) also noted that in the early stages it is hard even for well-trained transcribers to label the phonemes of infants' speech. Although Kent and Murray (1982) used unlabeled speech data for infants and indicated the vowel space expansion with age, their unlabeled data analysis focused only on the development of vowel space expansion, not on categorically separated vowel acquisition. Thus, to support the results of labeled data analyses, we need other transcription-independent analyses based solely on acoustical properties to reveal the process by which speech production changes from being vague and random in infants to categorically separated and planned in adults.

In summary, previous studies involving acoustic analyses of infants' vowels in the first few years have the following shortcomings: (1) Although a wide interindividual variability stimulates a need for longitudinal observation, there have been few studies that analyzed the acoustical change in vowels quantitatively with sufficient observation points and numbers of data. (2) Most previous studies lacked the unlabeled (transcription-independent) analysis needed to detect when and how infants become able to produce categorically separated vowels.

To deal with the above shortcomings, this study provides labeled and unlabeled longitudinal analyses of the spectral peaks of vowels for two Japanese infants at more than 25 observation points between the ages of 4 and 60 months. In particular, we obtained large numbers of data for the first 2 years to allow us to understand the great change expected during that period. In addition, a sufficiently large number of data were employed to allow both quantitative and qualitative analyses. In Sec. II, the infant speech data properties are described, and the acoustic analysis method used for extracting spectral peaks is presented. In Sec. III, to reveal the longitudinal trends in the development process, the changes in the values of spectral peaks with age are investigated using phoneme labeled vowel data as employed in previous studies. In this section, the changes in the phonemic distribution of the spectral peaks on the vowel space are also investigated. In Sec. IV, an unlabeled analysis technique is introduced that was derived from the linear discriminant analysis method, and the changes in the distribution of the spectral peaks on the vowel space were measured without phoneme labels. The analysis revealed when and how two infants attain the ability to produce categorically separated vowels in the same way as adults. Section V is a general discussion of the results of labeled and unlabeled data analyses.

II. SPEECH DATA AND SPECTRAL PEAK EXTRACTION

A. Infant speech data

Speech data were obtained from normally developing male and female infants drawn from the NTT Japanese infant speech database (Amano *et al.*, 2002). These two infants were chosen because they account for the largest individual amounts of speech data in this database. The utterances of the infants and their parents were recorded in a room in their house with a digital audio recorder (SONY TCD-D10) and a stereo microphone (SONY ECM-959) with 16-bit quantization at a sampling rate of 48 kHz. The microphone was held by a parent or placed in a microphone stand during the recording. The infant and parents were not required to undertake any particular task for the recording so that their utterances would occur in natural situations in their daily life. The speech data were recorded for at least an hour per month. The total recording time and recording periods for the male and female infants were 139 hours from 0 to 54 months of age and 68 hours from 0 to 60 months of age, respectively.

In this study, speech data were used that were obtained at 4–6, 8–20, 22, 24, 25, 30, 34, 40, 44, and 52 months for the male infant, and at 4–22, 24, 25, 30, 35, 40, 45, 50, 55, and 60 months for the female infant. The speech data were down-sampled to a sampling rate of 16 kHz and phoneme labeled by two Japanese transcribers. The primary transcriber was highly experienced at phoneme labeling (over 10 years of experience) with a proficient knowledge of phonetics and acoustics, and the secondary transcriber was trained in phoneme labeling in advance by the primary transcriber. Using both auditory and visual (spectral) cues, the secondary transcriber first segmented and labeled the data according to the

TABLE I. Number of analyzed vowel data by month of age.

	Male infant						Female infant					
Age (month)	/a/	/e/	/i/	/o/	/u/	Total	/a/	/e/	/i/	/o/	/u/	Total
4	266	34	331	264	25	920	128	9	262	191	10	600
5	411	88	216	275	17	1007	267	22	154	146	5	594
6	1339	134	570	876	38	2957	213	26	274	232	26	771
7							461	33	251	179	16	940
8	156	62	41	107	5	371	510	88	410	283	43	1334
9	650	226	241	325	52	1494	467	106	205	199	21	998
10	406	235	199	357	36	1233	152	51	111	118	32	464
11	250	147	190	189	25	801	305	75	164	185	59	788
12	1213	484	993	623	144	3457	233	83	110	120	14	560
13	441	212	444	204	60	1361	121	28	79	26	25	279
14	669	307	466	272	114	1828	315	113	178	93	86	785
15	234	83	122	71	19	529	252	110	121	183	78	744
16	532	322	290	342	135	1321	304	175	143	155	127	904
17	336	157	91	95	66	745	237	106	107	122	133	705
18	606	363	220	159	129	1477	449	282	249	133	206	1319
19	218	152	92	71	56	589	385	179	149	116	192	1121
20	413	444	230	206	244	1537	547	308	181	147	270	1453
21							437	275	196	216	285	1409
22	198	123	84	91	81	577	737	381	278	270	445	2111
24	1026	611	395	460	480	2972	579	288	167	304	403	1741
25	584	318	189	168	268	1527	427	240	140	167	280	1254
30	442	275	193	171	228	1309	111	88	26	63	64	352
34	461	194	196	201	248	1300						
35	82	31	25	23	51	212	297	115	109	156	173	850
40	158	78	66	90	65	457	612	322	162	369	394	1859
44	249	116	99	105	130	699						
45	18	10	11	16	12	67	1627	654	411	761	883	4336
50							285	110	47	135	169	746
52	399	171	152	213	189	1124						
55							268	131	43	103	156	701
60							285	125	48	192	187	837
Total	11 757	5377	6146	5974	2917	32 171	11 011	4623	4775	5364	4782	30 555

Japanese phoneme inventory. The primary transcriber then checked and corrected this work. The label onsets and offsets were specified in millisecond units. The two transcribers were asked to provide phoneme labels even when the sample was ambiguous. To exclude ambiguous utterances, this study only used the phoneme labels that were agreed on by the two transcribers. Based on these phoneme labels, we automatically extracted the five Japanese vowels /a/, /e/, /i/, /o/, and /u/ with durations of over 50 ms. Speech data that included crying, laughing, reading, or singing were eliminated from this analysis. Table I shows the number of vowel data for each month of age.

B. Spectral peak extraction

To avoid as far as possible the effect of coarticulations, we analyzed 512 sampling points (32 ms) that began at the first quarter position of the total duration of each labeled vowel. The samples were analyzed with the 12-order linear predictive coding (LPC) analysis method (Itakura, 1975). Frequency spectra were estimated based on the LPC coefficients, and the two lower spectral peaks were extracted automatically from each spectrum. Henceforth in this paper, the lower of the two spectral peaks is denoted as "$F1$" and the other as "$F2$." This paper considers that $F1$ and $F2$ basically correspond to the first and second formants of the vowels. However, it should be noted that automatic estimation may result in subtle differences from the precise formant peaks (Vallabha and Tuller, 2002), especially when used for vowels with high fundamental frequencies such as infant speech. To avoid error when estimating the resonant peaks, we eliminated the vowel data whose estimated $F1$ or $F2$ was the same as the fundamental frequency ($F0$). We estimated $F0$ by the robust $F0$ estimation method (Nakatani and Irino, 2004) and a trained operator corrected the misestimated results (Amano *et al.*, 2006).

III. ANALYSIS OF PHONEME LABELED DATA

A. Changes in spectral peaks with age

We first investigated the changes in spectral peak values with age. Figure 1 shows the mean $F1$ and $F2$ values as a function of age in months for each infant. To reveal the trend in the values with age, the mth-order polynomial function

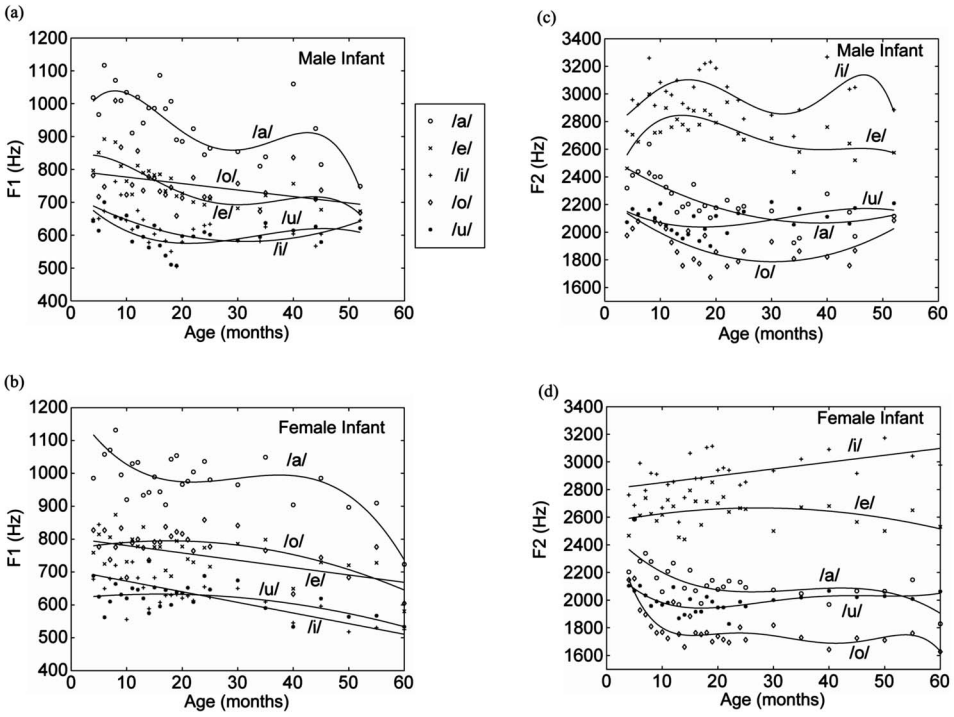


FIG. 1. (a) Means of $F1$ values of the male infant as functions of month of age and fitted polynomial curves for $F1$. (b) Means of $F1$ values of the female infant and fitted curves. (c) Means of $F2$ values of the male infant and fitted curves. (d) Means of $F2$ values of the female infant and fitted curves.

was fitted to the data with the months of age as explanatory variables. The polynomial function is as follows:

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m, \tag{1}$$

where y represents the object variables (mean $F1$ or $F2$ values), x represents the explanatory variables (the ages in months), and a_i represents the ith polynomial coefficients. The coefficients a_i can be estimated by the least squared method for polynomial function fitting after fixing the order m. The optimal order m of the polynomial function for each vowel was decided by using the Akaike information criterion (AIC; Akaike, 1974) for polynomial regression analysis as follows:

$$\mathbf{AIC} = n \log 2\pi + n \log \hat{\sigma}^2 + n + 2(m+2), \tag{2}$$

where **AIC** is the AIC value, n is the total number of data (number of mean $F1$ or $F2$ values), and $\hat{\sigma}^2$ is the residual variance calculated by the least squared method. A model with a small AIC value is considered more appropriate than one with a higher value. We selected the order with the minimum AIC value for each vowel.

Figure 1 also shows the fitted polynomial functions for each vowel as a function of age in months. These results reveal the trend of $F1$ [Figs. 1(a) and 1(b)] as follows. (1) Both infants' $F1$ values for vowel /a/ decrease gradually until around the age of 20 to 30 months, then remain almost unchanged from 25 to 40 months, and finally decrease again after 40 months. Both polynomial regressions were significant ($p<0.001$). (2) Both infants' $F1$ values for vowel /e/ decrease until the age of 30 months. After this age, the $F1$ values of the female infant continue to decrease, whereas those of the male infant remain almost unchanged. Both polynomial regressions were significant ($p<0.001$). (3) Both infants' $F1$ values for vowel /i/ decrease until around the age of 20 months. Subsequently, the $F1$ values of the female infant continue to decrease, whereas those of the male infant remain almost unchanged. Both polynomial regressions were significant ($p<0.01$). (4) The $F1$ values of the male infant for vowel /o/ remain unchanged for the whole period. The polynomial regression was not significant. On the other hand, the $F1$ values of the female infant for vowel /o/ remain unchanged until around the age of 20 months and then decrease. The polynomial regression was significant ($p<0.005$). (5) The $F1$ values of the male infant for vowel /u/ first decrease until around the age of 20 months, then remain unchanged. On the other hand, the $F1$ values of the female infant for vowel /u/ first remain unchanged until around the age of 20 months and then decrease. Both polynomial regressions were significant ($p<0.05$).

As shown in Figs. 1(c) and 1(d), the results obtained from polynomial curve fitting also reveal the trend of $F2$ as follows. (1) Both infants' $F2$ values for vowel /a/ decrease rapidly until around the age of 20 months, then remain almost unchanged. Both polynomial regressions for vowel /a/ were significant ($p<0.0005$). (2) The $F2$ values of the male infant for vowel /e/ first remain unchanged until the around the age of 20 months, then decrease until the age of 30 months, and finally remain unchanged. The polynomial regression was significant ($p<0.005$). On the other hand, the $F2$ values of the female infant for vowel /e/ remain unchanged from 4 to 60 months. The polynomial regression was not significant. (3) The $F2$ values of the male infant for vowel /i/ remain unchanged for the entire period. The polynomial regression was not significant. On the other hand, the $F2$ values of the female infant for vowel /i/ increase with age. The polynomial regression was significant ($p<0.005$). (4) Both infants' $F2$ values for vowel /o/ first decrease until around the age of 18 months and then remain unchanged. Both polynomial regressions were significant ($p<0.005$). (5)

TABLE II. Significance and goodness-of-fit (coefficient of determination, R^2) for fitted polynomial functions shown in Fig. 1.

		Male Infant		Female Infant	
	Vowel	Significance	R^2	Significance	R^2
F1	/a/	$p<0.0005$	0.604	$p<0.001$	0.508
	/e/	$p<0.0001$	0.741	$p<0.001$	0.353
	/i/	$p<0.01$	0.362	$p<0.0001$	0.528
	/o/	n.s.	0.132	$p<0.005$	0.381
	/u/	$p<0.05$	0.381	$p<0.005$	0.361
F2	/a/	$p<0.0001$	0.593	$p<0.0005$	0.539
	/e/	$p<0.005$	0.541	n.s.	0.117
	/i/	n.s.	0.298	$p<0.005$	0.285
	/o/	$p<0.005$	0.429	$p<0.0001$	0.729
	/u/	$p<0.005$	0.591	$p<0.005$	0.482

Both infants' $F2$ values for vowel /u/ decrease until around the age of 18 months and then remain unchanged. Both polynomial regressions were significant ($p<0.005$).

The significance and the goodness-of-fit of the fitted polynomial functions described above are shown in Table II. The goodness-of-fit was measured by the coefficient of determination (R^2) obtained from the fitting.

B. Phonemic distribution of spectral peaks

To investigate the phonemic developmental process of the spectral peaks in more detail, we plotted the vowel data on a plane whose abscissa and ordinate indicate $F1$ and $F2$, respectively (henceforth we call this the "$F1-F2$ plane"). Figure 2 shows 50% probability ellipses for each vowel corresponding to the phoneme labels plotted on the $F1-F2$ plane. $F1$ and $F2$ for each vowel have an approximately normal distribution. Thus, a 50% probability ellipse includes at least half of the number of the corresponding vowel, which confirms that the region of the ellipse provides a good representation of the typical $F1/F2$ region of each vowel. As shown in Fig. 2, the range in which the probability ellipses were depicted increases with age, especially from 4 to 18 months. Although there are large overlaps between the density ellipses at around the age of 12 months, the overlaps become smaller with age due to the spread of the centroids and the reduction in the ranges of each density ellipse. In particular, overlaps between the ellipses of vowels /a/ and /i/, /o/ and /i/, and /o/ and /e/ disappear at around the age of 18 months. However, it should be noted that the overlaps between the ellipses of the male infant often increase even after 18 months.

To evaluate the reduction in the overlaps quantitatively, we calculated the Euclidean distances between the pairs of vowel distribution centroids, and the Mahalanobis generalized distances between the pairs of vowel distributions. The Mahalanobis generalized distance $D_M(m_1,m_2)$ between two vowel categories ω_1 and ω_2 whose mean vectors $[F1,F2]$ are m_1 and m_2 was calculated as follows (Duda *et al.*, 2001):

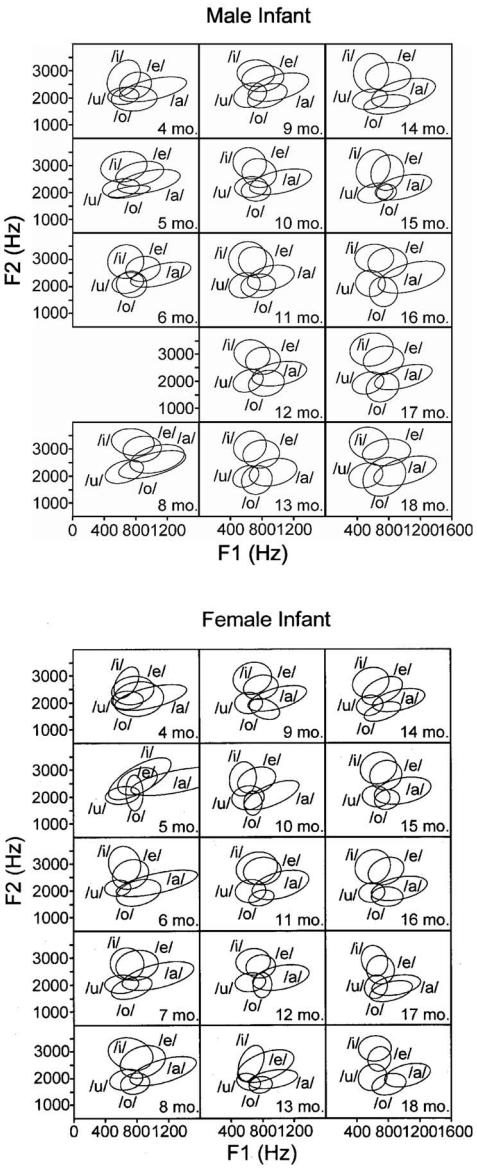


FIG. 2. Fifty percent probability ellipses for each vowel on the $F1-F2$ plane by month of age.

$$D_M^2(m_1,m_2)=(m_1-m_2)^t\Sigma_w^{-1}(m_1-m_2) \tag{3}$$

where Σ_w^{-1} is the inverse of the within-class covariance matrix calculated as follows:

$$\Sigma_w=\sum_{i=1,2}P(\omega_i)\Sigma_i, \tag{4}$$

where $P(\omega_i)$ is the *a priori* probability of vowel category ω_i (i.e., the proportion of a vowel category to all vowel categories for a given month of age) and Σ_i is the covariance matrix calculated from all the vowels included in vowel category ω_i. The Euclidean distance can only measure the distance between the centroids of two distributions without considering the scatter of the distributions and their overlaps, whereas the Mahalanobis distance can measure the distance between two distributions taking them into account. Therefore, the Mahalanobis distance becomes larger when the overlaps between two distributions become smaller even if the Euclidean distance between the centroids remains same.

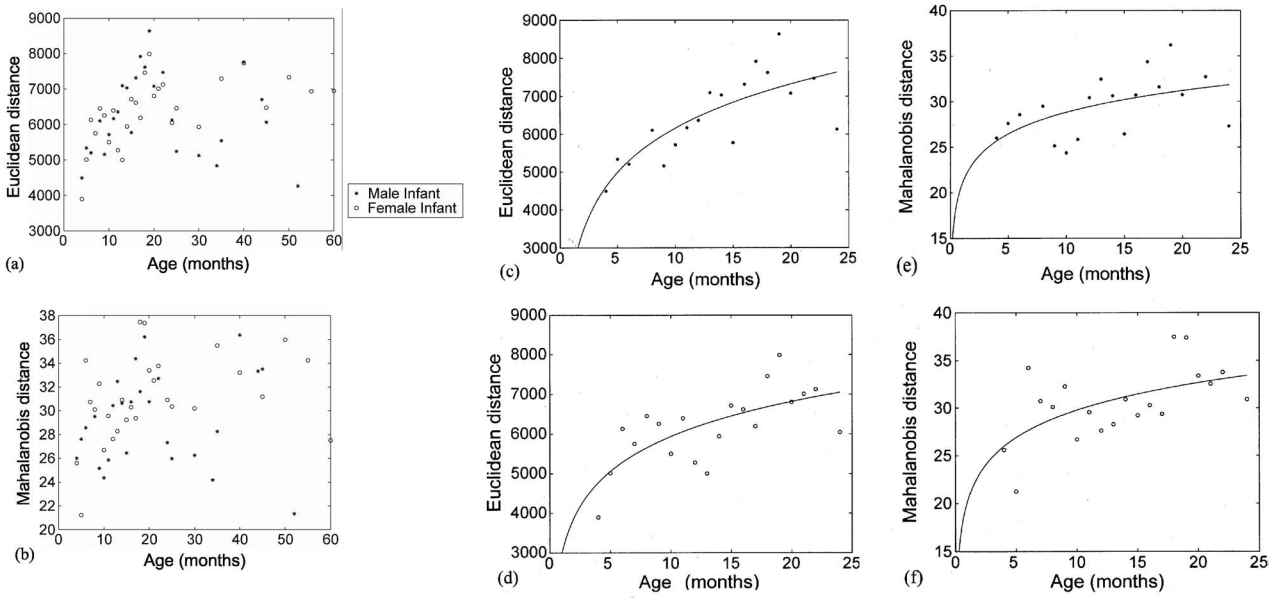


FIG. 3. (a) The sum of the Euclidean distances between the vowel centroids for each month of age. The increment in distance indicates an expansion among the center of each vowel distribution. (b) The sum of the Mahalanobis distances between the vowel distributions for each month of age. The increment in distance supports the idea of a reduction in the density overlaps of the vowel distributions. (c) The sum of the Euclidean distances of the male infant until the age of 24 months and the fitted logarithmic curve. (d) The sum of the Euclidean distances of the female infant until the age of 24 months and the fitted logarithmic curve. (e) The sum of the Mahalanobis distances of the male infant until the age of 24 months and the fitted logarithmic curve. (f) The sum of the Mahalanobis distances of the female infant until the age of 24 months and the fitted logarithmic curve.

Figure 3(a) shows the sums of the Euclidean distances between each pair of vowel centroids for each month of age, and Fig. 3(b) shows the sums of the Mahalanobis distances between each pair of vowel distributions for each month of age. The same tendencies can be observed between the ages of 4 and 24 months, namely the distances increase greatly with age. Figures 3(c) and 3(d) show logarithmic curve fittings to the sums of the Euclidean distances from the age of 4 to 24 months, and the result revealed significant increases with age ($t_{(16)}=5.06$, $p<0.0005$ for the male infant; $t_{(18)}=4.13$, $p<0.001$ for the female infant). Figures 3(e) and 3(f) show logarithmic curve fittings to the Mahalanobis distances from the age of 4 to 24 months, and the result also revealed significant increases with age ($t_{(16)}=2.51$, $p<0.05$ for the male infant; $t_{(18)}=2.89$, $p<0.01$ for the female infant). The increment in distance supports the idea that there is a reduction in the overlapping between the vowel distributions. On the other hand, interindividual differences can be observed between the two infants after the age of 25 months. While the distances remain at higher values for the female infant, the distances for the male infant become smaller between the ages of 25 and 35 months.

C. Discussion

Although we used LPC analysis for the spectral peak estimations, most of the $F1$ and $F2$ values are in the same range as the formant values in previously reported results based on phoneticians' labeling (Eguchi and Hirsh, 1969; Lieberman, 1980; Buhr, 1980; Kent and Murray, 1982; Bond *et al.*, 1982; Gilbert *et al.*, 1997). This fact allows us to assume that the spectral peaks $F1$ and $F2$ basically correspond to the first and second formants.

Although small interindividual variability can be observed, the trends of the changes in the $F1$ values are largely similar for the two infants. As shown in Figs. 1(a) and 1(b), the $F1$ values are separated corresponding to the height of the articulation position with age in the early stages. In particular, the difference between the vowel groups becomes larger until around the age of 20 months. The $F1$ values for high vowels, i.e., /i/ and /u/, decrease until around 25 months of age. The $F1$ values for a low vowel, i.e., /a/, first decrease, then the values remain unchanged between 20 and 40 months, maintaining their higher values despite decrements in the $F1$ values for other vowels. The values for middle vowels, i.e., /e/ and /o/, decrease gradually while maintaining middle values.

The $F1$ value is dependent on the vocal tract length and the height of the articulation position of the tongue, and so high (low) vowels have a low (high) $F1$ value in mature production. Thus, the differentiation in the $F1$ values of vowels until around 20 months possibly reflects the vocal tract lengthening and vertical tongue elevation, and this leads to the distinct articulation of high/middle/low vowels. We speculate that the vocal tract length increment and the tongue elevation were realized by the rapid growth of the pharyngeal cavity and the rapid descent of the larynx and hyoid between 15 and 20 months of age (Vorperian *et al.*, 2005). In addition, decrements in the $F1$ values in later periods of development can be considered a reflection of the gradual increase in the vocal tract length after 20 months.

The trends of the changes in the $F2$ values are also largely similar for the two infants. As shown in Figs. 1(c) and 1(d), the $F2$ values are separated corresponding to the horizontal articulation position with age in the early stages.

In particular, until around the age of 16 months, the difference between the vowel groups is large for both infants. The $F2$ values for the front vowels, i.e., /e/ and /i/, increase or remain unchanged until 16 months of age. The $F2$ values for the central vowels /a/ and /u/,[1] and the values for the back vowel /o/, decrease until 20 months of age. The decrement in the values for vowel /o/ is larger than that for the central vowels.

The $F2$ values correspond to the horizontal tongue advancement to form an articulation position. The tongue lengthens rapidly until 16 months (Vorperian *et al.*, 2005), and this age coincides with the increase in the difference between the $F2$ values of the front (/i/ and /e/) and the central (/a/ and /u/) and back (/o/) vowels. Therefore, these acoustic changes, which correspond to the horizontal articulation position of the tongue until around 16 months, may reflect the enlargement of the tongue until 16 months, and this leads to the distinct articulation of front/central/back vowels. Interestingly, different $F2$ values were also observed within the same vowel groups. The $F2$ values for vowel /e/ decrease after around 30 months, although the values for vowel /i/ increase, whereas the $F2$ values of both vowels show the same tendency as regards the developmental process until that time. The differentiation within the front vowels suggests that the separation caused by the horizontal tongue position first occurs as the result of relatively rough positioning and is then fine-tuned after 30 months.

In particular, the phonemic vowel distributions on the $F1-F2$ plane indicated that the vowel space expands until around the age of 18 months. This finding suggests that the infant rapidly learned to produce vowels that can be perceived by adults (transcribers) until around that age. Logarithmic curve fitting to the Euclidean and Mahalanobis distances revealed that such expansion occurs exponentially with age until around 24 months. These results also agree with anatomical findings showing that articulatory organs mature rapidly until around 18 months and then mature gradually (Vorperian *et al.*, 2005). In our study, we were able to measure the degree of expansion more quantitatively using a large amount of vowel data for the two infants and thus support the partial findings presented in previous studies (Kent and Murray, 1982; Lieberman, 1980; Buhr, 1980; Bond *et al.*, 1982; Kuhl and Meltzoff, 1996).

However, there are interindividual differences in the vowel space expansion after 24 months. With the female infant, the vowel space continued to expand. On the other hand, with the male infant, the vowel space contracted once after 24 months and then expanded again. Such a contraction of the vowel space with age was observed for some subjects in a previous study (Lieberman, 1980). This interindividual variability is also discussed in Sec. V.

Although these results revealed the vowel space expansion from a transcriber's point of view, it is still unclear how the vague vowel space of the early stages becomes categorically separated from a transcription-independent physical view. This is because, as Lieberman (1980) pointed out, the phoneme labels, especially in the early period, were subjectively provided by the transcribers and involved the difficulty of labeling immature speech. To investigate how infants become capable of producing categorically separated vowels in terms of transcription-independent characteristics, we employed an unlabeled analysis method derived from the discriminant analysis method. This is described in the next section.

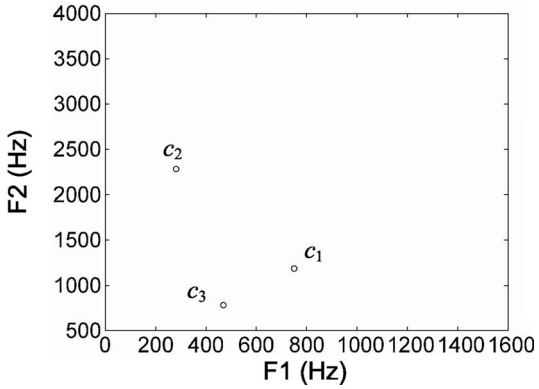


FIG. 4. Reference points for the discriminant function on the $F1-F2$ plane. c_1, c_2, and c_3 correspond to the centroids of Japanese adult vowels /a/, /i/, and /o/, respectively.

IV. ANALYSIS OF UNLABELED DATA

A. Process of vowel space expansion

To investigate the developmental process in vowel distribution based on the physical, transcription-independent characteristics of $F1$ and $F2$, we employed an unlabeled analysis technique derived from the linear discriminant analysis method. In this analysis, we used only speech data that were labeled as vowels by the transcribers and did not use the other data, which were labeled as voiced or unvoiced consonants. First, we fixed three reference points as the centroids of the typical "corner" vowels of the Japanese vowel space, i.e., /a/, /i/, and /o/.[2] Henceforth, these centroid vectors are called $c_i(i=1,2,3)$, which have $[F1,F2]$ values of [750, 1187] for vowel /a/, [281, 2281] for vowel /i/, and [468, 781] for vowel /o/ in Hz as suggested by Hirahara and Kato (1992). The vowel space of adults rather than children was employed because what infants actually learn to speak is the ambient input of adults' speech, and the adult vowel space has a more categorically separated canonical shape than that of children. The positions of c_i on the $F1-F2$ plane are plotted in Fig. 4. Using these vectors as a basis, we investigated the vowel space expansion with age quantitatively by studying the problem of classifying the vowel vector d_j, which places each vowel (j is a randomly assigned index of the vowel data for a month of age whose value is $j=1,\ldots,J$, J is the total number of vowels in a month age) in one of three categories;

Π_1: vowel /a/ whose centroid is c_1,
Π_2: vowel /i/ whose centroid is c_2, and
Π_3: vowel /o/ whose centroid is c_3.

First, to avoid the effect of the difference between spectral peak frequencies caused by the difference between the vocal tract lengths of the infants and Japanese adults, before calculating the divergence, we normalized d_j so that the center of

the vowel data for each age coincides with the center of the adult's vowel data. This normalization allows us to focus on the developmental process of forming a categorically separated vowel space. The normalization was conducted as follows. First, the center of the adult's centroids C_c and the center of the infant's vowels C_d were calculated with Eqs. (5) and (6). Then, all the vowel data were normalized as $C_c = C_d$ in Eq. (7):

$$C_c = \frac{\sum_{i=1}^{3} c_i}{3}, \tag{5}$$

$$C_d = \frac{\sum_{j=1}^{J} d_j}{J}, \tag{6}$$

$$\bar{d}_j = d_j - (C_d - C_c). \tag{7}$$

Then, classification was performed by the divergence $\sqrt{\Sigma|\bar{d}_j - c_i|^2}$. We introduced the discriminant function:

$$D_i^j = \sqrt{\sum |\bar{d}_j - c_i|^2} - \sqrt{\sum |\bar{d}_j - c_k|^2}, \quad k = \mathrm{mod}(i/3) + 1, \tag{8}$$

where "mod" is the modulus operator that returns the remainder. The actual classification procedure is, if D_1^j (where $i=1$, $k=2$) >0, then we consider that vowel data $\bar{d}_j$ belongs to category Π_2, because $D_1^j > 0$ means that the first term of Eq. (8) is larger than the second term, i.e., $\bar{d}_j$ is closer to c_2 than c_1, for example. Based on Eq. (8), all vowels were classified into two categories (Π_i and Π_k), and the classification boundary was the line where $D_i^j = 0$. The classification was performed for pairs of Π_1 and Π_2, Π_2 and Π_3, and Π_3 and Π_1. We calculated all the D_i^j values for each month of age and generated histograms of the values for each i. Figure 5 shows the histograms. These histograms reveal that the D_i^j values are first broadly distributed, and then they become biased towards the two ends of the histograms. This means that the vowels are clustered around either of the two reference points, that is, the vowels are categorically separated to the corner vowels. This analysis qualitatively confirms that the vowels form a categorically separated vowel space with age solely through the use of transcription-independent acoustic features.

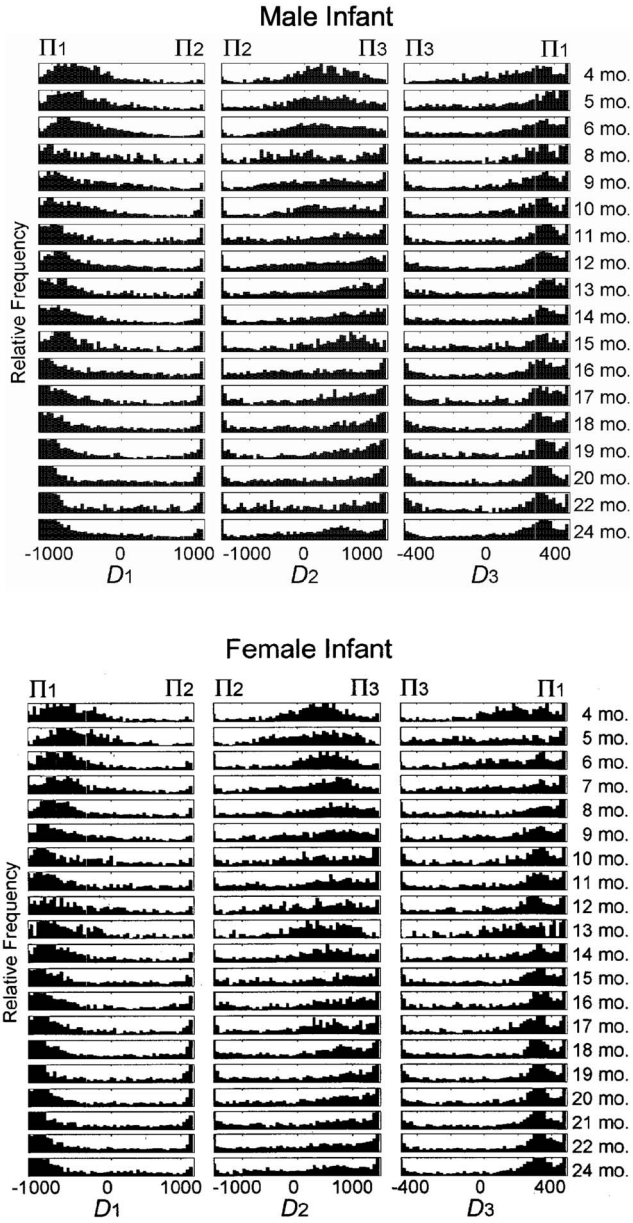


FIG. 5. Histograms of the discriminant function values by month of age. Π_i indicates the categories for the function (Π_1: vowel /a/; Π_2; vowel /i/; Π_3: vowel /o/). The discriminant function values are first broadly distributed until around 15 months of age, then they become biased towards the two ends of the histograms. A biased distribution means that the vowels are clustered around either of the two reference points (adult corner vowels).

B. Speed of vowel space expansion

To investigate the size of the developmental changes in the vowel distribution, we employed the temporal differences in the histograms from the first observed month (4 months) as sums of the differences of each histogram bin as a measure. The temporal differences in the histograms S_i were calculated as follows:

$$S_i^m = \sum_{k=1}^{N} |b_i^m(k) - b_i^4(k)|, \tag{9}$$

where $b_i^m(k)$ is the value of the kth bin of the histogram for D_i at m months of age, and N is the number of bins for the histogram. S_i^m indicates the size of the developmental changes in the vowel space from the age of 4 months. Namely, it shows the way in which the vowel space is categorically separated because changes in the histograms indicate that vowels become categorically separated with age as shown in Fig. 5.

Figure 6 shows the difference as a function of age in months. As shown in Fig. 6, the S_i^m values were small when

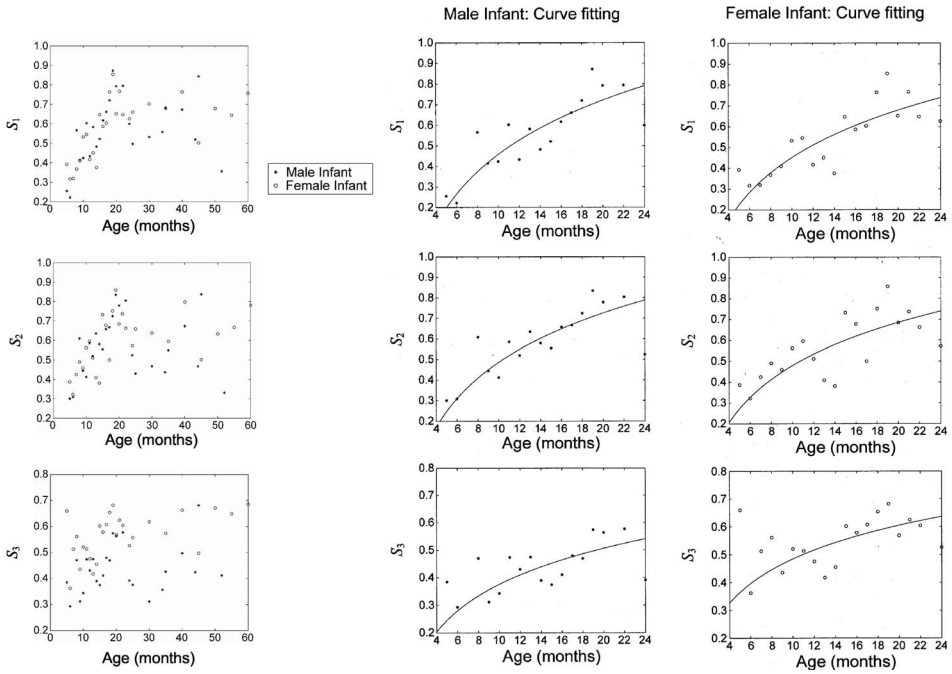


FIG. 6. Differences in the histograms shown in Fig. 5 from the first histogram (4 months) by month of age. The differences show how the vowel space is categorically separated.

the month of age was small, whereas the difference becomes larger with age. It should be noted that the developmental changes depicted by S_i^m are very similar to the Euclidean and Mahalanobis distances shown in Fig. 3. Logarithmic curve fitting with the age in months as an explanatory variable until the age of 24 months reveals that S_i^m increases rapidly with age ($t_{(15)}=6.04$, $p<0.0001$ for S_1 of the male infant; $t_{(15)}=5.25$, $p<0.0001$ for S_2 of the male infant; $t_{(15)}=2.87$, $p<0.05$ for S_3 of the male infant; $t_{(17)}=6.04$, $p<0.0001$ for S_1 of the female infant; $t_{(17)}=4.26$, $p<0.001$ for S_2 of the female infant; $t_{(17)}=3.93$, $p<0.005$ for S_3 of the female infant). In other words, the vowel space is rapidly categorically separated until the age of 24 months. This result coincides with the result of the labeled data analysis in Sec. III B. Interestingly, the interindividual differences in the tendencies after the age of 24 months are also similar to the results in Sec. III B and Fig. 4.

C. Discussion

The changes in the histograms of the values obtained from the linear discriminant functions revealed the process of vowel space expansion. The infants' vowel space expands toward the edges of the adults' vowel space with age. This suggests that the vowels expand from a small region, and cluster in the vowel space particularly until around 24 months. The vowel data distribution of infants places the region containing $F1$ and $F2$ higher than that of adults because of the infants' shorter vocal tract length (e.g., Kuhl and Melzoff, 1996). However, the above result suggests that the developmental process forms a categorically separated vowel space similar to that of an adult because the vowel data distributions of the infant and adult were normalized as their centers became the same.

In addition, an analysis of the differences between the first histogram (4 months) and later histograms revealed the speed of vowel space expansion. The result confirmed that the change occurs particularly in the early stages of the developmental process. The goodness-of-fit of the logarithmic curves to the differences indicates that the developmental speed of vowel space expansion is rapid in the early stages and then becomes slower. This result supports the exponential expansion of phonemic distribution as described in Sec. III. This analysis of the vowel distributions suggested that the vowels expand from a small region, and cluster in the vowel space particularly until 24 months in terms of their pure acoustic characteristics independent of the phoneme labels.

V. GENERAL DISCUSSION

This study investigated when and how infants become able to produce categorically separated vowels by analyzing the longitudinal development of the spectral peaks of vowels using labeled and unlabeled data obtained from two infants. The labeled data analysis in Sec. III revealed the developmental change of the spectral peak values (i.e., $F1$ and $F2$), as well as the vowel space expansion in the phonemic distribution on the $F1-F2$ plane. The spectral peak values of each vowel were categorically separated by around 18 months of age as described in Sec. III A. An analysis of the vowel distribution on the $F1-F2$ plane in Sec. III B showed that the overlaps of the probability ellipses of each labeled vowel become smaller with age particularly until around 24 months. These findings suggest that the infants become able to produce categorically separated vowels by that age, and the speed of expansion is rapid, particularly in the early stages, presumably in the first 2 years. To confirm the vowel space expansion revealed in Sec. III, in Sec. IV we undertook an unlabeled data analysis of the spectral peaks. The approach was derived from the linear discriminant analysis method. The results further support the view that the infant vowel space becomes similar to that of an adult by 24 months of age as described in Sec. IV A. In addition, as

shown in a labeled data analysis of the distance between vowel distributions, the unlabeled data analysis in Sec. IV B also supported the idea that the categorically separated expansion is especially rapid before 24 months.

Although there are interindividual differences after the age of 24 months, the developmental processes of the two infants are very similar until around that age. The rapid development in the first 2 years without noteworthy interindividual differences suggests that the process was caused mostly by some universal developmental change such as the growth of the articulatory organs. This period indeed corresponds to the age at which the articulatory organs grow rapidly. Vorperian *et al.* (2005) investigated the anatomical development of the vocal tract using MRI and indicated that the vocal tract structures exhibit accelerated growth, particularly between birth and 18 months. This correspondence suggests that the infants acquire a categorically separated vowel space, which is the same as that of an adult, by the rapid development of the articulatory organs during this period.

On the other hand, the development processes of categorically separated vowel production of the two infants described above may not result solely from the simple anatomical development of the articulatory organs, but may also require the development of the articulation skills of vowel production. Therefore, the correspondence between the ages of categorically separated vowel production and rapid maturation of articulation organs suggests that the development of articulation skills also begins in the early stages of an infant's development. This is supported by the fact that infants as young as 5 months old imitate the vowels they hear (Kuhl and Meltzoff, 1996; Patterson and Werker, 1999). Despite the immaturity of their articulation organs, infants tend to imitate the vowels they hear by using their articulation skills.

The way in which infants acquire their articulation skills remains ambiguous. However, without the need to regard any given acquisition model as true (e.g., Guenther, 1994, 1995; Markey, 1994; Bailly, 1997), speech feedback from the auditory system (i.e., perception of the vowels produced by adults and the infants themselves) is apparently essential in terms of learning to produce speech. Previous perception research on infants (e.g., Trehub, 1973; Kuhl, 1983) has revealed that they can discriminate vowels at a very early stage (before 6 months).In addition, even before 12 months of age, perception adapts to the inherent vowel code of the infant's native language (Kuhl *et al.*, 1992; Pegg and Werker, 1997). Therefore, the early perceptual sensitivity to vowels may facilitate the smooth and rapid development of the articulation skills needed for vowel production in the auditory feedback system.

After 24 months of age, we observed a wide interindividual variability between the two infants, particularly in the developmental speed of the vowel space expansion. At around this age, the infants have passed through the babbling period and become able to produce meaningful words. Although more work is needed in this area, it can be speculated that, as an infant becomes a proficient speaker of his/her native language, the articulation tends to be influenced by the ambient speech provided by caregivers. Such influences must be present even in earlier periods. However, the influences of anatomical developments are presumably predominant until 24 months of age, and they conceal the influence of ambient speech. Given the fact that the manner of vowel articulation in adults has wide interindividual variability, it is plausible that each infant develops a particular speech style according to his/her individual environment. If this is the case, the wide interindividual difference between the two infants we observed in the later period possibly reflected the specificity of the speech style in each infant.

We should point out that the findings in this paper have the limitations of a generalization because the speech data were from just two infants. Nonetheless, because of the large amount data available for the infants, this study revealed the long-term trend of vowel development quantitatively. In addition, this longitudinal study could reveal the interindividual variability of the developmental speed of each infant, which may reflect both the development of the articulatory organs and of the articulation skills of vowel production. Moreover, by comparing the two longitudinal data analyses, this study also clearly revealed the similarity between the development processes of the two infants. Longitudinal studies using speech data from a larger number of infants, and cross-linguistic comparisons of infant vowel production, constitute future work.

In summary, we conducted a longitudinal study of the vowel development of two Japanese infants between the ages of 4 and 60 months in terms of changes in the spectral peaks of vowels to investigate when and how the infants become able to produce categorically separated vowels in the early development stages. By using a large number of data for each infant, our phoneme labeled data analysis revealed longitudinal trends of vowel development that correspond to the articulation positions of the tongue and the vocal tract length. In addition, the labeled data analysis of the vowel distributions of spectral peaks revealed that the categorically separated vowel space is formed by around the age of 20 months. To support the results of the labeled data analysis, we employed an unlabeled data analysis technique derived from linear discriminant analysis. This unlabeled (transcription-independent) data analysis confirmed that the vowel space expands with age and becomes more similar to that of an adult by an age of around 20 months. The results of the labeled and unlabeled data analyses suggest that the early acoustic development of the infant's vowels corresponds to the rapid anatomical development of the vocal tract during the early stages of the speech development process.

ACKNOWLEDGMENTS

The authors thank Dr. Tadahisa Kondo (NTT Communication Science Laboratories, NTT Corporation) for recording the infant speech data, and Dr. Tomohiro Nakatani (NTT Communication Science Laboratories, NTT Corporation) for extracting the fundamental frequencies of the infants' speech. The authors also thank Mr. Ken-ichi Sakakibara (Dept. of Communication Disorders, Health Sciences University of Hokkaido) for valuable comments on speech production, and Professor Kazuhiko Kakehi (School of Computer and Cog-

nitive Sciences, Chukyo University) for his strong recommendation of transcription-independent analysis.

[1]The horizontal articulation position of the tongue for vowel /u/ in Japanese differs from that in English and is the almost same as that for vowel /a/ (Vance, 1987). Therefore, vowel /u/ in Japanese can be categorized as a central vowel.

[2]As described in footnote 1, the horizontal articulation position for Japanese vowel /u/ is further forward than that for English (Vance, 1987). Therefore, we chose vowel /o/ as a corner vowel of the vowel space rather than vowel /u/.

Akaike, H. (**1974**). "A new look at statistical model identification," IEEE Trans. Autom. Control **19**, 716–723.

Amano, S., Kato, K., and Kondo, T. (**2002**). "Development of Japanese infant speech database and speaking rate analysis," *Proc. of the 7th International Conference on Spoken Language Processing (ICSLP)*, Vol. **1**, pp. 317–320.

Amano, S., Nakatani, T., and Kondo, T. (**2006**). "Fundamental frequency of infants' and parents' utterances in longitudinal recordings," J. Acoust. Soc. Am. **119**, 1636–1647.

Bailly, G. (**1997**). "Learning to speak. Sensori-motor control of speech movements," Speech Commun. **22**, 251–267.

Bennett, S. (**1981**). "Vowel formant frequency characteristics of preadolescent males and females," J. Phonetics **10**, 417–422.

Bond, Z. S., Petrosino, L., and Dean, C. R. (**1982**). "The emergence of vowels: 17 to 26 months," J. Phonetics **10**, 417–422.

Buhr, R. D. (**1980**). "The emergence of vowels in an infant," J. Speech Hear. Res. **23**, 73–94.

Busby, P. A., and Plant, G. L. (**1995**). "Formant frequency values of vowels produced by preadolescent boys and girls," J. Acoust. Soc. Am. **97**, 2603–2606.

Duda, R. O., Hart, P. E., and Stork, D. G. (**2001**). *Pattern Classification Second Edition* (Wiley, New York), pp. 33–36.

Eguchi, S., and Hirsh, I. J. (**1969**). "Development of speech sounds in children," Acta Oto-Laryngol., Suppl. **257**, pp. 1-51.

Ferguson, C. A. (**1979**). "Phonology as an individual access system: some data from language acquisition," in *Individual Differences in Language Ability and Language Behavior*, edited by C. J. Fillmore, D. Kempler, and W. S.-Y. Wang (Academic, New York), pp. 189–201.

Ferguson, C. A., and Farwell, C. (**1975**). "Words and sounds in early language acquisition," Language **51**, 419–439.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Gilbert, J. H. (**1973**). "Acoustical features of children's vowel sounds: Development by chronological age versus bone age," Lang Speech **16**, 218–223.

Gilbert, H. R., Robb, M. P., and Chen, Y. (**1997**). "Formant frequency development: 15 to 36 months," J. Voice **11**, 260–266.

Goodell, E. W., and Studdert-Kennedy, M. (**1993**). "Acoustic evidence for the development of gestural coordination in the speech of 2-year-olds: A longitudinal study," J. Speech Hear. Res. **36**, 707–727.

Guenther, F. H. (**1994**). "A neural network model of speech acquisition and motor equivalent speech production," Biol. Cybern. **72**, 43–53.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hirahara, T., and Kato, H. (**1992**). "The effect of F0 on vowel identification," in *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 89–112.

Itakura, F. (**1975**). "Minimum prediction residual principle applied to speech recognition," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-23**, 67–72.

Kent, R. D., and Murray, A. D. (**1982**). "Acoustic features of infant vocalic utterances at 3, 6, and 9 months," J. Acoust. Soc. Am. **72**, 353–365.

Kuhl, P. K. (**1983**). "Perception of auditory equivalence classes for speech in early infancy," Infant Behav. Dev. **6**, 263–285.

Kuhl, P. K., and Meltzoff, A. N. (**1996**). "Infant vocalizations in response to speech: Vocal imitation and developmental change," J. Acoust. Soc. Am. **100**, 2425–2438.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experiences alter phonetic perception in infants by 6 months of age," Science **255**, 606–608.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Lieberman, P. (**1980**). "On the development of vowel production in young children," in *Child Phonology, Volume 1 Production*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, London), pp. 113–142.

Markey, K. L. (**1994**). "The sensorimotor foundations of phonology: A computational model of early childhood articulatory and phonetic development," Ph.D. thesis, University of Colorado, Boulder, CO.

Nakatani, T., and Irino, T. (**2004**). "Robust and accurate fundamental frequency estimation based on dominant harmonic components," J. Acoust. Soc. Am. **116**, 3690–3700.

Patterson, M. L., and Werker, J. F. (**1999**). "Matching phonetic information in lips and voice is robust in 4.5-month-old infants," Infant Behav. Dev. **22**, 237–247.

Pegg, J. E., and Werker, J. F. (**1997**). "Adult and infant perception of two English phones," J. Acoust. Soc. Am. **102**, 3742–3753.

Trehub, S. E. (**1973**). "Infants' sensitivity to vowel and tonal contrasts," Dev. Psychol. **9**, 91–96.

Vallabha, G. K., and Tuller, B. (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State University of New York, New York).

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.

Whiteside, S. P. (**2001**). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," J. Acoust. Soc. Am. **110**, 464–478.

Age, sex, and vowel dependencies of acoustic measures related to the voice source[a)]

Markus Iseli,[b)] Yen-Liang Shue,[c)] and Abeer Alwan[d)]
Department of Electrical Engineering, University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90095

(Received 22 February 2006; revised 23 January 2007; accepted 24 January 2007)

The effects of age, sex, and vocal tract configuration on the glottal excitation signal in speech are only partially understood, yet understanding these effects is important for both recognition and synthesis of speech as well as for medical purposes. In this paper, three acoustic measures related to the voice source are analyzed for five vowels from 3145 CVC utterances spoken by 335 talkers (8–39 years old) from the CID database [Miller *et al.*, Proceedings of ICASSP, 1996, Vol. **2**, pp. 849–852]. The measures are: the fundamental frequency (F_0), the difference between the "corrected" (denoted by an asterisk) first two spectral harmonic magnitudes, $H_1^*-H_2^*$ (related to the open quotient), and the difference between the "corrected" magnitudes of the first spectral harmonic and that of the third formant peak, $H_1^*-A_3^*$ (related to source spectral tilt). The correction refers to compensating for the influence of formant frequencies on spectral magnitude estimation. Experimental results show that the three acoustic measures are dependent to varying degrees on age and vowel. Age dependencies are more prominent for male talkers, while vowel dependencies are more prominent for female talkers suggesting a greater vocal tract-source interaction. All talkers show a dependency of F_0 on sex and on F_3, and of $H_1^*-A_3^*$ on vowel type. For low-pitched talkers ($F_0 \leqslant 175$ Hz), $H_1^*-H_2^*$ is positively correlated with F_0 while for high-pitched talkers, $H_1^*-H_2^*$ is dependent on F_1 or vowel height. For high-pitched talkers there were no significant sex dependencies of $H_1^*-H_2^*$ and $H_1^*-A_3^*$. The statistical significance of these results is shown.
 [DOI: 10.1121/1.2697522]

PACS number(s): 43.70.Gr [BHS] Pages: 2283–2295

I. INTRODUCTION

For almost half a century, research has been conducted on the nature of the glottal voice source signal, and glottal source parameters have been estimated using various procedures and algorithms. In the past, the study of the voice source signal has mainly centered on voice synthesis and speech coding applications. However, recent studies (Fant *et al.*, 2000; Sluijter and Van Heuven, 1996; Sluijter *et al.*, 1997) have shown that a relationship exists between the characteristics and/or parameters of the glottal voice source signal, and voice quality. A better knowledge of the relationship of acoustic measures that characterize the voice source with speaker properties such as sex and age, and with context or sound type such as vowel, would benefit the understanding of the human voice production mechanism and help improve voice analysis for a variety of speech processing and medical applications.

The human voice production mechanism can be roughly divided into three parts: lungs, vocal folds, and vocal tract. Air pressure from the lungs causes air to flow through the glottis, which is the airspace between the vocal folds. In voiced speech the vocal folds open and close quasiperiodically and thus convert the glottal air flow (air volume velocity) into a train of flow pulses, called the voice source excitation signal. This signal then passes through the vocal tract, which functions as an acoustic filter that shapes the spectrum of the sound, and at the end of the vocal tract the volume velocity signal is modified by the lip impedance. The speech pressure waveform measured in front of the lips can be approximated by the time derivative of the volume velocity signal (Rabiner and Schafer, 1978). This radiation effect is typically included in the source function, i.e., the source signal is modeled as the derivative of the glottal flow volume velocity. Sounds produced with nonvibrating vocal folds, such as in the fricative /f/, are called unvoiced sounds and are not studied in this paper.

Here, we use the linear source-filter model of speech production (Fant, 1960), in which the derivative of the glottal flow volume velocity acts as the source, sometimes also referred to as the excitation, and the vocal tract acts as the linear filter. The fact that source and filter are assumed to be independent of each other is one reason that this is the simplest and commonly used model for speech production. Early models of the source signal used a simple impulse train for modeling voiced signals. More recent studies model the shape of the glottal airflow or its derivative in the time-domain (Ananthapadmanabha, 1984; Fant *et al.*, 1985; Hedelin, 1984; Holmes, 1973; Klatt and Klatt, 1990; Rosenberg, 1971). Frequency-domain representations for some of those models were presented in Fant (1995) and Doval and d'Alessandro (1999). In this paper, the Liljencrants-Fant

[a)]Portions of this paper were presented at ICASSP04 and ICASSP06.
[b)]Electronic mail: iseli@ee.ucla.edu
[c)]Electronic mail: yshue@ee.ucla.edu
[d)]Electronic mail: alwan@ee.ucla.edu

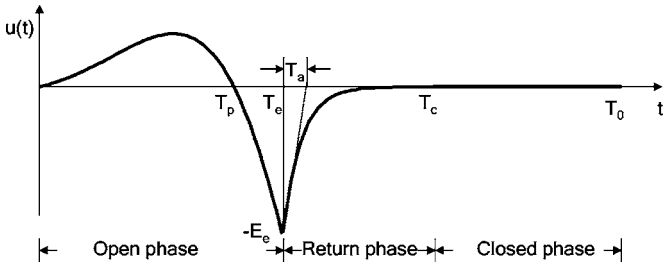


FIG. 1. The LF model and its parameters: instant of maximum airflow (T_p), instant of maximum airflow derivative (T_e), effective duration of return phase (T_a), beginning of closed phase (T_c), fundamental period (T_o), and amplitude of maximum excitation of glottal flow derivative (E_e).

(LF) model by Fant *et al.* (1985) is used to generate synthetic stimuli. It models the glottal volume velocity derivative, hence incorporating the effect of lip radiation, and is illustrated in Fig. 1. Vocal tract models, on the other hand, evolved from electric circuit models (Miller, 1959) and acoustic tube models (Fant, 1960) to all-pole autoregressive representations (Markel and Gray, 1976). For vowels, the vocal tract is typically modeled as an all-pole filter, where each complex-conjugate pole-pair represents a resonance frequency (formant) and its bandwidth.

To recover glottal source parameters from the acoustic speech signal, vocal tract resonances need to be removed by an "inverse filtering" process. The linear source-filter model assumes that the vocal tract is a linear filter and that it is independent and linearly separable from the source, all of which facilitates inverse filtering. Inverse filtering was first presented by Miller (1959), who applied analog electronic filters to cancel the two lowest formants and the lip radiation effect from the speech pressure waveform captured by a microphone. Rothenberg (1973) introduced a different inverse filtering technique that measures the airflow at the mouth and nose with a special mask. This method allows the estimation of absolute flow values, including the dc component, as opposed to the inverse filtering of the pressure signal captured by a microphone, which loses the absolute zero level of flow due to the lip radiation effect. The flow measurement mask is also less sensitive to low-frequency noise and the mask's frequencies are band limited at approximately 1.6 kHz (Hertegård and Gauffin, 1992). For all recording equipment, be it mask or microphone, it is important that its frequency magnitude response is flat and its phase response is linear from very low frequencies up to high frequencies. Compared to analog filtering, digital sampling, storage, and filtering techniques provide obvious advantages over analog techniques, since they are flexible, repeatable, easy to implement, and cause no phase distortion. Because of these advantages, today, digital inverse filtering methods are almost always used.

To find vocal tract filter parameters, typically a linear predictive coding based analysis is applied (Hertegård and Gauffin, 1992). However, more accurate results can usually be achieved with the method of discrete all-pole modeling (DAP) introduced by El-Jaroudi and Makhoul (1991). DAP uses a cost function which is based on the Itakuro-Saito distance evaluated at the discrete frequencies of the signal power spectrum. A recent publication which uses the DAP method in combination with a code book of source functions, generated with the LF model, and an iterative optimization algorithm is described in Fröhlich *et al.* (2001). These approaches to obtaining the glottal flow waveform are computationally expensive, and often need manual correction and tuning. Instead of trying to estimate the time domain parameters of the source models, researchers can study acoustic measures which are correlated with these parameters. This typically involves analyzing the harmonic frequencies in the speech spectrum, such as the magnitudes of the first two spectral harmonics of the source spectrum, located at the fundamental frequency F_0 and at $2F_0$, and the spectral magnitude of various formant peaks. This is less computationally intensive and less prone to error than finding the glottal flow waveform, and is therefore suited for analyzing the extensive amount of data needed for a reliable statistical evaluation. Spectral harmonics, however, are affected by both the source characteristics and by vocal tract resonances (formants). Hence, if one needs only to characterize the source signal properties, then the influence of vocal tract resonances, or formant frequencies, need to be compensated for (Fant, 1982, 1995; Hanson, 1995; Mártony, 1965). The correction in this paper is done using both formant frequencies and their bandwidths (Iseli and Alwan, 2004). The formula can be applied to voices produced with high fundamental frequency and/or low first formant frequency.

Holmberg *et al.* (1995) showed that the difference between the corrected (denoted by as asterisk hereafter) magnitudes of the first two harmonics ($H_1^*-H_2^*$) is correlated with the open quotient (OQ). On the other hand, Henrich *et al.* (2001) showed that $H_1^*-H_2^*$ is dependent on both OQ and glottal flow asymmetry. Hanson (1997) found that $H_1^*-A_3^*$, where A_3^* is the corrected spectrum level at the frequency of the third formant, is correlated with the source spectral tilt. The correction accounted for the first two formants (F_1 and F_2) and the bandwidth of the third formant (F_3), and tokens were normalized with respect to a neutral vowel. In addition, Hanson and Chuang (1999) showed that the acoustic characteristics of the glottal excitation signal are gender dependent. Their study compared the effects of gender on voice source parameters for about 21 adult male and adult female talkers for the three vowels /eh/, /ae/, and /ah/. It analyzed three acoustic cues—open quotient (as shown by the magnitude difference between the first two spectral harmonics), first formant bandwidth, and source spectral tilt (as shown by the difference between the magnitude of the first spectral harmonic and the corrected spectrum level at F_3)—and showed that open quotient and source spectral tilt are generally higher for adult female than for adult male talkers. Speech acoustics are also affected by age, which was shown in a study by Lee *et al.* (1999). It analyzed fundamental frequency (F_0) and formant frequencies for a large speech database (Miller *et al.*, 1996) with about 490 subjects in the age range of 5–50 years. The study showed that children have higher F_0 and formant frequencies, and greater temporal and spectral variability than adults. These findings are attributed to vocal-tract anatomical differences and possible differences in the ability to control speech articulators.

This paper has two specific aims. The first is to introduce and evaluate, through error analysis, a spectral magnitude correction formula, which uses both bandwidth and frequency estimates of the resonant frequencies of the vocal tract. This formula can be used to reliably estimate acoustic measures related to the voice source signal, such as the difference between the magnitude of the first two source spectral harmonics. The second aim is to use the correction formula to uncover age, sex, and vowel dependencies for the source parameter F_0 (fundamental frequency) and two acoustic measures: $H_1^*-H_2^*$ (related to open quotient), and $H_1^*-A_3^*$ (related to source spectral tilt). The dependencies are analyzed using speech signals recorded from 335 people (185 males, 150 females) in ten age groups from the CID database (Miller *et al.*, 1996).

The paper is organized as follows: In Sec. II a spectral magnitude correction formula is presented and its accuracy is evaluated through error analysis. Results on age, sex, and vowel dependencies of the three acoustic measures $(F_0, H_1^*-H_2^*, H_1^*-A_3^*)$ are presented in Sec. III. A summary in Sec. IV concludes the paper.

II. CORRECTION FORMULA AND ERROR ANALYSIS

In the following, a spectral magnitude correction formula, which uses both bandwidth and formant frequencies, is presented and evaluated. The formula can be used to reliably estimate acoustic measures related to the voice source signal such as the difference between the magnitude of the first two spectral harmonics.

A. A correction formula to compensate for the effects of formant frequencies in the speech spectrum

The spectral magnitude of the speech signal is the result of interactions from both the voice source and the vocal tract. The spectral magnitude formant correction formula (Iseli and Alwan, 2004), which requires no explicit inverse-filtering techniques, assumes the linear source-filter model of speech production (Fant, 1960) and is derived in the Appendix. The purpose of this correction formula is to "undo" the effects of the formants on the magnitudes of the source spectrum. This is done by subtracting the amount by which the formants boost the spectral magnitudes. Theoretically, if the formant frequencies and their respective bandwidths were known exactly and the linear source-filter model is applicable, then the corrected spectral magnitudes should represent the actual magnitudes of the source spectrum. For example, the corrected magnitude of the first spectral harmonic located at frequency ω_0, where $\omega_0=2\pi F_0$ and F_0 is the fundamental frequency, is given by

$$H^*(\omega_0) = H(\omega_0) - \sum_{i=1}^{N} 10\log_{10}\frac{(1-2r_i\cos(\omega_i)+r_i^2)^2}{(1-2r_i\cos(\omega_0+\omega_i)+r_i^2)(1-2r_i\cos(\omega_0-\omega_i)+r_i^2)} \quad (1)$$

with $r_i=e^{-\pi B_i/F_s}$ and $\omega_i=2\pi F_i/F_s$ where F_i and B_i are the frequencies and bandwidths of the ith formant, F_s is the sampling frequency, and N is the number of formants to be corrected for. $H(\omega_0)$ is the magnitude of the first harmonic from the speech spectrum and $H^*(\omega_0)$ represents the corrected magnitude and should coincide with the magnitude of the source spectrum at ω_0. Note that all magnitudes are in decibels. A less general form of this equation ($N=2$) is used in Sec. III, where only the first two formants are corrected for.

B. Error analysis of the correction method

To evaluate the accuracy of the correction formula (with and without bandwidth information) in estimating harmonic spectral magnitudes, error analysis is performed. In Secs. II B 1 and II B 2 error analysis is done using synthetic single-, and three-formant vowels, respectively. Specifically, error analysis is evaluated for the H_1-H_2 parameter. For the synthetic stimuli, the LF voice source signal is filtered with an all-pole model of the vocal tract. The LF shape is defined by $T_p=0.48$, $T_e=0.6$, and $T_a=0.05$, with $T_c=T_o=1$.

Analysis errors are calculated without using correction (NoC); with correction for the influence of only the first formant, F_1, without using bandwidth information, that is, by setting B_1 in Eq. (1) to zero, (F1noB1); and correction for the influence of F_1 using exact bandwidth information (F1B1). It is important to note that when $\omega=\omega_i$ for the F1noB1 case ($B_i=0$), the correction yields an infinite value [see Eq. (A5)].

1. Error analysis for single-formant synthetic signals

Formant correction is applied to single-formant synthetic signals with F_0 varying between 100 and 300 Hz, and F_1 between 200 and 800 Hz with constant bandwidth (B_1) of 75 Hz. Since the signals are synthetic, the actual values for H_1 and H_2 are known and the correction error between the actual and estimated harmonics' magnitudes can be calculated.

Figure 2 compares the H_1-H_2 error at $F_0=250$ Hz for the cases NoC, F1noB1, and F1B1. Maximum errors for NoC and F1noB1 occur at $F_1=F_0$ and $F_1=2F_0$, where the absolute NoC error is about 24 dB and the F1noB1 error is infinite. The error for F1B1 is zero, which is expected.

2. Error analysis for three-formant synthetic vowels

The vowels /a/, /i/, and /u/, are synthesized using the first three formant frequencies specified in Peterson and Barney (1952). Formant bandwidths are calculated according to the formula in Mannell (1998):

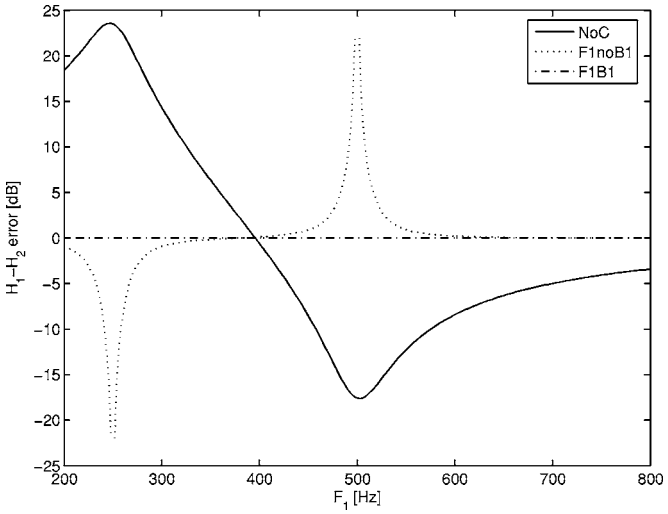


FIG. 2. H_1-H_2 error in decibels with $F_0=250$ Hz and $B_1=75$ Hz for synthetic one-formant signals. The three curves represent: NoC, no correction (solid line); F1noB1, correction for F_1 not using bandwidth information (dotted line); and F1B1, correction for F_1 using exact bandwidth information (dash-dotted line). The maximum NoC error is about 24 dB. The absolute error for the F1noB1 correction at $F_1=F_0$ and $F_1=2F_0$ is infinite, and the F1B1 error is zero.

$$B_i = (80 + 120F_i/5000). \quad (2)$$

These values are shown in Table I.

F_0 is chosen from the ranges provided by Baken (1987): For male talkers, F_0 ranges between 85 and 154 Hz, for female talkers F_0 is between 164 and 256 Hz, and for children F_0 is between 208 and 256 Hz. The sampling frequency (F_s) is at 10 kHz.

For each sex, vowel, and correction method, the minimum, average, and maximum absolute estimation errors for $|H_1-H_2|$ are calculated over the appropriate range of F_0. The results are shown in Table II. F1noB1 introduces the highest errors especially when F_1 is close to F_0 or $2F_0$. For the vowel /a/, on the other hand, F1noB1 performs similarly to F1B1 because /a/ has a very high F_1, which is greater than $2F_0$, and hence, the influence of F_1 on the first two harmonics is small. The errors for F1B1 are lower but are not zero since F1B1 does not correct for F_2 and F_3. The highest F1B1 errors are measured for /u/, which has the lowest F_2 of the three vowels.

TABLE II. Min/Mean/Max $|H_1-H_2|$ error in decibels without correction (NoC), correction for F_1 without bandwidth information (F1noB1), and correction for F_1 using bandwidth information (F1B1). Synthesis included three formants. As a reference, F_1 is given in parentheses for each of the vowels. It can be seen that the errors for NoC and F1noB1 are high when F_1 is close to F_0 or $2F_0$. The error for F1noB1 where $F_1=F_0$ or $F_1=2F_0$ is infinite.

Vowel (F_1 in Hz)	Min/Mean/Max Error in decibels		
	NoC	F1noB1	F1B1
	Male talkers (F_0: 85–154 Hz)		
/a/ (730)	0.57/1.06/1.99	0.20/0.38/0.69	0.20/0.38/0.69
/i/ (270)	3.04/5.58/8.15	0.41/∞/∞	0.07/0.13/0.23
/u/ (300)	2.66/5.61/9.67	0.00/∞/∞	0.30/0.56/1.04
	Female talker (F_0: 164–256 Hz)		
/a/ (850)	1.71/2.84/4.73	0.64/1.02/1.63	0.63/1.02/1.63
/i/ (310)	0.14/5.31/12.20	0.08/1.90/7.82	0.21/0.33/0.52
/u/ (370)	0.05/7.29/11.47	0.03/∞/∞	0.98/1.62/2.67
	Child talkers (F_0: 208–256 Hz)		
/a/ (1030)	2.05/2.59/3.25	0.86/1.07/1.33	0.83/1.04/1.30
/i/ (370)	0.36/2.91/5.97	0.01/0.73/2.15	0.29/0.36/0.44
/u/ (430)	4.60/9.59/12.48	0.23/∞/∞	1.08/1.36/1.71

Figures 3 and 4 show the absolute $|H_1-H_2|$ error as a function of F_0 for the methods NoC, F1noB1, and F1B1 for synthetic /a/ and /u/ vowels, respectively. Figure 3 shows the error for the synthetic female /a/ ($F_1=850$ Hz) where correction without using bandwidth information (F1noB1) works well. As mentioned earlier, this is due to F_1 being much higher than F_0 or $2F_0$, hence, the first formant does not have a significant effect on the magnitudes of the first two harmonics. However, for the female /u/ (Fig. 4), bandwidth information becomes important in the correction since $F_1=2F_0=370$ Hz when $F_0=185$ Hz. Hence, F1B1 yields significantly better results than F1noB1.

Since it is difficult to estimate bandwidths accurately (Hanson and Chuang, 1999), we also compare these results with another case, F1B50, which applies the correction formula using a constant bandwidth, $B_1=50$ Hz. The average

TABLE I. Formant frequencies (Peterson and Barney, 1952) and bandwidths (Mannell, 1983) in Hertz used to synthesize the three corner vowels appropriate for male, female, and child talkers.

Vowel	F_1	F_2	F_3	B_1	B_2	B_3
			Male talker			
/a/	730	1090	2440	98	106	139
/i/	270	2290	3010	86	135	152
/u/	300	870	2240	87	101	134
			Female talker			
/a/	850	1220	2810	100	109	147
/i/	310	2790	3310	87	147	159
/u/	370	950	2670	89	103	144
			Children			
/a/	1030	1370	3170	105	113	156
/i/	370	3200	3730	89	157	170
/u/	430	1170	3260	90	108	158

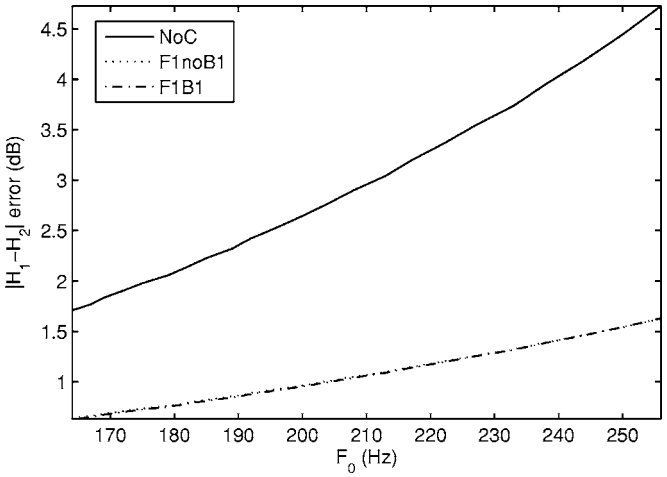


FIG. 3. $|H_1-H_2|$ error in decibels for a three-formant synthetic female /a/ (F_1=850 Hz, F_2=1220 Hz, F_3=2810 Hz) as a function of F_0. Error using NoC (solid line), with F1noB1 correction (dotted line), and with F1B1 (dash-dotted line). In this case, using bandwidth information is not critical since F_1 is much higher than $2F_0$.

absolute errors for the four cases NoC, F1noB1, F1B1, and F1B50, are shown in Fig. 5. It can be seen that the largest error occurs for the high back vowel /u/, since there is no correction for the low F_2. Using exact bandwidth information (F1B1) or using a fixed B_1 of 50 Hz improves significantly over F1noB1 for /i/ and /u/, which have low F_1. Interestingly, using a bandwidth estimate of 50 Hz (F1B50) yields similar results to using exact bandwidth information. These results imply that for reducing errors, it is better to use some bandwidth information, even if it is only an educated guess of the true bandwidth.

III. ESTIMATION OF ACOUSTIC MEASURES FOR NATURALLY PRODUCED SOUNDS

In the following we apply the correction formula to estimate age, sex, and vowel dependencies of three acoustic measures, F_0, $H_1^*-H_2^*$, $H_1^*-A_3^*$, on a relatively large speech database.

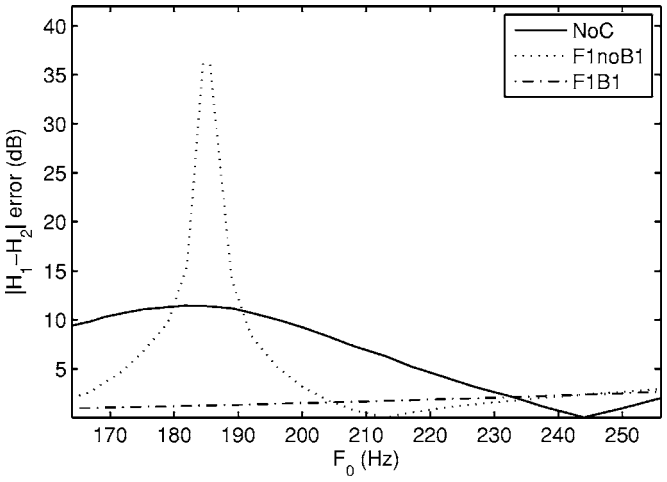


FIG. 4. $|H_1-H_2|$ error in decibels for a three-formant synthetic female /u/ (F_1=370 Hz, F_2=950 Hz, F_3=2670 Hz) as a function of F_0. Error using NoC (solid line), with F1noB1 correction (dotted line), and with F1B1 (dash-dotted line). F1B1 performed significantly better than F1noB1 since F_1 is quite low. The error for F1noB1 where $F_1=2F_0$ is infinite.

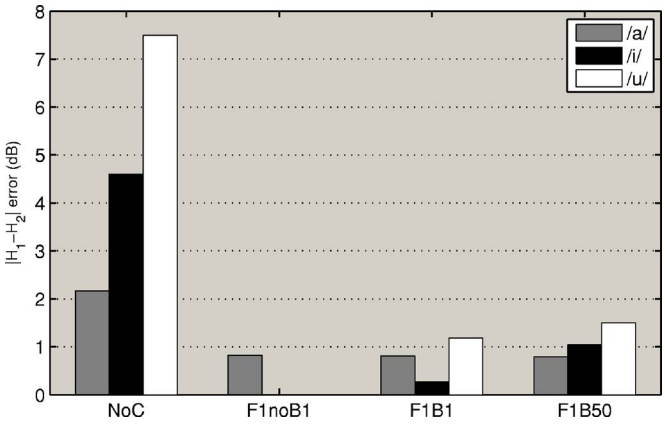


FIG. 5. (Color online) A bar diagram comparison of average $|H_1-H_2|$ error measurements for the three synthetic, three-formant vowels (averaged over both sexes, age groups, and corresponding F_0 values.) Results for NoC, F1noB1, F1B1, and F1B50, which is a correction for F_1 with B_1=50 Hz. No error bars are shown for F1noB1 for /i/ and /u/ since for some values of F_0 they can be infinite.

A. Speech data

Speech signals recorded from 335 people (185 males, 150 females) in ten age groups, ages 8, 9, 10, 11, 12, 13, 14, 15, 18, and 20–39, from the CID database (Miller *et al.*, 1996) were analyzed. The carrier sentence was "I say uh, bVt again," where the vowel was /ih/ (bit), /eh/ (bet), /ae/ (bat), and /uw/ (boot). "uh" was used before the target word to maximize vocal tract neutrality. The corner vowel /iy/ in "bead" was also analyzed. Most utterances were repeated twice by each speaker. Recordings were made at normal habitual speaking levels with a sampling frequency of 16 kHz. In total, 3145 utterances were analyzed. The age and sex distribution of the analyzed talkers is shown in Table III.

B. Methods

The voice source parameter F_0 and the acoustic cues $H_1^*-H_2^*$, and $H_1^*-A_3^*$ were estimated. As mentioned earlier, these measures are of significant importance in the areas of voice perception and voice synthesis (Fant and Kruckenberg, 1996; Holmberg *et al.*, 1995). $H_1^*-H_2^*$, the difference between the spectral magnitudes of the first two source harmonics, is related to the OQ (Holmberg *et al.*, 1995). $H_1^*-A_3^*$, the difference between the spectral magnitudes of the first harmonic and the third formant peak, is related to the source spectral tilt (Holmberg *et al.*, 1995). The asterisk denotes that spectral magnitudes (H_1,H_2,A_3) were corrected for the effects of formants. For H_1^* and H_2^*, the correction was for the first and second formant (F_1 and F_2) influence with

TABLE III. Number of analyzed talkers in each age group separated by sex (males: M; females: F).

Age	M	F	Age	M	F
8	25	11	13	16	13
9	24	25	14	11	10
10	25	14	15	11	11
11	24	19	18	10	10
12	22	21	20–39	17	16

$N=2$ in Eq. (A5). For A_3^*, the first three formants were corrected for ($N=3$) and there was no normalization to a neutral vowel; recall that our correction accounts for formant frequencies and their bandwidths.

The calculation of the three acoustic measures requires the estimation of the first three formant frequencies (F_1,F_2,F_3), their respective bandwidths (B_1,B_2,B_3), and F_0. Formant frequencies F_1, F_2, and F_3, as well as F_0 were estimated using the "SNACK SOUND TOOLKIT" software (Sjölander, 2004). The main parameters that can be changed in SNACK are frame length, frame shift, and analysis methods. For formant estimation, the covariance method was chosen because of its accuracy. F_0 can be extracted with either the ESPS (Entropic Signal Processing System), or the AMDF (Average Magnitude Difference Function (Ross *et al.*, 1974)) method. Both methods are based on conventional autocorrelation analysis. Since no significant estimation differences between the two methods were found, the ESPS method was used. Additional settings were: The preemphasis coefficient was 0.9, the length of the analysis window was 25 ms, and the window shift was 10 ms. Using the values extracted with SNACK, the amplitudes H_1, H_2, and A_3 were estimated from the speech spectrum. Since the SNACK bandwidth estimates varied greatly within the analysis segments and were sometimes unrealistic, all bandwidths were calculated from their corresponding formant frequency using Eq. (2). This reduced the bandwidth variance and therefore the variance of bandwidth-dependent results. Analysis segments were chosen at the steady-state part of the vowel, where the context influence was smaller than in other segments.

The estimates of F_0, F_1, F_2, and F_3 were manually checked for every utterance by viewing the spectrogram, time waveform, and LPC spectral slices. Most formant estimation errors occurred with child speech. For example, for high pitched /iy/, SNACK typically allocated two formants to the first spectral peak resulting in a much lower second formant frequency. The number of formant estimate corrections in percent, for 8 year old children, was: 86% for /iy/, 44% for /eh/, 32% for /ih/, and 2% for /uw/. The formant values are not listed here as the results are similar to those reported in Lee *et al.* (1999).

C. Results

In this section, we refer to males and females from ages 8 to 14, and females 15 years and older as "Group 1," and to male talkers age 15 and older as "Group 2." Group 1 talkers were typically high-pitched (with $F_0 > 175$ Hz) and Group 2 talkers were generally low-pitched (with $F_0 \leqslant 175$ Hz), although there were F_0 outliers within both groups as can be seen in the minimum/maximum F_0 values in Table VIII. The source parameter F_0, and acoustic measures $H_1^*-H_2^*$ and $H_1^*-A_3^*$ were analyzed as a function of age, sex, and vowel type, and their intercorrelations were studied.

1. Analysis of variance of the three acoustic measures

Statistical analysis was performed on the extracted acoustic measures by using the three-way analysis of variance (ANOVA) test in the software package SPSS (v13.0). The factors age (ages 8, 9, 10, 11, 12, 13, 14, 15, 18, and 20–39), sex (M, F) and vowel-type (/iy/, /ih/, /eh/, /ae/, and /uw/) were tested against the variables F_0, $H_1^*-H_2^*$ and $H_1^*-A_3^*$. These factors were tested with: (a) all the talkers, (b) the talkers separated by sex, and (c) the talkers separated into Group 1 and Group 2. Tests where the null hypothesis had a probability of $p<0.05$ were considered to be statistically significant. In addition, Pearson correlation coefficients were calculated to test for statistically significant intercorrelations between the three acoustic measures.

Table IV shows results for all the talkers for the F value (ratio of the model mean square to the error mean square) and partial η^2 (calculated as $SS_{effect}/(SS_{effect}+SS_{error})$, where SS_{effect} is the sum of squares of the effect and SS_{error} is the sum of squares of the error). Partial η^2 is a measure of effect size. For all three measures the effect size is greatest with age. For $H_1^*-H_2^*$ and $H_1^*-A_3^*$, the effect size of age is followed by vowel and sex, while for F_0, vowel type shows the smallest effect size.

Table V shows the ANOVA results when the talkers were separated by sex. It can be seen that across all three acoustic measures, the effect size of age is greater for males than for females. This was expected since speech acoustics, for example F_0 (Lee *et al.*, 1999), vary more significantly with age for male talkers. However, for vowel type, the effect size is greater for females than for males. This may suggest a greater vocal tract-source interaction for female talkers.

The results are also interesting when viewed in terms of the Group 1 (children and females, generally high-pitched) and Group 2 (older males, generally low-pitched) talkers. Table VI shows the F and partial η^2 results for Group 1 and Group 2 talkers. For Group 1 talkers, it can be seen that nearly all the entries are statistically significant except when sex is tested against $H_1^*-H_2^*$ and $H_1^*-A_3^*$. This result is interesting, since it suggests that females of all age groups have a similar OQ and source spectral tilt compared to boys (ages 8–14). More notable are the results for the Group 2 talkers which only have one significant entry: vowel type versus $H_1^*-A_3^*$. The lack of any age effect for Group 2 talkers is

TABLE IV. ANOVA results for all talkers showing F and partial η^2 values (in parentheses). All entries are statistically significant.

	F_0	$H_1^*-H_2^*$	$H_1^*-A_3^*$
Age	235.0 (0.410)	23.9 (0.066)	35.0 (0.094)
Sex	1012.3 (0.250)	57.7 (0.019)	4.1 (0.001)
Vowel	28.0 (0.036)	52.7 (0.065)	68.9 (0.083)

TABLE V. ANOVA results for female and male talkers showing F and partial η^2 values (in parentheses). All entries are statistically significant.

	F_0	$H_1^*-H_2^*$	$H_1^*-A_3^*$
Age (F)	26.4 (0.145)	2.8 (0.018)	8.8 (0.058)
Age (M)	310.3 (0.0618)	30.7 (0.138)	32.4 (0.144)
Vowel (F)	19.2 (0.052)	48.2 (0.121)	38.2 (0.098)
Vowel (M)	4.8 (0.011)	16.6 (0.037)	35.8 (0.076)

TABLE VI. ANOVA results for Group 1 (children and females) and Group 2 (older males) talkers showing F and partial η^2 values (in parentheses) for statistically significant entries. "⋯" denotes a nonsignificant entry. Sex is not included in the analysis for Group 2 since that group contains only male talkers.

	F_0	$H_1^*-H_2^*$	$H_1^*-A_3^*$
	Group 1		
Age	78.7 (0.208)	3.9 (0.013)	17.2 (0.054)
Sex	167.9 (0.059)	⋯	⋯
Vowel	26.1 (0.037)	75.9 (0.101)	65.1 (0.088)
	Group 2		
Age	⋯	⋯	⋯
Vowel	⋯	⋯	6.5 (0.069)

likely due to the fact that source characteristics for males do not change significantly with age above 15 years old; this has been shown for F_0 in Lee *et al.* (1999). Sex was not included for the Group 2 analysis since all the talkers in that group were male.

Table VII shows the Pearson correlation coefficients (PCCs) when the three acoustic measures were tested against each other. Although the intercorrelations are statistically significant, there is only one PCC greater than 0.7, indicating a strong correlation. This occurs for the relationship between $H_1^*-H_2^*$ and F_0 for Group 2 talkers.

2. F_0

Table VIII shows the range of F_0 values for all talkers. Note that F_0 was not normalized for stress. For males the mean F_0 drops by about 130 Hz between ages 8 and 20 with the largest drop between ages 12 and 15 (105 Hz), while the change is less dramatic for female talkers (overall about 50 Hz). These changes are reflected in Table V which shows that age has a greater effect size on F_0 for males (F/partial $\eta^2=310.3/0.618$) than for females (F/partial $\eta^2=26.4/0.145$). As expected, adult females exhibit higher F_0 values than adult male talkers: The difference in the means is about 110 Hz. These trends agree with the results in Lee *et al.* (1999). We noticed that a few very high F_0 values (above 300 Hz) were due to high stress on the target word. In those cases, F_0 was around 300 Hz for the rest of the sentence, but increased for the target word.

Average F_0 values are highest for /uw/, and higher for /iy/ than for /eh/ and /ae/. The trend of increasing F_0 as the tongue moves from a front to a back position and from open to closed vowels has been reported for German talkers by Marasek (1996). This trend can be seen for all ages and genders for the vowels in this study and may partly be explained by vowel-dependent intrinsic pitch (Lehiste and Peterson, 1961). ANOVA results in Table V indicate that although these trends are statistically significant for both males and females, the partial η^2 values, and hence the effect sizes of vowel type, are relatively small for both sexes: F/partial $\eta^2=19.2/0.052$ for females and $4.8/0.011$ for males. Interestingly, the vowel effect size on F_0 is five times higher for females. A further analysis into the vowel dependency was done by performing an ANOVA test on the effects of high and low formant frequencies (thresholds at the formant means) on F_0. It was found that F_0 was positively correlated only with F_3 for all talkers and this correlation was statistically significant (F/partial $\eta^2=133.1/0.041$); again the effect size was relatively small. This positive correlation can be explained by the fact that F_3 is typically correlated with vocal tract length (Wakita, 1977). Hence, a higher F_3, which typically results from a shorter vocal tract, coincides with a higher F_0.

TABLE VII. Pearson correlation coefficients (PCCs) for F_0, $H_1^*-H_2^*$ and $H_1^*-H_3^*$ for Group 1 and Group 2 talkers. Correlation coefficients greater than 0.7 indicate strong correlations. All results are statistically significant.

	F_0	$H_1^*-H_2^*$	$H_1^*-A_3^*$
	Group 1		
F_0	1	−0.471	−0.356
$H_1^*-H_2^*$	−0.471	1	0.532
$H_1^*-A_3^*$	−0.356	0.532	1
	Group 2		
F_0	1	0.767	0.268
$H_1^*-H_2^*$	0.767	1	0.473
$H_1^*-A_3^*$	0.268	0.473	1

TABLE VIII. Min/Mean/Max of F_0 (in Hz) per age group for vowels in the target syllables.

Age	F_0 males (Hz)	F_0 females (Hz)
8	170/255/420	152/283/423
9	160/264/454	187/267/437
10	141/256/407	146/266/367
11	167/256/378	185/254/494
12	125/230/328	178/236/338
13	119/190/285	180/251/394
14	101/177/272	169/228/293
15	95/125/251	179/228/310
18	84/129/239	199/246/310
20–39	88/127/191	156/235/356

3. $H_1^*-H_2^*$

The effects of age and sex on $H_1^*-H_2^*$ (related to open quotient) are shown in Fig. 6. Comparing the values, it is interesting to observe that the $H_1^*-H_2^*$ (mean value) separation between the genders is the clearest at age 15 (5.8 dB). Between ages 8 and 20–39, the mean $H_1^*-H_2^*$ value drops by about 4 dB for male talkers, whereas for female talkers it remains relatively unchanged. Having smaller changes in $H_1^*-H_2^*$ with age is reflected in the statistical analysis of Table V where the effects of age are less pronounced for females: F/partial $\eta^2=2.8/0.018$ vs $30.7/0.138$ for male talkers. The difference between genders may be related to the fact that F_0 drops significantly between age 12 and 15 for males while it does not change as much for females (Lee *et al.*, 1999). Adult females exhibit higher mean $H_1^*-H_2^*$ values (about 3.4 dB) than adult male talkers. A similar difference (3.1 dB) between adult male and adult female talkers was found in Hanson and Chuang (1999). When the talkers are split into Group 1 and Group 2 categories (see Table VI), it is interesting to note that the dependence on sex is not significant for Group 1 talkers (children and females).

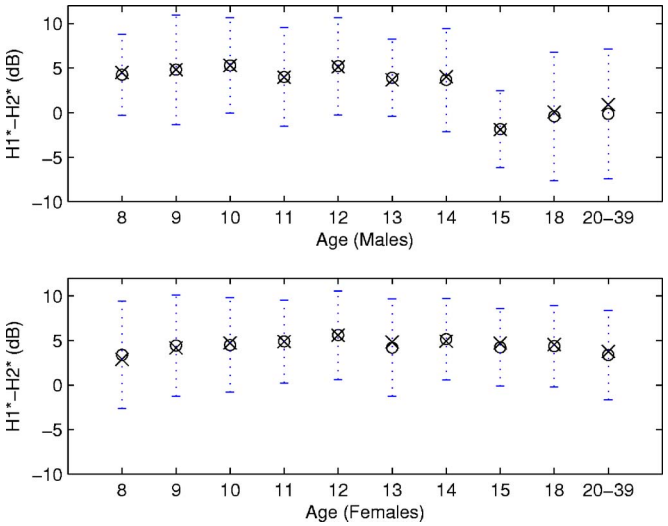


FIG. 6. (Color online) $H_1^*-H_2^*$ vs age, separated by sex. Between age 8 and 20–39, $H_1^*-H_2^*$ drops by about 4 dB for males, while for females there is little change. The largest difference between the sexes appears at age 15 where the difference in the means approaches 6 dB. Mean, median, and standard deviation are represented by circles, crosses, and whiskers, respectively.

Vowel effects are larger for female talkers than for males as shown in Table V (F/partial η^2=48.2/0.121 for females vs 16.6/0.037 for males). When analyzed against Group 1 and Group 2, the results in Table VI indicate that only Group 1 talkers exhibit a dependence on vowel (F/partial η^2 =75.9/0.101) whereas Group 2 (older male) talkers do not exhibit a significant dependence on vowel or on age.

ANOVA tests were also done to study the effects of formant values (thresholds at the formant means). The only statistically significant result is for F_1 with Group 1 talkers (F/partial η^2=91.4/0.034). No significant correlation between $H_1^*-H_2^*$ and F_1 (vowel height) can be observed for Group 2, or can a correlation with F_2 and F_3 be shown for any group. This effect can be seen in Fig. 7, which depicts

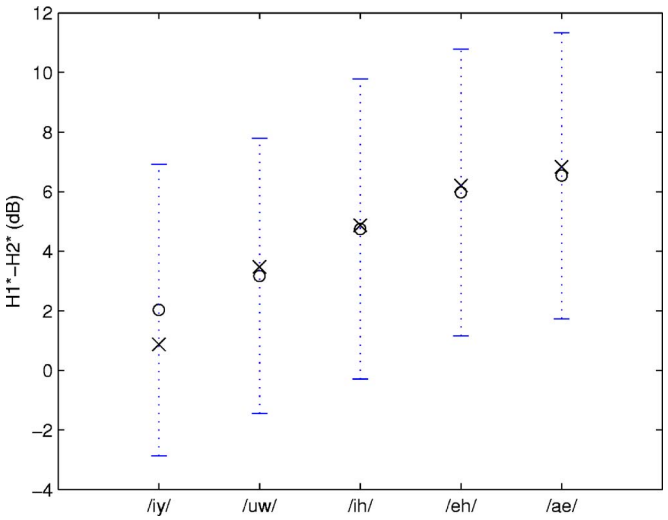


FIG. 7. (Color online) $H_1^*-H_2^*$ as a function of vowel for Group 1 talkers (females and children). Vowels are sorted according to their F_1 value from low to high. Note that the lowest values occur for the high and tense vowels /iy/ and /uw/.

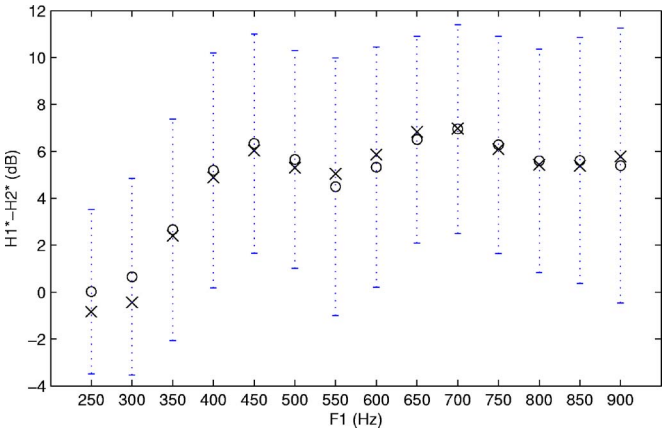


FIG. 8. (Color online) $H_1^*-H_2^*$ vs F_1 for Group 1 talkers. $H_1^*-H_2^*$ monotonically increases, on average, by about 6 dB when F_1 increases between 250 and 450 Hz.

$H_1^*-H_2^*$ as a function of vowel for the Group 1 talkers. Vowels are sorted from left to right as a function of their average F_1 value. $H_1^*-H_2^*$ values for /iy/ and /uw/ are the lowest, suggesting that high vowels have lower OQ. As F_1 increases for /iy/, /uw/, /ih/, /eh/, and /ae/, $H_1^*-H_2^*$ becomes larger. Figure 8 shows $H_1^*-H_2^*$ as a function of F_1 and agrees with Fig. 7 trends. Hanson (1997) showed that, for adult female voices, the mean value of $H_1^*-H_2^*$ was slightly lower for /eh/ than /ae/ which agrees with our results.

The lack of significant trends of $H_1^*-H_2^*$ values with F_1 for Group 2 talkers may be due to the physiology associated with voice production in different genders. This difference could be due to increased vocal tract-source interaction when F_0 or its harmonics are close to F_1 (Titze, 2004), which is often the case for low F_1 and high F_0.

For both sexes $H_1^*-H_2^*$ for /iy/ is about 3 dB lower than for /ih/. This could be due to the tense/lax difference. For four minority languages in China, Maddieson and Ladefoged (1985) reported that the amplitude difference between the first two harmonics was smaller for tense vowels than lax ones, which would agree with our findings.

Relationship of $H_1^*-H_2^*$ with F_0 and $H_1^*-A_3^*$

Figure 9 shows the relationship between $H_1^*-H_2^*$ and F_0 for both groups. As can be seen in Table VII, the PCC between $H_1^*-H_2^*$ and F_0 yields a value of 0.767 for Group 2 and a weak negative correlation (PCC=−0.471) for Group 1. An approximate mapping for $H_1^*-H_2^*$ and F_0 for Group 2 is

$$H_1^* - H_2^* \approx 0.22F_0 - 28 \quad \text{for } F_0 \text{ between 80 and 175 Hz.} \tag{3}$$

A possible interpretation for this result is that the Group 1 talkers (females and children, generally high-pitched) and the Group 2 talkers (older males, generally low-pitched) use OQ differently during the phonation of vowels. In a study by Esposito (2005) utilizing electroglottography of Zapotec talkers, females were shown to produce phonation differences by altering OQ while males did not. It has also been observed in Koreman (1996) that increased tension of the cricothyroid muscle in the larynx induces a simultaneous increase of F_0 and OQ, and therefore also of $H_1^*-H_2^*$. However,

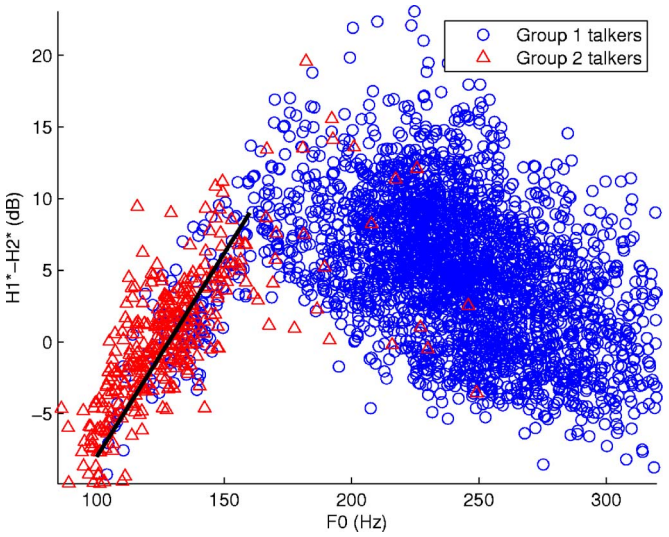


FIG. 9. (Color online) $H_1^*-H_2^*$ vs F_0 for Group 1 and Group 2 talkers. A linear relationship for F_0 between 80 and 175 Hz is observed.

we observed a strong positive correlation only for low F_0 values. Swerts and Veldhuis (2001) also found similar results for some of their speakers.

As seen in Table VII, the intercorrelation between $H_1^*-H_2^*$ and $H_1^*-A_3^*$ for both groups is weak: 0.532 (Group 1), 0.473 (Group 2). A weak correlation was also reported in Hanson (1997) for adult female talkers.

4. $H_1^*-A_3^*$

The age and sex effects on $H_1^*-A_3^*$ (related to source spectral tilt) are shown in Fig. 10. Between ages 8 and 20–39, the mean $H_1^*-A_3^*$ value drops for male talkers by about 10 dB, whereas for female talkers it drops by about 4 dB resulting in higher values (by about 4 dB) for adult females than for adult males. The higher effect size for males (F/partial η^2=32.4/0.144) compared to females

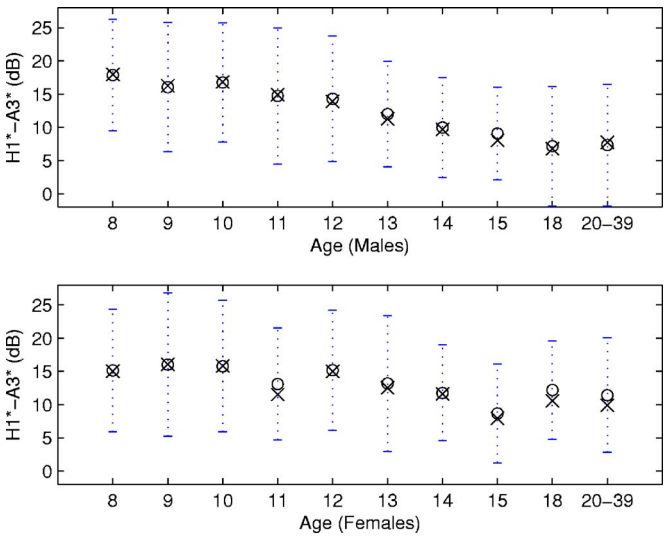


FIG. 10. (Color online) $H_1^*-A_3^*$ vs age; the top panel represents data for male talkers and the lower panel represents data for female talkers. For both sexes there is a drop of $H_1^*-A_3^*$ between age 8 and age group 20–39: The drop is about 4 dB for females, and 10 dB for males.

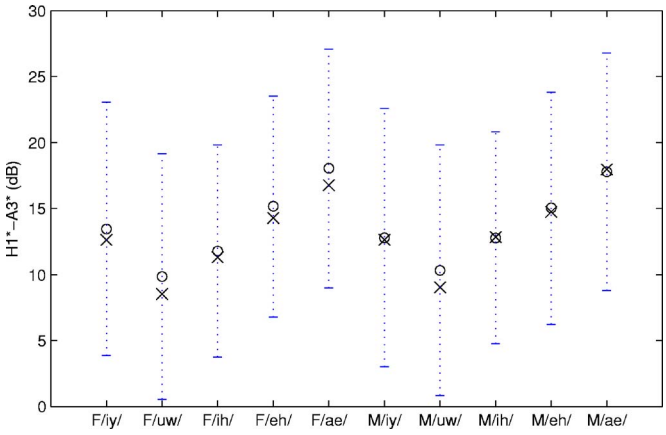


FIG. 11. (Color online) $H_1^*-A_3^*$ as a function of vowel for all talkers; M and F indicate data from male and female talkers, respectively. /ae/ and /eh/ have the highest values, while /uw/ has the lowest value. This result might be related to the dependence of the parameter on formant values.

(F/partial η^2=8.8/0.058) in Table V confirms this result. When the talkers are split into groups (see Table VI), Group 1 shows a dependence on age (F/partial η^2=17.2/0.054), whereas Group 2 does not. It is also interesting to note that the dependence on sex is not significant for Group 1. These trends are similar to those shown for $H_1^*-H_2^*$ (see Sec. III C 3), thus they can be interpreted similarly. That is, females (children and adults) and young males (8–14 years old) exhibit statistically similar OQ and source spectral tilt characteristics.

In Fig. 11, $H_1^*-A_3^*$ is depicted as a function of vowel and sex. The largest difference is observed between the vowels /ae/ and /uw/ where /ae/ is a low front vowel (high F_1, high F_2) and /uw/ is a high back vowel (low F_1, low F_2). Values for $H_1^*-A_3^*$ for /ae/ and /eh/ are the highest, and for /uw/ they are the lowest. These trends are similar for both sexes and indeed it can be seen from ANOVA analysis that the effect sizes of vowel are similar when male talkers are compared with females (Table V).

To find the effects of formants on $H_1^*-A_3^*$, an ANOVA analysis based on high and low values of F_1, F_2, and F_3 (thresholds at the formant means) yielded F/partial η^2 values of 210/0.063, 42.7/0.013, and 100.0/0.031, respectively. Thus, the first three formants have an effect on $H_1^*-A_3^*$ for all talkers. To visualize these effects, Figs. 12–14 show $H_1^*-A_3^*$ gradually rising for increasing F_1, F_2, and F_3. Since /uw/ on average has lower F_2 and F_3 compared to the other vowels used in this study, this can explain why $H_1^*-A_3^*$ values for /uw/ are lowest.

The dependency of $H_1^*-A_3^*$ on F_1 is somewhat similar to the dependency of $H_1^*-A_3^*$ on $H_1^*-A_1$ (related to F_1) which was observed in Hanson and Chuang (1999). The dependency of the measure on F_2 and F_3 was expected since a high F_2 is normally associated with a high F_3, which in term will affect the source spectral tilt. Since A_3^* represents the magnitude of the source spectrum at F_3, it is affected by the position of F_3 due to the source spectral tilt. A_3^* can also be influenced by the presence of higher formants, such as F_4, for which the parameter was not corrected for, and which would boost the value of A_3^* when evaluated close to F_4.

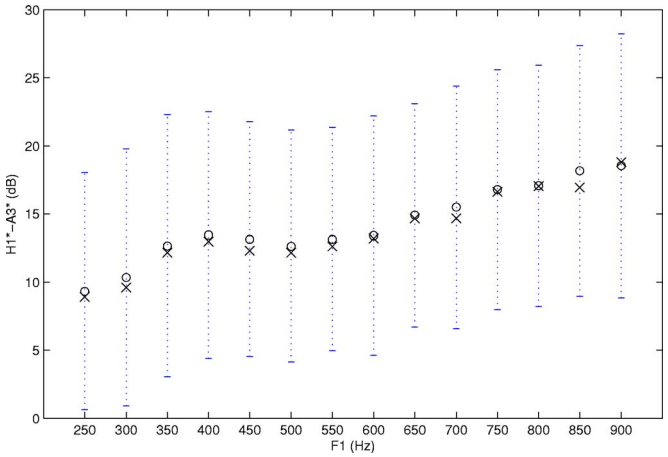


FIG. 12. (Color online) $H_1^*-A_3^*$ vs F_1 for all talkers. $H_1^*-A_3^*$ increases for increasing F_1.

IV. SUMMARY

In this paper the effects of age, sex, and vocal tract configuration on three acoustic measures related to voice source parameters: F_0, $H_1^*-H_2^*$, and $H_1^*-A_3^*$ are studied.

In order to estimate the acoustic measures $H_1^*-H_2^*$, and $H_1^*-A_3^*$, the vocal tract influence on the source spectrum needs to be compensated for. A correction formula which corrects for the influence of the vocal tract resonances is presented in Sec. II A. The importance of using the correction formula to estimate the magnitudes of the first two harmonics, H_1 and H_2, for vocal-tract influences, especially for high vowels and for high-pitched voices, is shown in Sec. II B. Synthetic speech is produced with formant frequencies from Peterson and Barney (1952) data and formant bandwidths are estimated from corresponding formant frequencies using Mannell's (1998) formula.

For synthetic speech, analysis errors are calculated without correction and with correction for the influence of only the first formant: (a) with bandwidth information, (b) without bandwidth information, and (c) with a bandwidth estimate of 50 Hz.

Error analysis results show that it is better to use an educated guess of formant bandwidth when correcting for the vocal tract influence, rather than using no bandwidth information [i.e., setting $B_i=0$ in Eq. (A5)] as in Hanson (1995). Examples of synthetic vowels show that correction without using bandwidth information can yield larger errors than no correction at all.

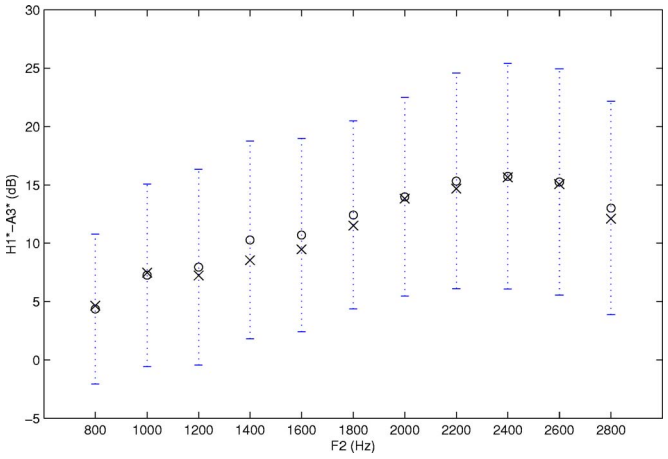


FIG. 13. (Color online) $H_1^*-A_3^*$ vs F_2 for all talkers. $H_1^*-A_3^*$ monotonically increases for F_2 increasing between 800 and 2400 Hz.

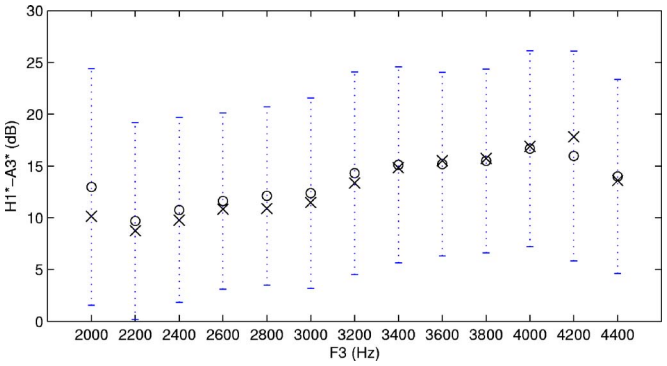


FIG. 14. (Color online) $H_1^*-A_3^*$ vs F_3 for all talkers. $H_1^*-A_3^*$ monotonically increases for F_3 increasing between 2200 and 4000 Hz.

The correction formula is then used to analyze acoustic measures related to source parameters for a relatively large speech database. The five vowels /iy/, /ih/, /eh/, /ae/, and /uw/ recorded from 335 people (185 males, 150 females) in ten age groups, ages 8, 9, 10, 11, 12, 13, 14, 15, 18, and 20–39, from the CID database (Miller *et al.*, 1996) are analyzed. F_0, as well as the formant frequencies, are extracted using the "SNACK SOUND TOOLKIT" software (Sjölander, 2004) and manually corrected if necessary. Bandwidth values are estimated from their corresponding SNACK formant frequencies again using Mannell's 1998 formula.

Statistical analysis of variance (ANOVA) is performed for all three acoustic cues and the three factors, age, sex, and vowel. These factors are tested with: (a) all talkers, (b) talkers separated by sex, and (c) talkers separated into Group 1 (children ages 8–14 and females ages 15 and older: generally high-pitched) and Group 2 (males ages 15 and older: generally low-pitched). In addition, where applicable, Pearson correlation coefficients are calculated for the different measurements. For Group 1, all effects are statistically significant except when sex is tested against $H_1^*-H_2^*$ and $H_1^*-A_3^*$. This result suggests that females of all age groups and boys (ages 8–14) have similar OQ and source spectral tilt values. For Group 2 the only significant result occurs when $H_1^*-A_3^*$ is tested against vowel type.

F_0 for male talkers drops between ages 8 and 20–39 (by about 130 Hz), whereas the overall drop for females is only about 50 Hz. F_0 is shown to be vowel dependent, with the highest values for /uw/, and higher for /iy/ than for /eh/ and /ae/. This trend may be attributed to intrinsic pitch. Furthermore, F_3 is shown to have a statistically significant relationship with F_0 which can be explained by the dependency of F_3 on vocal tract length.

$H_1^*-H_2^*$ (hence, the open quotient) is age dependent and for male talkers a drop by about 4 dB between the ages of 9 and 20–39 is found. For females, there is less dependency on age. On average, $H_1^*-H_2^*$ values are higher by about 3 dB for adult female compared to male talkers. There is no significant dependency on age and vowel for Group 2 talkers. H_1^*

TABLE IX. Summary of key results.

	Age (from 8 to 39 years old)		Vowel dependencies and intercorrelations
	Females	Males	
F_0	↓50 Hz	↓130 Hz	Linearly related to $H_1^*-H_2^*$ for low-pitched talkers, and to F_3 for all talkers
$H_1^*-H_2^*$	⋯	↓4 dB	Linearly related to F_0 for low-pitched talkers, and to F_1 for high-pitched talkers
$H_1^*-A_3^*$	↓4 dB	↓10 dB	Dependent on F_1, F_2, and F_3 for all talkers

$-H_2^*$ is proportional to F_0 for F_0 below 175 Hz. Above that frequency a weak negative correlation with F_0 could be found. For Group 1 talkers and for F_1 below 450 Hz, $H_1^*-H_2^*$ is proportional to F_1, resulting in low $H_1^*-H_2^*$ values for high vowels. For Group 2 talkers, on the other hand, no significant correlations between the $H_1^*-H_2^*$ values and vowel height could be observed. The different OQ dependencies between females and children (ages 8–14), and older males (ages 15 and older) could be due to physiological differences, to phonological differences, where females alter OQ to signal acoustic differences while males do not (Esposito, 2005), and/or to vocal tract-source interaction when F_0 or its harmonics are close to F_1 (Titze, 2004), which is often the case for low F_1 and high F_0 values. For both sexes $H_1^*-H_2^*$ for /iy/ is about 3 dB lower than for /ih/ which could be due to a tense/lax difference.

$H_1^*-A_3^*$ (hence source spectral tilt) values drop by about 10 dB between ages 8 and 20–39 for males, whereas for females the values drop by only about 4 dB within the same age period. This results in generally lower values for adult males (by about 4 dB) compared to adult females. Until age 10, the values are similar for both sexes. Statistical analysis shows a high dependence of the measure on age and vowel for all talkers. Also, $H_1^*-A_3^*$ shows a strong dependence on all formant frequencies for all talkers and age groups: Increasing F_1, F_2, or F_3 yields an increase in $H_1^*-A_3^*$. These findings imply that source spectral tilt is vowel dependent and, in fact, it can be seen that tilt values are highest for /ae/ and /eh/ and lowest for /uw/. The dependence of $H_1^*-A_3^*$ on F_3 can be explained by the dependence of A_3^* on F_3: Increasing F_3 will yield decreasing A_3^*.

Key dependencies are summarized in Table IX. Results show that all three acoustic measures are dependent to varying degrees on age and vowel. Age dependencies are more prominent for males than for females while vowel dependencies are more prominent for female talkers suggesting a greater vocal tract-source interaction. For $H_1^*-H_2^*$ vowel dependencies are only significant for Group 1 (generally high-pitched) talkers. F_0 shows a dependency on sex and on F_3, $H_1^*-H_2^*$ on F_1 (Group 1 talkers only), and $H_1^*-A_3^*$ on all three formants. For Group 2 (generally low-pitched) talkers F_0 is positively correlated with $H_1^*-H_2^*$.

The methods and results presented in this paper may contribute to a better understanding of speech production and may be useful for applications such as speech synthesis, speech recognition, and speaker identification.

Future work will study the effects of context, prosody, stress, and accentedness on acoustic measures related to source parameters.

ACKNOWLEDGMENTS

We thank Dr. Story and two anonymous reviewers for their helpful suggestions. This material is based upon work supported by NSF Grant No. 0326214 and by a Radcliffe Fellowship to Abeer Alwan. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

APPENDIX: DERIVATION OF THE CORRECTION FORMULA

The derivation of the spectral magnitude formant correction formula presented in this appendix is based on the linear source-filter model of speech production (Fant, 1960). Assuming a vocal tract all-pole model, the normalized transfer function $T(s)$ with N formants can be written as

$$T(s)=\prod_{i=1}^{N}\frac{\sigma_i^2+\Omega_i^2}{(s-(\sigma_i+j\Omega_i))(s-(\sigma_i-j\Omega_i))}. \tag{A1}$$

The numerator is chosen such that $T(s=0)=1$. $s_i=\sigma_i+j\Omega_i$, $\sigma_i=-\pi B_i$, $\Omega_i=2\pi F_i$, where B_i and F_i are the ith formant bandwidth and frequency, respectively.

Assuming that the axis $s=j\Omega$ lies in the region of convergence (ROC), the Fourier transform of the magnitude of Eq. (A1) becomes

$$|T(j\Omega)|=\prod_{i=1}^{N}\left|\frac{\sigma_i^2+\Omega_i^2}{\sigma_i^2+\Omega_i^2-\Omega^2+j2\sigma_i\Omega}\right|,$$

$$|T(j\Omega)|^2=\prod_{i=1}^{N}\frac{(\sigma_i^2+\Omega_i^2)^2}{(\sigma_i^2+\Omega_i^2-\Omega^2)^2+(2\sigma_i\Omega)^2}.$$

Using the definitions of σ_i and Ω_i produces

$$|T(f)|^2=\prod_{i=1}^{N}\frac{(\pi^2B_i^2+4\pi^2F_i^2)^2}{(\pi^2B_i^2+4\pi^2F_i^2-4\pi^2f^2)^2+16\pi^4B_i^2f^2}.$$

Finally, the total contribution of N formants to the vocal tract power spectrum at frequency f is

$$|T(f)|^2=\prod_{i=1}^{N}\frac{((B_i/2)^2+F_i^2)^2}{((B_i/2)^2+F_i^2-f^2)^2+B_i^2f^2}. \tag{A2}$$

Note: For $B_i \ll F_i$ the terms $(B_i/2)^2$ can be neglected (Fant, 1995). In this paper, however, we will account for these terms.

The aforementioned analysis was done in the continuous frequency domain. For sampled signals (sampling frequency F_s) the contribution of N formants to the vocal tract transfer function can be written in the z domain as

$$T(z) = \prod_{i=1}^{N} \frac{1 - 2\Re(z_i) + |z_i|^2}{(z - z_i)(z - z_i^*)}, \tag{A3}$$

where $T(z)$ is normalized so that $|T(z=1)|=1$. $z_i = r_i e^{j\omega_i}$ with $\omega_i = 2\pi F_i / F_s$.

Assuming that the unit circle $z=e^{j\omega}$ lies in the ROC, the Fourier transform of the squared magnitude of Eq. (A3) becomes

$$|T(\omega)|^2 = \prod_{i=1}^{N} \frac{(1 - 2r_i \cos(\omega_i) + r_i^2)^2}{(1 - 2r_i \cos(\omega - \omega_i) + r_i^2)(1 - 2r_i \cos(\omega + \omega_i) + r_i^2)}, \tag{A4}$$

with $r_i = e^{-\pi B_i / F_s}$ and $\omega_i = 2\pi F_i / F_s$.

Equation (A4) specifies the amount by which the spectral magnitude at a particular frequency, ω, is boosted by the effects of formants located at frequencies ω_i. Therefore, to obtain the source spectral magnitudes, the effects of the formants need to be subtracted from the magnitudes of the speech spectrum. For example (Iseli and Alwan, 2004),

$$H^*(\omega) = H(\omega) - \sum_{i=1}^{N} 10 \log_{10} \frac{(1 - 2r_i \cos(\omega_i) + r_i^2)^2}{(1 - 2r_i \cos(\omega + \omega_i) + r_i^2)(1 - 2r_i \cos(\omega - \omega_i) + r_i^2)}, \tag{A5}$$

where $H(\omega)$ is the magnitude of the original signal spectrum (in dB) at frequency ω, N is the number of formants, and $H^*(\omega)$ is the corrected magnitude (i.e., the magnitude of the source spectrum) at frequency ω. Note that for $B_i=0$ and $\omega=\omega_i$ this formula is undefined.

Ananthapadmanabha, T. V. (**1984**). "Acoustic analysis of voice source dynamics," STL-QPSR **25**, 1–24.

Baken, R. J. (**1987**). *Clinical Measurement of Speech and Voice* (Taylor and Francis, London).

Doval, B., and d'Alessandro, C. (**1999**). "The spectrum of glottal flow models," Technical Report, LIMSI-CNRS, Orsay, France.

El-Jaroudi, A., and Makhoul, J. (**1991**). "Discrete all-pole modeling," IEEE Trans. Signal Process. **39**, 411–423.

Esposito, C. (**2005**). "An acoustic and electroglottographic study of phonation in Santa Ana del Valle Zapotec," Poster at the 79th Meeting of the Linguistic Society of America, 2005.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, Paris).

Fant, G. (**1982**). "The voice source-acoustic modeling," STL-QPSR **23**, 28–48.

Fant, G. (**1995**). "The LF model revisited. Transformations and frequency domain analysis," STL-QPSR **36**, 119–156.

Fant, G., and Kruckenberg, A. (**1996**). "Voice source properties of speech code," TMH-QPSR **37**, 45–56.

Fant, G., Kruckenberg, A., Liljencrants, J., and Hertegård, S. (**2000**). "Acoustic-phonetic studies of prominence in Swedish," TMH-QPSR **41**, 1–52.

Fant, G., Liljencrants, J., and Lin, Q. (**1985**). "A four-parameter model of glottal flow," STL-QPSR **26**, 1–13.

Fröhlich, M., Michaelis, D., and Strube, H. W. (**2001**). "Sim-Simultaneous inverse filtering and matching of a glottal flow model for acoustic speech signals," J. Acoust. Soc. Am. **110**, 479–488.

Hanson, H. M. (**1995**). "Glottal characteristics of female speakers," Ph.D. dissertation, Harvard University, Cambridge, MA.

Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.

Hanson, H. M., and Chuang, E. S. (**1999**). "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**, 1064–1077.

Hedelin, P. (**1984**). "A glottal LPC-vocoder," in Proc. IEEE 1.6.1–1.6.4.

Henrich, N., d'Alessandro, C., and Doval, B. (**2001**). "Spectral correlates of voice open quotient and glottal flow asymmetry: Theory, limits and experimental data," in Proceedings of EUROSPEECH, Scandinavia, pp. 47–50.

Hertegård, S., and Gauffin, J. (**1992**). "Acoustic properties of the Rothenberg mask," STL-QPSR **33**, 9–18.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., and Goldman, S. L. (**1995**). "Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice," J. Speech Hear. Res. **38**, 1212–1223.

Holmes, J. N. (**1973**). "Influence of the glottal waveform on the naturalness of speech from a parallel formant synthesizer," IEEE Trans. Audio Electroacoust. 298–305.

Iseli, M., and Alwan, A. (**2004**). "An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation," in Proceedings of ICASSP, Montreal, Canada, Vol. **1**, pp. 669–672.

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.

Koreman, J. (**1996**). "Decoding linguistic information in the glottal airflow," Ph.D. thesis, University of Nijmegen.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of childrens speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.

Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–425.

Maddieson, I., and Ladefoged, P. (**1985**). "Tense and lax in four minority languages of China," J. Phonetics **13**, 433–454.

Mannell, R. H. (**1998**). "Formant diphone parameter extraction utilising a labelled single speaker database," in *Proceedings of the ICSLP* (ASSTA, Sydney, Australia), Vol. **5**, pp. 2003–2006.

Marasek, K. (**1996**). "Glottal correlates of the word stress and the tense-lax opposition in German," in Proceedings ICSLP, Philadelphia, PA, pp. 1573–1576.

Markel, J. D., and Gray, A. H., Jr. (**1976**). *Linear Prediction of Speech* (Springer, New York).

Mártony, J. (**1965**). "Studies of the voice source," STL-QPSR **6**, 4–9.

Miller, J., Lee, S., Uchanski, R., Heidbreder, A., Richman, B., and Tadlock, J. (**1996**). "Creation of two children's speech databases," in Proceedings of ICASSP, Vol. **2**, pp. 849–852.

Miller, R. L. (**1959**). "Nature of the vocal cord wave," J. Acoust. Soc. Am. **31**, 667–677.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice Hall, Englewood Cliffs, NJ).

Rosenberg, A. E. (**1971**). "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590.

Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. (**1974**). "Average magnitude difference function pitch extractor," IEEE Trans. Acoust., Speech, Signal Process. **22**, 353–362.

Rothenberg, M. (**1973**). "A new inverse-filtering technique for deriving the glottal airflow during voicing," J. Acoust. Soc. Am. **53**, 1632–1645.

Sjölander, K. (**2004**). "Snack sound toolkit," KTH Stockholm, Sweden, http://www.speech.kth.se/snack/ (last viewed January 2007).

Sluijter, A., and Van Heuven, V. (**1996**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.
Sluijter, A., Van Heuven, V., and Pacilly, J. (**1997**). "Spectral balance as a cue in the perception of linguistic stress," J. Acoust. Soc. Am. **101**, 503–513.
Swerts, M., and Veldhuis, R. (**2001**). "The effect of speech melody on voice quality," Speech Commun. **33**, 297–303.
Titze, I. R. (**2004**). "A theoretical study of f0-f1 interaction with application to resonant speaking and singing voice," J. Voice **18**, 292–298.
Wakita, H. (**1977**). "Normalization of vowels by vocal-tract length and its application to vowel identification," IEEE Trans. Acoust., Speech, Signal Process. **25**, 183–192.

Time course of speech changes in response to unanticipated short-term changes in hearing state

Joseph S. Perkell[a)]
Speech Communication Group, Research Laboratory of Electronics, and Department of Brain and Cognitive Sciences, MIT, Room 36–511, 50 Vassar Street, Cambridge, MA 02139 and Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215

Harlan Lane
Department of Psychology, Northeastern University, Boston, and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Margaret Denny
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Melanie L. Matthies
Department of Communication Disorders, Boston University, and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Mark Tiede
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139 and Haskins Laboratories, New Haven, Connecticut 06510

Majid Zandipour
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Jennell Vick
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98015

Ellen Burton
Johns Hopkins School of Public Health and the Maryland Association of County Health Officers, Baltimore, Maryland 21205

(Received 3 May 2006; revised 19 January 2007; accepted 19 January 2007)

The timing of changes in parameters of speech production was investigated in six cochlear implant users by switching their implant microphones off and on a number of times in a single experimental session. The subjects repeated four short, two-word utterances, /$dV_1n\sharp SV_2d$/ (S=/*s*/ or /ʃ/), in quasi-random order. The changes between hearing and nonhearing states were introduced by a voice-activated switch at V_1 onset. "Postural" measures were made of vowel sound pressure level (SPL), duration, *F*0; contrast measures were made of vowel separation (distance between pair members in the formant plane) and sibilant separation (difference in spectral means). Changes in parameter values were averaged over multiple utterances, lined up with respect to the switch. No matter whether prosthetic hearing was blocked or restored, contrast measures for vowels and sibilants did not change systematically. Some changes in duration, SPL and *F*0 were observed during the vowel within which hearing state was changed, V_1, as well as during V_2 and subsequent utterance repetitions. Thus, sound segment contrasts appear to be controlled differently from the postural parameters of speaking rate and average SPL and *F*0. These findings are interpreted in terms of the function of hypothesized feedback and feedforward mechanisms for speech motor control. [DOI: 10.1121/1.2642349]

PACS number(s): 43.70.Mn, 43.70.Dn, 43.70.Bk, 43.66.Ts [BHS] Pages: 2296–2311

I. INTRODUCTION

Multiple parameters of speech production change when auditory feedback is lost or restored (Kishon-Rabin *et al.*, 1999; Lane and Webster, 1991; Waldstein, 1990; Cowie and Douglas-Cowie, 1983) or modified in some way (discussed below). Both the nature and the timing of these changes can shed light on the role of hearing in maintaining adult speech production. There is a considerable literature concerning the nature of changes in speech brought about by changes in auditory feedback, but less is known about the precise time course of such changes.

A. Suprasegmental vs segmental changes

Speakers use auditory information to monitor listening conditions and adjust suprasegmental aspects of their speech accordingly. When listening conditions are degraded by the imposition of loud masking noise, speakers with normal hearing increase vocal amplitude (the Lombard effect-Lane and Tranel, 1971), fundamental frequency (Bond *et al.*, 1989; Clark *et al.*, 1987) and the duration of speech segments (Tartter *et al.*, 1993; van Summers *et al.*, 1988; Hanley and Steer, 1949). Similar changes have been observed in postlingually deaf cochlear implant users in whom auditory feedback has not been masked but blocked: temporarily by turning off their implant speech processors. This results in louder, slower speech (Svirsky *et al.*, 1992). Such changes tend to make speech more intelligible (van Summers *et al.*, 1988; Dreher and O'Neill, 1958; Peters, 1955; Draegert, 1951).

Another strategy that speakers might use to compensate for perceived degradations in speaking conditions would be to increase contrasts at the segmental level. Instead, decreases in vowel contrast have been observed when implant users' auditory feedback is blocked (cf. Perkell *et al.*, 2001) or the auditory feedback of normal-hearing speakers is masked (cf. Bond *et al.*, 1989; van Summers *et al.*, 1988). It appears contradictory that speakers would enhance some aspects of speech intelligibility while allowing others to degrade.

In search of the reason that speaking sound level and durations increase when auditory feedback is interrupted whereas vowel contrast decreases, Perkell *et al.*, (2007) exposed both speakers with normal hearing and those with cochlear implants to masking noise that ranged in intensity from just detectable to maximally tolerable. Measures included vowel duration, sound pressure level (SPL), and average vowel spacing (the mean separation of all possible vowel pairs in the formant plane; Lane *et al.*, 2001). Similar trends were observed for both subject groups, although the results from the implant users were more variable. Vowel duration and SPL increased with noise level as expected. Average vowel spacing also increased with noise intensity, but only for low to moderate noise levels. At the higher levels, vowel spacing declined. Perkell *et al.*'s interpretation was that subjects tended to hyperarticulate their vowels for as long as they could hear the resulting contrast enhancement. As noise levels increased and they could no longer perceive vowel contrasts, an influence of economy of effort (Lindblom, 1990) prevailed, which resulted in the decreased vowel spacing at the highest noise levels.

This pattern of results suggests that listeners with normal hearing and postlingually deafened cochlear implant users respond similarly to degradation and to loss of auditory feedback and seek to increase their intelligibility with both suprasegmental and segmental cues, provided that they can hear themselves speak well enough. When they cannot, segmental contrasts decrease. This is true when (1) normal-hearing speakers speak in loud masking noise; (2) cochlear implant users speak in loud masking noise; and (3) cochlear implant users speak without benefit of prosthetic hearing. The difference between segmental contrasts on the one hand and durations and SPL on the other indicates that the two types of parameters might be controlled by somewhat separate mechanisms in which auditory feedback plays different roles.

The characteristics of these two types of parameters merits a brief discussion here. Based on a study of vowel production in cochlear implant users, Perkell *et al.* (1992) argued for the existence of a distinction between postural and segmental parameters. They suggested that speaking rate (which is reflected inversely in vowel durations), overall SPL, and average $F0$ are acoustic manifestations of postural settings that are adjusted rapidly to maintain intelligibility in the face of changing acoustic transmission conditions, for both normal-hearing speakers and users of cochlear implants. A possible alternative term for this class of speech parameters, "suprasegmental," usually refers to linguistically salient manifestations of prosody. Use of the term "postural" is intended to focus attention on nonlinguistic, relatively long-term average aspects of the speech signal. (The term postural was first introduced in this context by Stevens, Nickerson and Rollins, 1983). As to the segmental parameters, for each phonemic contrast in our research, a distance is calculated between the members of a contrast pair—which we call *contrast distance*. For example, the contrast distance of /ɑ/-/ʌ/ is the mean separation (in Hz) in the $F1 \times F2$ plane of tokens of those two phonemes; while the contrast distance of /s/-/ʃ/ is the difference in Hz between the average value of tokens of each phoneme in spectral median (Matthies *et al.*, 1994; 1996).

B. Timing of postural changes

All of the effects on speech parameters discussed above have been observed in the course of single experimental sessions but those studies were not designed to determine how long it takes for a feedback-based compensatory response to emerge from normal variability in speech production. Svirsky *et al.* (1992) examined how quickly speech production parameters can change as a result of loss and restoration of auditory feedback. Three postlingually deafened cochlear implant users turned their speech processors off for 24 h prior to testing. On arriving in the laboratory, they read

[a]Author to whom correspondence should be addressed. Electronic mail: perkell@speech.mit.edu

words in carrier phrases, first with their processors off, then with their processors turned on. Their processors were then turned off again, and a final repetition of the reading task was performed. The investigators measured vowel duration, SPL, fundamental frequency ($F0$), and first and second formant frequencies.

In all cases, restoring hearing after 24 h of deprivation resulted in a significant decrease in SPL and $F0$. The effects of turning the processor off again were inconsistent across subjects. Thus, compensatory adjustments in the postural parameters of sound level and $F0$ (cf. Perkell *et al.*, 1992) were more consistent when the speakers experienced a change from a nonhearing state to a hearing state than for a hearing change in the opposite direction. On the other hand, changes in durations (also a postural parameter) and segmental vowel formant frequencies were highly variable across subjects in both direction and time course. Svirsky *et al.* (1992) describe the time course of SPL changes as follows: "We can say with some certainty that SPL had reached a lower level by the first three to four occurrences of each vowel after turning the speech processor on, although inherent variability makes it difficult to indicate precisely when this change took place—it may have been well under way by the first or second occurrence of each token after turning the processor on" (p. 1290).

An important question raised by these observations is how long it takes for a compensatory adjustment to develop in response to a change in auditory feedback. In studies employing pitch-shifted feedback during the production of prolonged vowels, compensatory responses typically occur at a latency of 100–150 ms (Burnett *et al.*, 1997, 1998; Kawahara and Williams, 1996). Natke and Kalveram (2001) reported the effects of pitch shifts for randomly selected trials when normal-hearing subjects produced the nonsense word /tatatas/ with different stress patterns. The subjects' voice $F0$ changed to compensate for the pitch shift in the first syllable, but only if that syllable was stressed. If the first syllable was unstressed, compensatory changes in $F0$ did not take place until the second syllable. The mean duration for stressed syllables was 325 ms and for unstressed syllables, it was 125 ms. Thus this result, indicating a delay of around 125 ms for feedback-based compensatory adjustments, is consistent with those reported by Burnett *et al.* (1997, 1998) and Kawahara and Williams (1996). Xu *et al.* (2004) also perturbed $F0$, but in speakers of Mandarin, in which tone contours are used to differentiate CV words from one another. They found within-syllable compensatory responses with latencies as short as 100 ms. When contrasted with results showing longer delays, these findings were interpreted as indicating that the system for "regulation of voice $F0$ may be task dependent" (Xu *et al.*, 2004, p. 1168).

C. Timing of segmental changes

Experiments in which speakers experienced unexpected changes in vowel formants in their auditory feedback throw further light on the timing of segmental vs postural parameters. Tourville *et al.* (2005) introduced abrupt unanticipated shifts of $F1$ (with an 18 ms delay) in the feedback of speakers' vowels embedded in /CεC/ words. Subjects responded rapidly (within 100–200 ms) with partial compensatory changes in $F1$ (i.e., $F1$ changes in the opposite direction). Such results demonstrate that the speaker is able under certain conditions to generate corrective motor commands during the current articulatory movement and that these changes can be observed experimentally if the movement lasts long enough. Purcell and Munhall (2006a) unexpectedly altered $F1$ in the auditory feedback of steady-state vowels and also found partial compensatory responses; however with a longer latency (up to 460 ms).

Most of the experiments cited above were designed mainly to look for closed-loop compensatory responses to modifications of auditory feedback—responses that are manifested during the production of the sound in which the modification is introduced. On the other hand, several recent "sensorimotor adaptation" experiments have been reported in which the focus was on compensation revealed in the production of subsequent sounds, and on "adaptation"—persistence of compensatory adjustments when auditory feedback is masked or when the perturbation is no longer present. In these studies, the experimental apparatus introduces incremental modifications of acoustic parameters in nearly real time and the subjects are unaware of the perturbations. Jones and Munhall (2002) introduced pitch shifts to speakers of Mandarin and found compensatory responses that persisted when the feedback was returned to normal. Houde and Jordan (1998, 2002) shifted $F1$ and $F2$ of vowels in whispered CVC (consonant-vowel-consonant) words to effectively change vowel quality that speakers heard themselves uttering. The speakers partially compensated for these changes in repeated elicitations, and the compensations persisted in the presence of masking noise. Using a similar technique, Villacorta *et al.* (2004, 2005) shifted vowel $F1$ in the feedback of speakers' voiced VCV (vowel-consonant-vowel) utterances (also see Purcell and Munhall, 2006b). Partial compensatory adjustments were found; they persisted when feedback was masked and also for a short time after the perturbation was removed. The results of such studies may be interpreted as follows: speakers generate feedback-based error corrections, and if the movement lasts long, closed-loop corrections may be observed during the movement. Regardless of movement duration, the error corrections are incorporated into feedforward commands for the production of subsequent sounds (cf. Guenther *et al.*, 2006).

D. Goal and hypotheses of this study

To our knowledge, there have been no studies designed to simultaneously examine segmental and postural responses to unexpected modifications of acoustic feedback in a way that could separate those responses from one another. The goal of the current study was to address this issue by examining how rapidly both kinds of parameters of speech production change when a speaker's hearing is switched between blocked and unblocked states. The study uses repeated elicitations and multiple switches in hearing state, so that hearing-related effects can emerge from normal background variability in speech parameters. In order to examine the la-

TABLE I. Subject Characteristics. Male speakers are indicated by M, female, F. All but one of the subjects (FK) have participated in other studies in this laboratory. Subject identifiers are consistent across publications to facilitate comparisons across studies. (L designated left ear, R right.)

Speaker code	FI	FJ	FK	MM	MO	MP
Vowel perception	0.29	0.26	0.14	0.32	0.51	0.27
Consonant perception	0.26	0.25	0.31	0.32	0.49	0.21
Etiology	Auto-immune response	Infection	Infection	Noise (WWII)	Blood clot	Hereditary
Age at onset of change in hearing	19	5	48	20	60	Birth
Age at onset of profound loss	54	45	49	72	67	26
Age at cochlear implantation	56	46	50	78	72	36
Hearing aid used pre-CI: L, R, both	None	None	Both	Both	Left	Both
Implant: clarion/ nucleus	Nucleus-24	Clarion	Nucleus-24	Clarion	Clarion	Clarion
Processor strategy	Spectral Peak Coding	Simultaneous Analog Stimulation	Advanced Combination Encoders	Continuous Interleaved Sampling	Continuous Interleaved Sampling	Continuous Interleaved Sampling

tency of compensatory responses with respect to the time of the change in hearing state, test utterances consist of sequences of two single-syllable CVC words and parameters in both words are examined. The study employs a group of postlingually deafened speakers who had had cochlear implants for one year. With implant users, it was possible to unexpectedly and completely block and restore prosthetic hearing by switching the input to their speech processors off and on at unpredictable intervals.

Based on the preceding background, the experiment is designed to test the following hypotheses:

1. Segmental contrasts will decrease when hearing is blocked, and they will increase when hearing is restored. Conversely, the postural variables (SPL, *F*0 and sound-segment duration) will increase when hearing is blocked, and they will decrease when hearing is restored.
2. If the switch in hearing state is introduced unexpectedly at the beginning of the vowel in the first of two CVC words spoken in sequence, changes in contrast distance and the postural variables will be evident in the second word, unless the duration of the vowel in the first word exceeds 150 ms. In that case, changes will be observed during that first vowel.
3. The latency of parameter changes following the switch will differ, depending on whether the parameter indexes a segmental or a postural variable.

II. METHOD

A. Subjects

Subjects were three male and three female postlingually deaf, adult, paid cochlear implant users. Informed consent procedures were carried out as approved by the Committee on the Use of Humans as Experimental Subjects at MIT. The implant was either the Clarion (Advanced Bionics, Wilson *et al.*, 1995) or the Nucleus 24 device, (Cochlear Corp., Blamey *et al.*, 1987; McKay and McDermott, 1993). The subjects were referred by the Massachusetts Eye and Ear Infirmary or the University of Massachusetts Memorial Medical Center. Other studies from our laboratory (Lane *et al.*, in press; Lane *et al.*, 2005; Perkell *et al.*, 2007) include data from five of these individuals; to facilitate comparison of results, the same subject codes are used across all the studies. Pertinent characteristics of the subjects are summarized in Table I. To provide an approximate indication of subjects' ability to perceive speech, consonant and vowel identification scores are reported in Table I. All subjects participated in a vowel and consonant recognition test. Subjects were tested in a forced-choice task with eight vowels and another with 11 consonants. There were two recorded talkers, a male and a female, both speakers of Standard American English. With implant users' perception of their auditory feedback in mind, female subjects were presented recordings from the female talker and male subjects from the male talker. A practice test was given in which the stimuli were presented once each orthographically on a monitor and audibly with loudspeakers so the subject heard the stimuli knowing what the corpus would be. There were eight practice trials at the start of each session in which each of the syllables was presented once. During the practice, the listener was encouraged to adjust the volume control on the speakers to a comfortable loudness level. Then, subjects listened to four repetitions of each of three productions of each of the eight vowels for a total of 96 trials per block in random order; two blocks were presented in each testing session. (For more detail see Lane *et al.*, in press).

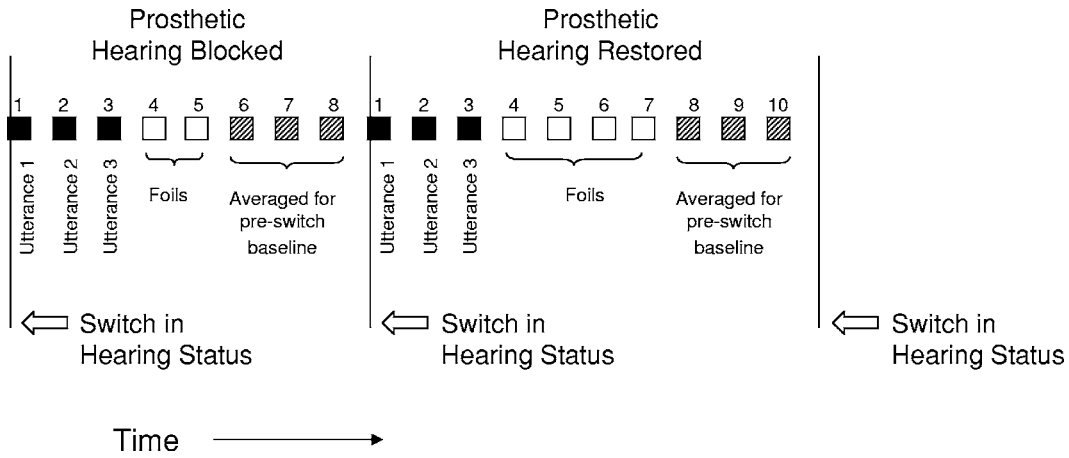


FIG. 1. Schematic of the experimental procedure. The squares represent individual two-word utterances (e.g., "Dun shed"), presented as a function of time. The vertical lines represent the switching of hearing state at 20 ms postvoicing onset in some utterances. Unfilled squares represent foils that were included to make the timing of the switches unpredictable, but were not analyzed; the number of foils varied from one to seven. Filled squares represent the utterances that were entered into the analysis. Results from the three utterances immediately prior to each switch (hatched squares) were averaged to obtain a pre-switch base line. For the three trials following a switch (solid black squares numbered from one to three) results were analyzed separately for each of the three postswitch utterances.

B. Elicitation

The speech elicitation set consisted of four /dV1n#SV2d/ utterances of two words each: *Don shad, Don sad, Dun shed*, and *Dun said*. Thus, there were two vowel contrasts, /ɑ/-/ʌ/ in the first word position and /æ/-/ɛ/ in the second, and the sibilant contrast /s/-/ʃ/ in the second. The subject was instructed to read the stimuli with neutral stress, at a comfortable loudness level. There was a brief practice period, in which the rate of stimulus presentation was adjusted to be comfortable for the subject, typically 3–3.5 s per stimulus item. The stimuli were presented in quasi-random order. In order to prevent the subject from anticipating the switches in hearing state, the number of stimulus presentations between switches was varied, with a minimum of eight items between switches.

Figure 1 illustrates the variable of utterance position; the term refers to the position in time of an utterance relative to a preceding or following switch in hearing status. Tokens with the same utterance position (and the same phonetic content) were averaged in order to determine, for example, the average duration of a vowel produced as the first utterance following a switch in hearing state from prosthetic hearing blocked to prosthetic hearing restored. Each square represents an utterance (e.g., *Dun shed*). Vertical lines represent switches in hearing status. The three solid black squares following each switch in hearing status are labeled utterance 1, utterance 2, and utterance 3. Blank squares represent foils, utterances that were not analyzed but were elicited to insure that subjects could not anticipate the timing of the switches in hearing status. The three utterances prior to a switch in hearing status (shaded squares) were averaged for a pre-switch base line.

A minimum of 15 repetitions of utterances containing each of the four vowels /ɑ/, /ʌ/, /ɛ/ and /æ/ was recorded for each direction of switch (hearing blocked or restored) and for each of the three utterances immediately preceding and immediately following the switch. For example, the vowel /ɑ/ in the first word in *Don sad* or *Don shad* was pronounced 15 times immediately following a switch that blocked hearing, 15 times as the second utterance following a switch that blocked hearing, and 15 times as the third utterance following a switch that blocked hearing. The need to provide multiple foils and to randomize vowels as well as sibilants resulted in a very large initial stimulus set. To avoid fatiguing the subjects, for the sibilants /s/ and /ʃ/, a minimum of 12 repetitions was recorded for each direction of switch and each of the three utterances immediately preceding and immediately following the switch in hearing state.

C. Equipment

Each subject's speech processor program currently in use was uploaded from his or her own processor, then downloaded to a laboratory-owned processor from the same manufacturer. Subjects then adjusted the processor controls until their own speech sounded "normal."

The subject was seated in a sound-attenuating room in a comfortable office chair. A head-mounted electret microphone (Audio-Technica, model AT803B) was placed at a fixed distance of 20 cm from the subject's lips and was connected through a preamplifier for recording the speech signal. A second microphone was placed near the subject's ear. Its output was connected to a custom-built feedback controller (Technical Collaborative, Lexington, MA). The output of the controller was connected to the input of the speech processor, which in turn delivered the stimulation signal to the subject's implant. The feedback controller included a voice-activated switch that turned the input to the speech processor on or off with a delay of 20 ms from the onset of a vowel. The feedback controller's switching function was ramped to avoid the generation of abrupt changes in amplitude that would be heard as clicks. The stimuli were presented in text form on a computer monitor. Stimulus presentation and arming of the voice-activated switch were under computer control.

For calibration of sound pressure level, an electrolarynx (Cooper-Rand Sound Source; Luminaud, Inc.; Mentor, OH)

was placed in front of the speaker's lips while an experimenter observed the sound pressure level on a sound level meter (C scale) placed next to the microphone. The calibration signal and the subjects' speech were low-pass filtered at 7.2 kHz and digitized in real time with a 16 kHz sampling rate.

D. Data extraction

Extracted data consisted of the postural variables of vowel *F*0, SPL, duration, and the segmental variables of vowel *F*1 and *F*2 and sibilant spectral mean. Working with a display of the digitized speech signal of each utterance, an experimenter placed markers at the following points in each /*d*V$_1$*n*♯SV$_2$*d*/ utterance: (1) at the onset of V$_1$ (/ɑ/ or /ʌ/); (2) at the offset of V$_1$; (3) at the onset of S (the sibilant /s/ or /ʃ/); (4) at the offset of S; (5) at the onset of V$_2$ (/æ/ or /ɛ/); and (6) at the offset of V$_2$. Vowel data were extracted from the 15 repetitions at each of the three utterances immediately preceding each switch in hearing state, and of each of the three utterances immediately following.

For the vowels, *F*1, *F*2, and *F*3 were extracted algorithmically from an LPC (linear-predictive coding) spectrum around midvowel using a 25 ms analysis window. The LPC filter order was chosen to optimize formant extraction for each speaker. The algorithm displayed, for each vowel token, the initial measurements at the exact midpoint of the vowel; a broadband spectrogram on which was superimposed the formant trajectories that were detected; and, finally, the spectral cross section at the measurement time. If the first three formants were detected unambiguously in the regions expected for that vowel target, the experimenter accepted that token with those values. If not, the experimenter adjusted the measurement offset time slightly (as much as three glottal cycles in either direction) until the formants could be detected unambiguously. *F*0 was also estimated at the same offset over a centered 40 ms window, based on the filtered error signal autocorrelation sequence to minimize formant interaction (modified autocorrelation analysis; cf. Markel and Gray, 1976). Duration and rms amplitude were extracted algorithmically at the same time. The previously recorded calibration signal was used to convert the rms amplitude to dB SPL. Contrast distances between the vowels in the two first words (/ɑ/, /ʌ/) and between the vowels in the two second words (/æ/, /ɛ/) were calculated as Euclidean distances in the formant 1 × formant 2 plane expressed in mels.

Sibilant data were extracted from the 12 repetitions at each of the three utterances immediately preceding each switch, and from each of the three utterances immediately following. For each sibilant, the spectral mean was extracted algorithmically. Contrast distance between the sibilants was calculated as the average separation of the spectral means of the tokens of /s/ and /ʃ/ (Jongman, Wayland, and Wong, 2000; Forrest *et al.*, 1988; Matthies *et al.* 1994, 1996).

E. Graphing and statistical analyses

Figure 2 illustrates how the data were averaged over both utterance content and utterance position. As in Fig. 1, squares represent individual utterances and vertical lines represent switches in hearing status. The postural parameters of the second utterance following restoration of prosthetic hearing (number 2—black squares) have been selected for illustration. Each parameter mean is derived from data collected over the entire experimental session for a given utterance position and word, while the surrounding content varies unpredictably.

Repeated-measures analyses of covariance (ANOVAs) for each of the dependent variables were performed with subjects as the categorical variable. The effects of blocking and of restoring hearing were tested in separate ANOVAs. In order to test the hypotheses concerning parameter changes following switches in hearing status, six planned contrasts were evaluated for each dependent variable and each change in hearing state—blocked and restored. For each of the two vowel contrasts /ɑ/-/ʌ/ and /æ/-/ɛ/, the mean contrast distance of the first, second and third utterances following the switch in hearing state were contrasted with the mean base line value, viz., the value averaged over the three utterances immediately preceding the switch. For the single sibilant contrast (/s/-/ʃ/), the effects of the switch on mean contrast distance in each of two vowel contexts (/æ/ and /ɛ/) were evaluated. For the three postural variables—duration, SPL and *F*0—planned contrasts assessed the change from the three base line utterances pooled to the first, second, and third postswitch utterances; this was done for each of the two vowel positions in an utterance (first word or second word). Data for the two different vowels in each of those two positions were pooled (first word: /ɑ/-/ʌ/ and second word: /æ/-/ɛ/). The $p < 0.01$ level of significance was adopted in testing these contrasts to reduce the chances of Type I errors.

III. RESULTS

A. Contrast distance

1. Vowels

Figure 3 illustrates the results of the analysis of vowel contrast distances, averaged across subjects. The vertical dashed lines represent the time of the switches, which actually occurred at the beginning of the vowel in the first of the two words in an utterance. The symbols to the left of these lines represent the parameter values averaged over the last three utterances prior to the switch. Vertical pairs of dotted lines represent the variable intervals in which stimulus items were presented, but not analyzed (called "foils" in the figure legend). For each panel, the question of interest is whether the results for each of the utterances labeled 1, 2, and 3 on the abscissa are different from the values measured in the pre-switch base line. Table II reports mean values and the outcome of statistical tests. Each row corresponds to a change in prosthetic hearing, an utterance position immediately following that change, and the vowel contrast embodied in the utterance. For each row, the table reports the average value of contrast distance in the three utterances preceding the switch, the value for the postswitch utterance, the *F* value of a planned comparison, and a measure of variance accounted for (eta-square). The changes in contrast distance following changes in hearing state were inconsistent and mostly quite small, median 1.6%. Six of the 12 contrasts

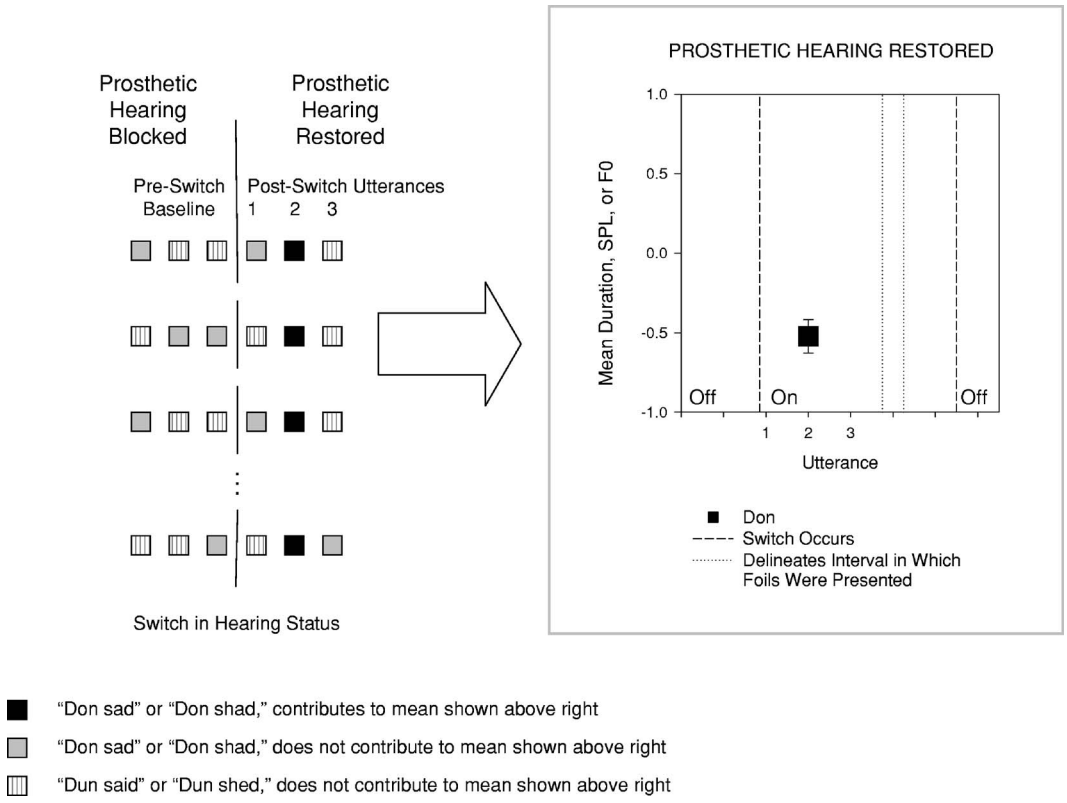


FIG. 2. Schematic illustrating how graphs of results were generated. The example shown here is for the /ɑ/ in *Don* occurring as the second utterance after a switch in hearing status from prosthetic hearing blocked to restored. As in Fig. 1, each square represents an individual two-word utterance (e.g., *Don sad*). Solid filled boxes represent utterances of *Don* (followed by either *sad* or *shad*). Only the black (not the gray) utterances contribute to the data point shown in the graph at the right side of the figure. Hatched squares represent utterances of *Dun* followed by *said* or *shed*. In order to generate the single data point shown in the graph at the right, values were averaged across all 15 utterances of *Don* occurring during the second utterance after the switch in hearing status from hearing blocked to hearing restored. The vertical ellipsis indicates that not all 15 switches in hearing state are pictured here; the utterances shown would have been scattered in time throughout the experimental session.

were not statistically reliable at $p < 0.01$. Most changes were increases both when hearing was blocked and when it was restored (cf. Fig. 3). In an exception worth noting, after hearing was blocked, the /æ/-/ɛ/ vowel contrast declined 8 mels in the first utterance after the switch (from 90 to 82 mels, row 4), and when hearing was restored that contrast increased 5 mels (row 10). That change in contrast distance after blocking was not sustained in the second utterance (row 5) but did appear in the third. The 5 mel increase after restoring feedback was sustained in the second utterance (row 11) but not the third (row 12).

2. Sibilants

Sibilant results from subjects FI and FJ were excluded from analysis because some data were lost due to technical difficulties. The results from the remaining four subjects are illustrated in Fig. 4.

Mean sibilant contrast distances and the outcome of statistical tests appear in Table III. Again, dashed lines represent the occurrence of a switch and dotted lines the interval in which utterances were elicited but not analyzed (foils). Upper panels report results from hearing blocked, lower from hearing restored. Left-hand panels are for the vowel context /æ/, right-hand panels for /ɛ/.

Table III shows that changes in hearing state had no statistically reliable effects on sibilant contrast distance. This was true no matter whether postswitch utterances were considered singly or contrasted as a group with base line utterances. There is a drop worth noting in contrast distance on the first utterance in the /æ/ environment when prosthetic hearing was blocked (row 1 and upper left-hand panel). However, because of the large amount of variability, the drop is not reliable on this or the following two utterances (rows 2 and 3) which are much closer to the preswitch base line. This drop in contrast distance is not replicated in the /ɛ/ environment (there are increases on the second and third utterances, rows 5 and 6), nor is there evidence of an effect in either vowel context when prosthetic hearing was restored.

B. Postural parameters

1. Duration

Figure 5 and Table IV summarize the effects of short-term changes in hearing state on vowel duration. Changes in hearing state were associated with large reliable changes in vowel duration. The changes in vowel duration following blocking or restoring of prosthetic hearing were all statistically significant at $p < 0.01$ in each of the three postswitch utterances and both vowel positions (first word and second word of each utterance). When hearing was blocked, the rise in vowel duration from the base utterances to the first utterance postswitch was 12% for the mean of the two vowels

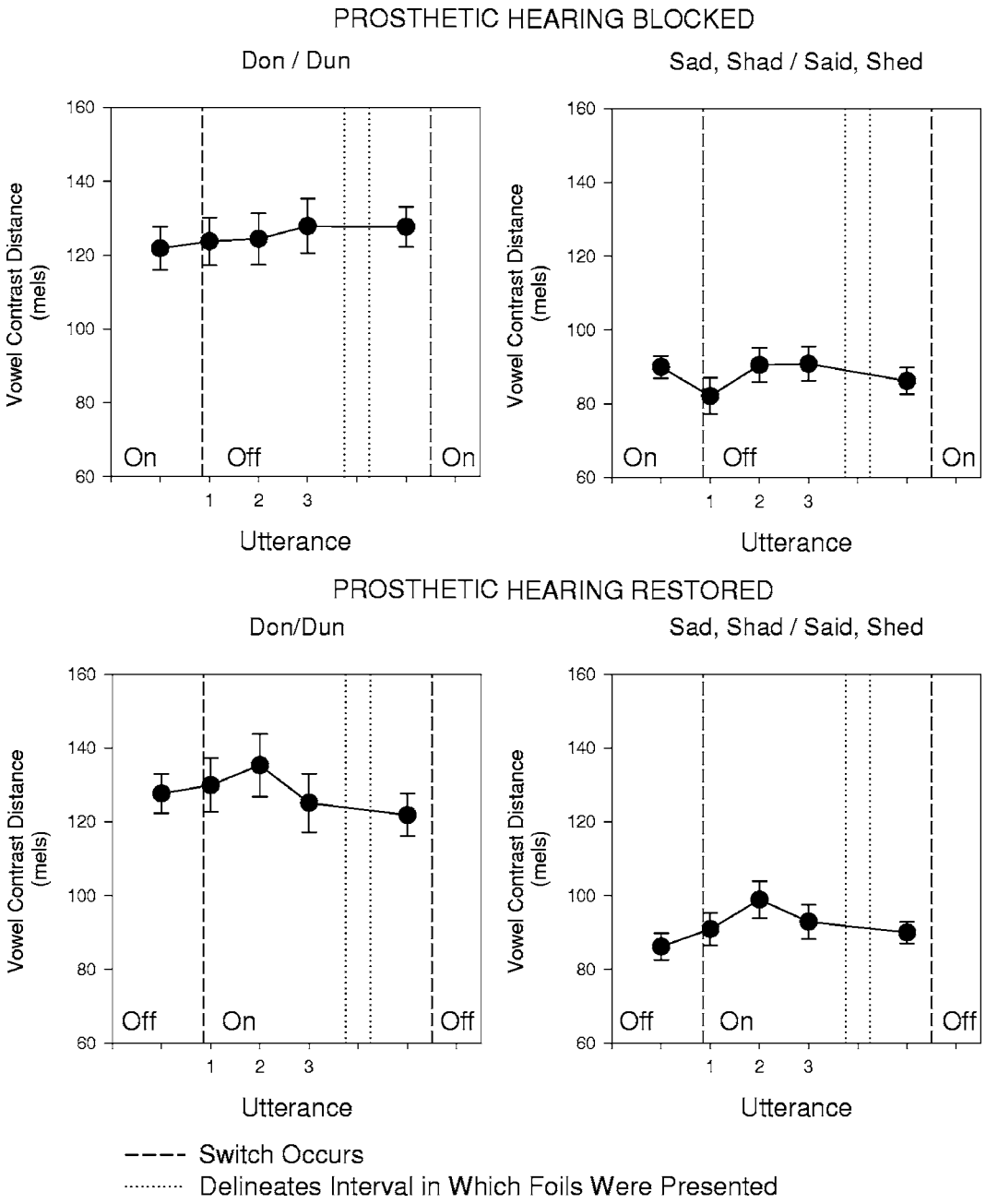


FIG. 3. The effects on vowel contrast distance of blocking (top) and restoring (bottom) prosthetic hearing, as a function of utterance relative to a switch in hearing state, for vowels in the first word (left) and vowels in the second word (right). The vertical dashed lines represent the relative timing of the switches (which actually occurred at the beginning of the vowel in the first of the two words in an utterance; the line has been shifted for clarity). The symbols immediately to the left of these lines represent the average over three utterances immediately prior to the switch. Vertical pairs of dotted lines represent the variable intervals in which foils were presented.

occurring in word 1 (row 1, either /ɑ/ or /ʌ/) and 17% for the two vowels occurring in word 2 (row 4, /æ/ or /ɛ/). In both cases, the increase over base line was sustained through all three utterances. When hearing was restored, the drop in vowel duration from the base utterances to the first utterance postswitch was 15% for the two vowels occurring in the second word (row 11). For those in the first word, there was no change detected in utterance 1 (row 7) but a drop of approximately 7% was found on the following two utterances. (Because the planned contrasts measure within-subject changes in speech parameters, the corresponding F values can be significant while means computed by averaging across subjects can show little or no change.)

Changes in hearing state can affect phonemic contrasts and postural variables during an ongoing segment only if the duration of that segment is long enough. Table IV shows that syllable duration means based on 30 determinations (15 trials × 2 vowels) were always well in excess of the 100–150 ms duration cited earlier in studies of $F0$ compensation responses. In individual vowel utterances prior to blocking hearing (6 Ss × 15 trials × 3 utterances) not a single one had duration less than 150 ms.

2. SPL

Figure 6 and Table V summarize the results of the analysis of vowel SPL. Changes in hearing state were associated

TABLE II. Vowel contrast distances when prosthetic hearing is blocked and restored for the contrasts/ɑ-ʌ/ and /æ-ɛ/ (df=6,84) All contrasts are significant at $p<0.01$ except where *ns* (not significant) is noted.

Prosthetic hearing state	Utterance contrast	Vowel contrast	Base mean (mels)	Utterance mean (mels)	Contrast *F*	Eta^2 ×100	Row
Blocked	1	ɑ-ʌ	122	124	0.7 ns	5	1
	2	ɑ-ʌ		124	5.8	29	2
	3	ɑ-ʌ		128	1.7 ns	11	3
	1	æ-ɛ	90	82	7.5	35	4
	2	æ-ɛ		91	2.8 ns	16	5
	3	æ-ɛ		91	4.2	23	6
Restored	1	ɑ-ʌ	128	130	1.0 ns	7	7
	2	ɑ-ʌ		135	3.3	19	8
	3	ɑ-ʌ		125	4.4	24	9
	1	æ-ɛ	86	91	2.0 ns	12	10
	2	æ-ɛ		99	7.5	35	11
	3	æ-ɛ		93	2.7 ns	16	12

with small reliable changes in vowel SPL. All average changes were less than 1 dB. The changes in vowel SPL following blocking or restoring prosthetic hearing were all statistically significant at $p<0.01$ in each of the three postswitch utterances and both component words with two exceptions: viz., there was no statistically significant change from base line to the first postswitch utterance for the vowels in word 1, both when hearing was blocked (row 1) and when it was restored (row 7). In both cases, the significant change in the second utterance was sustained in the third. For the vowels in word 2 there *was* a small but significant sustained drop on the first utterance with hearing blocked (rows 4–6). Although one might infer that an additional delay beyond the first word of the first utterance was required for SPL changes to be expressed, the significant drop in SPL on the second word of the second utterance with hearing restored (row 11) was not sustained.

The pattern of results is not consistent with the Lombard effect, which is, however, normally measured over longer time intervals and with masking noise to block hearing. In the present study, for the most part, SPL fell when hearing was blocked and rose when it was restored, whereas in the Lombard effect the opposite occurs—louder speech with hearing blocked by masking noise and softer when only ambient noise is present (Perkell *et al.*, 2007). Some subjects showed significant changes in SPL with changes in hearing state, but others did not. The pattern of the Lombard effect was observed for only one subject, MO. Three speakers produced SPL changes contrary to the Lombard effect for hearing blocked or hearing restored. Two of these three subjects had the greatest magnitudes of change in SPL among the six subjects; thus, their response patterns were clearly reflected in the group results.

3. *F*0

Changes in hearing state were associated with small reliable changes in vowel *F*0 (see Fig. 7). All changes were 6% or less. The *F*0 changes tended to correspond to those in SPL: lower *F*0 when hearing was blocked, higher when it was restored. As with SPL, there was no statistically significant change from base line to the first postswitch utterance for the vowels in word 1, both when hearing was blocked (row 1) and when it was restored (row 7). Consistent with the group results for SPL, the direction of change in fundamental frequency was opposite to that expected, that is, *F*0 was lowered when hearing was blocked and raised when hearing was restored. Comparing Tables V and VI, of the ten cases in which there was a change in *F*0 or SPL from base line to the first, second or third utterance, with either word 1 or word 2, nine out of the ten changes matched in their directions of change. The Pearson correlation coefficient between SPL and *F*0 for all vowels in both hearing conditions in all subjects and trials ($n=1440$) was 0.54, $p<0.01$.

IV. DISCUSSION

A. Contrast distance

The goal of this study was to establish the time course over which adjustments in speech output are made in response to an abrupt, unanticipated change in hearing state. The responses we observed can be classified as one of two types: transient or sustained. Changes in contrast distances among vowels and sibilants were of the transient type; they did not persist reliably beyond a single utterance following a switch. Since changes in contrast distance were not reliably sustained, it appears that those changes do not function to offset or compensate for the changes in hearing introduced experimentally. Recall that, on the contrary, several studies cited in the introduction found compensatory adjustments to auditory feedback modifications. The difference may lie in this: systematic changes with short latency are observed in segmental parameters when feedback is modified, but not when it is completely blocked or restored as in the present study.

A hypothetical explanation for the differences observed with the two types of paradigm may be found in the operation of a model of speech motor planning that consists of feedforward and feedback control subsystems (cf. Guenther *et al.*, 2006). In such a model, the articulatory movements of mature speech are normally controlled by feedforward com-

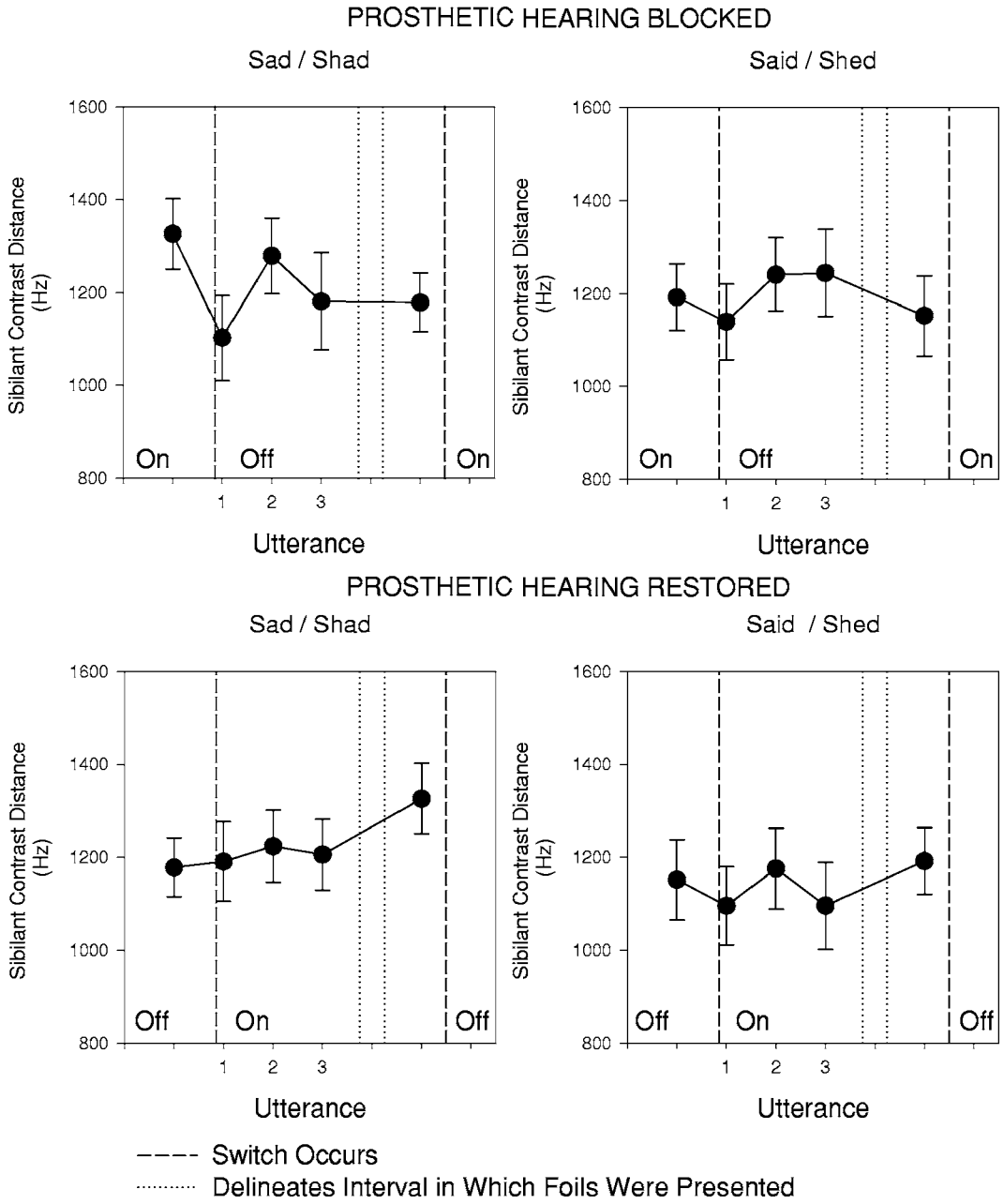


FIG. 4. The effects on sibilant contrast distance of blocking (upper panels) and restoring (lower) prosthetic hearing, as a function of utterance relative to a switch in hearing state. For details, see the caption for Fig. 2.

mands. The feedback control subsystem monitors the speech output for differences between the expected auditory (and somatosensory) consequences of producing the movements and the actual sensory results; if a large enough difference between expected and actual sensations is detected, a feedback-based corrective command is generated and, if the movement lasts long enough, the corrective motor command is expressed during that movement. As described in the introduction, such compensatory corrections have been observed in *F*0 (cf. Burnett *et al.*, 1997, 1998) and vowel formants (cf. Tourville *et al.*, 2005). The corrective motor commands also serve to update the feedforward motor commands for subsequent sounds (and syllables), as has been observed in the cited sensorimotor adaptation experiments (cf. Houde and Jordan, 1998, 2002; Villacorta *et al.*, 2004, 2005; Purcell and Munhall, 2006b).

Based on this view, when auditory feedback is blocked completely, there can be no feedback-based corrective motor commands; consequently, there is little or no change in produced segmental contrasts. There is also no updating of feedforward commands when feedback is blocked. Since feedforward commands are well ingrained and extremely robust, there is virtually no drift and thus the sound output remains "on target" during the time feedback remains blocked. In addition, the lack of drift in the feedforward commands makes it unlikely that any somatosensory feedback errors will be generated when hearing is blocked. When feedback is next restored, there are very few differences between auditory goals and the sound output, so again, very few segmental contrast changes are observed.

TABLE III. Sibilant contrast distances when prosthetic hearing is blocked and restored in the vowel environments /æ/ and /ɛ/. All contrasts are nonsignificant at $p<0.01$ (df=4,24).

Prosthetic hearing state	Utterance contrast	Vowel context	Base mean (Hz)	Utterance mean (Hz)	Contrast F All ns	Eta^2 ×100	Row
Blocked	1	æ	1326	1103	3.4	36	1
	2	æ		1279	0.9	13	2
	3	æ		1181	1.9	24	3
	1	ɛ	1192	1139	0.9	13	4
	2	ɛ		1240	0.4	6	5
	3	ɛ		1244	0.4	6	6
Restored	1	æ	1178	1191	0.5	7	7
	2	æ		1224	2.0	25	8
	3	æ		1205	2.8	32	9
	1	ɛ	1151	1095	2.9	33	10
	2	ɛ		1176	0.5	8	11
	3	ɛ		1096	0.5	8	12

B. Postural parameters

Measures of postural parameters revealed some changes that were sustained over multiple utterances and persisted through the following pre-switch utterances (i.e., extreme right-hand data points in Figs. 5–7).

1. Vowel duration

The postural variable that changed most consistently in response to a switch in hearing state, both within and between subjects, was duration. For all four of the vowels tested (i.e., both vowels in each of the word positions 1 and 2), duration increased when hearing was blocked in the first word (*Don* or *Dun*) for all three postswitch utterances (see Fig. 5). This made duration the only variable that changed systematically in the first word following the on-to-off switch and sustained the direction of change over all three utterances. The mean duration of the vowels in the three utterances prior to the switch that blocked hearing was greater than 150 ms, and as noted above, compensatory responses can be observed as early as about 100 ms following the introduction of an acoustic perturbation (e.g., Burnett *et al.* 1997, 1998). Comparing durations of the pre-switch vowels in this study to the shortest latencies of compensatory responses in other studies indicates that the prolongation of vowels when feedback was blocked in this study could be due to a closed-loop feedback mechanism.

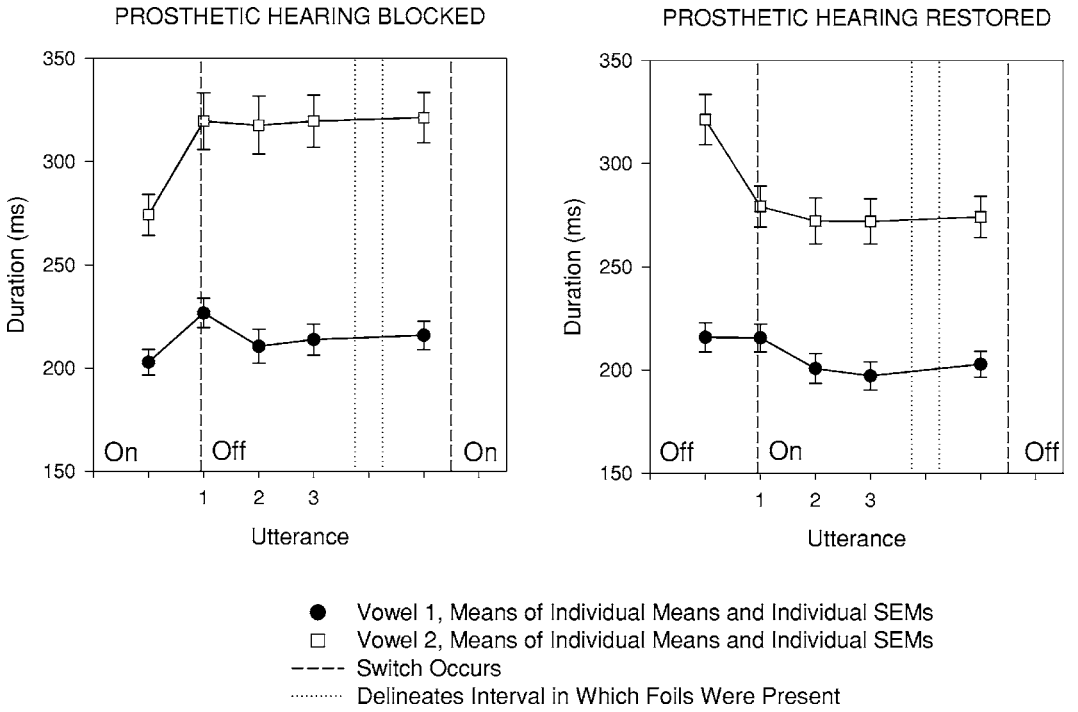


FIG. 5. The effects on vowel duration of blocking (left panel) and restoring (right panel) prosthetic hearing, as a function of utterance relative to a switch in hearing state. Vowels in the first word are represented by round symbols; vowels in the second word are represented by square symbols. Within an utterance, the vowel /ɑ/ was always followed by the vowel /æ/; the symbols for these vowels are both filled; the vowel /ʌ/ was always followed by the vowel /ɛ/; the symbols for these vowels are both unfilled. For further details, see caption for Fig. 2.

TABLE IV. Changes in vowel duration when prosthetic hearing is blocked and restored. All contrasts are significant at $p < 0.01$ (df=6,84).

Prosthetic hearing state	Utterance positions re: switch	Vowel positions re: utterance	Base mean (ms)	Utterance mean (ms)	Contrast *F*	Eta^2 ×100	Row
Blocked	1	1	203	227	30	68	1
	2	1		211	5	25	2
	3	1		214	8	36	3
	1	2	274	320	54	79	4
	2	2		318	58	81	5
	3	2		320	71	83	6
Restored	1	1	216	216	4	22	7
	2	1		201	22	61	8
	3	1		197	8	36	9
	1	2	321	279	78	85	10
	2	2		272	117	89	11
	3	2		272	125	90	12

Prolongation of a sound presumably involves a combination of delaying the onset of the motor commands for producing the next sound and sustaining the agonist commands involved in producing the current one. Regarding truncation, while it might be possible to initiate motor commands for an upcoming sound earlier than planned, it should be more difficult to truncate the effects of motor commands already issued because of delays due to neural conduction and muscle contraction mechanisms. As described in Guenther *et al.* (2006), the time it takes for an action potential in a motor cortical cell to affect the length of a muscle via a subcortical motoneuron consists of "(1) the delay between motor cortex activation and activation of a muscle as measured by EMG [electromyography], and (2) the delay between EMG onset and muscle length change." These two delays add up to about 40 ms (Guenther *et al.*, 2006, p. 284). Presumably, additional tens of milliseconds would be required for there to be enough change in muscle length to produce a perceptible change in the acoustic output. These delays could account for the observation that vowel duration did not decrease until Vowel 2, utterance 1 when hearing was restored (Fig. 5).

2. Vowel SPL and *F*0

Although there was considerable variability in the magnitude and direction of changes, if any, in SPL and *F*0, they were significantly correlated so they are both discussed in this section. A surprising result was the failure to observe an appreciable increase in vowel SPL when hearing was blocked (with one exception, subject MO), and an appre-

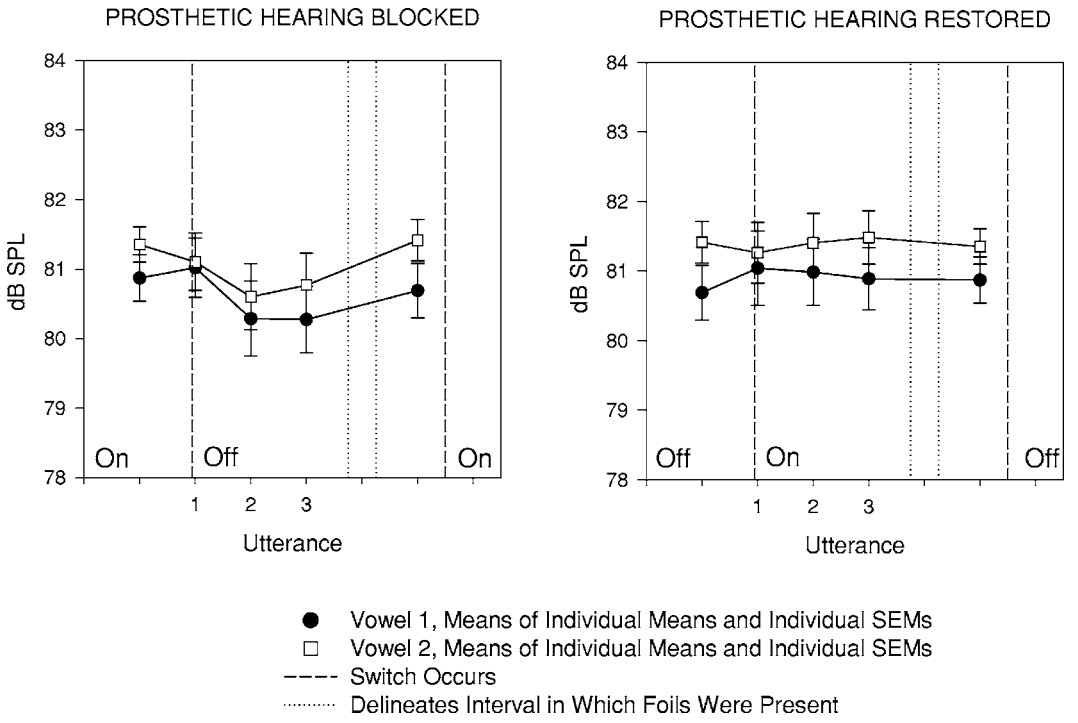


FIG. 6. The effects on vowel SPL of blocking (left panel) and restoring (right panel) prosthetic hearing, as a function of utterance relative to a switch in hearing state. For further details, see captions for Figs. 2 and 5.

TABLE V. Changes in vowel SPL when prosthetic hearing is blocked and restored. All contrasts are significant at $p<0.01$ except where noted; (df=6,84).

Prosthetic hearing state	Utterance positions re: switch	Vowel positions re: utterance	Base mean (dB SPL)	Utterance mean (dB SPL)	Contrast *F*	Eta^2	Row
Blocked	1	1		81.02	0.6 ns	4	1
	2	1	80.87	80.29	5	25	2
	3	1		80.27	10	41	3
	1	2		81.10	34	71	4
	2	2	81.35	80.60	24	64	5
	3	2		80.77	23	62	6
Restored	1	1		81.04	0.9 ns	6	7
	2	1	80.69	81.00	13	48	8
	3	1		80.89	15	52	9
	1	2		81.26	38	73	10
	2	2	81.41	81.40	39	74	11
	3	2		81.48	38	73	12

ciable decrease when it was restored. Two subjects responded by decreasing vowel SPL when hearing was blocked; these decreases occurred on the second utterance following the switch for vowels in both the first and second words. Three other subjects did not change vowel SPL in either direction when their speech processors were switched on or off.

When implant users' prosthetic hearing was altered in other studies from our laboratory, increased SPL was observed when prosthetic hearing was blocked and decreased when it was restored (Perkell *et al.*, 2001; Svirsky *et al.*, 1992). In those studies, however, the intervals in which auditory feedback was removed and restored—on the order of about 20 min—were far longer than the intervals of less than 1 min during the present study. We speculate that the disparity in outcomes may be attributable to the disparity in the durations of the change in hearing state. Such a conclusion would be compatible with the findings that the effects of changing auditory feedback of SPL are larger the more the task simulates real communication (Lane and Tranel, 1971). Then, too, some of our implant users may have learned, during their year's experience with their implant, to moderate their vocal level even when auditory feedback became unavailable.

The absence of auditory information is qualitatively different from the presence of loud masking noise; however, both have the effect of reducing the signal-to-noise (S/N) ratio and thus qualify as adverse speaking conditions. Lane and Tranel (1971) present evidence that speakers communicating in noise maximally increase their SPL about 5 dB for

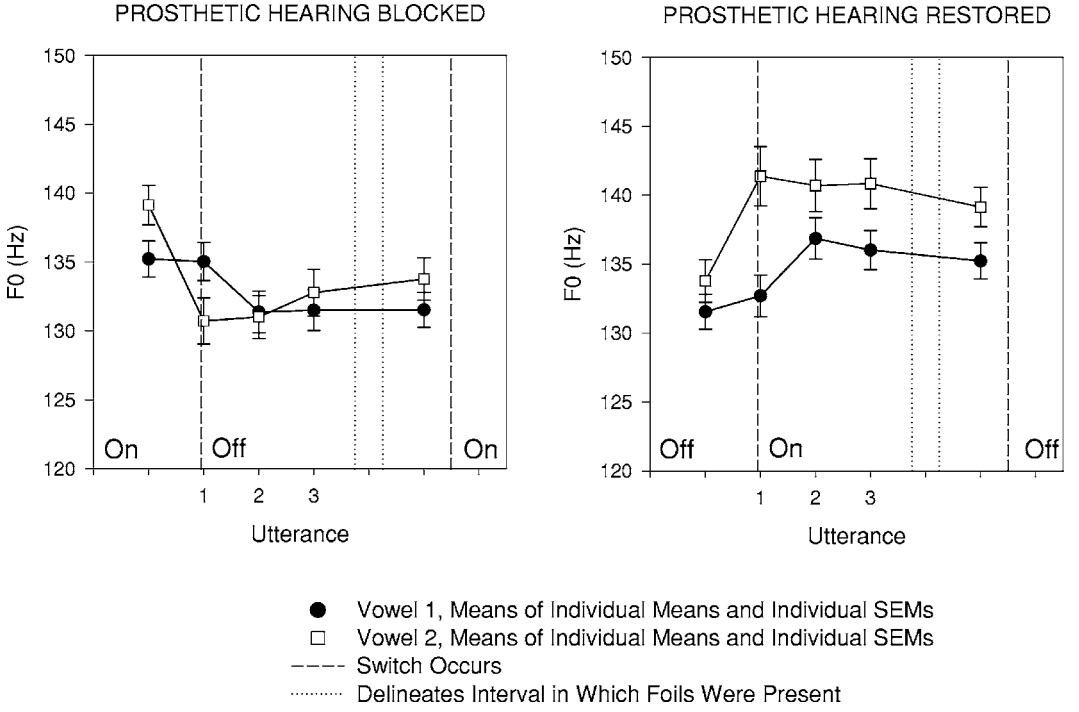


FIG. 7. The effects on vowel *F*0 of blocking (left panel) and restoring (right panel) prosthetic hearing, as a function of utterance relative to a switch in hearing state. For further details, see captions for Figs. 2 and 5.

TABLE VI. Changes in vowel fundamental frequency when prosthetic hearing is blocked and restored. All contrasts are significant at $p < 0.01$ except where nonsignificant (ns) is noted. (df=6,84).

Prosthetic hearing state	Utterance positions re: switch	Vowel positions re: utterance	Base mean (Hz)	Utterance mean (Hz)	Contrast *F*	Eta^2 ×100	Row
Blocked	1	1	135	135	3 ns	15	1
	2	1		132	54	80	2
	3	1		132	155	92	3
	1	2	139	131	72	84	4
	2	2		131	80	85	5
	3	2		133	70	83	6
Restored	1	1	132	133	2 ns	12	7
	2	1		137	149	91	8
	3	1		136	81	85	9
	1	2	134	141	20	59	10
	2	2		141	70	83	11
	3	2		141	108	88	12

every increase in noise of 10 dB and they term this relation the "Lombard function." It has been suggested that the tendency of the deaf to speak more loudly is likewise attributable to compensation for the apparent reduction of the S/N ratio. An experiment by Black (1951) supports this interpretation. Black had 144 college students read lists of words while their SPL was recorded. Then he reduced speakers' self-hearing by exposing them to 110 dB noise for 2 h. After the termination of the noise, he measured subjects' temporary hearing loss at 3 min intervals and again recorded their vocal level several times. Lane and Tranel (1971) report that the function relating the speaker's vocal level in Black's experiment to the amount of that speaker's temporary hearing loss had approximately the same slope (in dB coordinates) but opposite sign as the Lombard function. Lane, Tranel and Sisson (1970) present other evidence that this "sidetone compensation function" is approximately the reciprocal of the Lombard function. Although these studies were conducted with normal-hearing subjects, Perkell *et al.* (2007) found functions relating vocal level to signal-to-noise ratio that were somewhat similar for a group of speakers with normal hearing and one with cochlear implants, when subjected to masking of auditory feedback.

Speakers with normal hearing can learn to modulate the Lombard effect. Those studied by Pick *et al.* (1989) spoke during alternating periods of quiet and noise. When given visual feedback about their vocal level and training in regulating it, they were able to speak quietly in noisy backgrounds, even after the visual feedback was removed. Subjects who returned for retesting a day later were able to restrict the Lombard effect "sharply" without visual feedback. Interestingly, some of these subjects "overcompensated" by dropping their vocal intensity in noise below the levels they used in quiet. Two of the subjects in the current study also used lower vocal intensity when auditory feedback became unavailable.

Just as speakers can voluntarily minimize changes in SPL despite changing listening conditions, so too they can produce voluntary responses to pitch-shifted auditory feedback. Hain *et al.* (2000) instructed subjects to respond to a pitch shift either by raising or lowering their *F*0, or by ignoring it and keeping *F*0 constant. In addition to the expected compensatory change opposite to the direction of the shift, with a latency of 100–150 ms, Hain *et al.* observed a later response, which was almost always made in accordance with experimenters' instructions, hence in the same direction as the shift when so instructed. This provides another example of a response to a change in auditory feedback that can be modified by experimenter's instructions.

V. SUMMARY AND CONCLUSIONS

Returning to the hypotheses of the current study, the first part of Hypothesis 1 predicted that when hearing was blocked, segmental contrasts would decrease and they would increase when hearing was restored. This part of the hypothesis was disconfirmed. With minor exceptions, there were no reliable changes in segmental parameters. We attribute this result to the nature of the perturbation—removal and restoration of feedback, as opposed to modification of feedback parameters. According to Guenther's DIVA model of speech motor control (Guenther *et al.*, 2006), removal and restoration of auditory feedback would not result in any feedback-based error corrections to current movements or to feedforward commands for future movements, so feedforward commands should remain unchanged in the short run, regardless of whether or not feedback is present.

The second part of Hypothesis 1 predicted that blocking auditory feedback would result in increases in SPL, *F*0, and sound segment durations; restoration of hearing would reverse these changes. Such predictions are compatible with earlier results on the behavior of these speech postural parameters in response to changes in acoustic transmission conditions (cf. Lane and Tranel, 1971; Perkell *et al.*, 1992, 2007; Svirsky *et al.* 1992). In support of this prediction, the current results showed increases in sound segment durations when feedback was blocked and decreases when hearing was restored; the duration changes were sustained until the next switch in hearing state. SPL and *F*0 values were correlated, most likely due (at least partly) to the aerodynamic/

biomechanical interdependence of these two parameters (cf. Perkell *et al.*, 1992). In contrast to the findings on duration, changes in SPL were not systematic. They differed among subjects; some were opposite to the direction found in other studies; they were only observed on the first utterance after blocking hearing for one of the two vowel pairs. These anomalous findings may be due to the brevity of the changes in processor state that were employed or to the tendency of some subjects, when they do not have access to prosthetic hearing from their implants, to actively suppress increases in SPL in order to avoid adverse social consequences of speaking too loudly.

Hypothesis 2 predicted that parameter changes would take place in the word following the one in which the switch was made, unless the duration of the vowel in the first word exceeded 150 ms, in which case, responses could occur during that vowel. This hypothesis was supported for duration measurements (but not SPL and *F*0) when hearing was blocked and also when hearing was restored. With a switch in hearing state introduced unexpectedly at the beginning of the vowel in the first of the two CVC words, the duration of the vowel in the first word after the switch exceeded 150 ms, and indeed the first and subsequent vowels were altered in duration as soon as possible after the switch—lengthened during Vowel 1, Utterance 1 with hearing blocked and shortened during Vowel 2, Utterance 1 with hearing restored.

Finally, Hypothesis 3 predicted differences in the latency of parameter changes following a switch in hearing state, depending on whether the parameter indexes a segmental or postural speech variable. As it turned out, changes in contrast distances among vowels and sibilants were of the transient type; they did not persist beyond a single utterance following a switch, whereas duration, a postural variable, changed in the first word following the change in hearing state and sustained that change.

The production of *sound segments*, syllables and words under feedforward control very rarely encounters the kind of distorting perturbations that would cause mismatches between auditory goals and the produced sounds. On the other hand, changes in acoustic transmission conditions, such as in the level of environmental noise, occur frequently and elicit short-latency responses in the form of changes in speaking rate and level. As presented above, the results for segmental and postural parameters did differ from one another. Because unexpectedly blocking and restoring auditory feedback cannot engage the postulated feedback control system that is involved in segmental (and syllabic) sound production, no systematic changes in vowel and sibilant contrasts were observed. The contrary finding of some systematic postural changes that appear to be implemented by feedback control leads to the inference that the control of segmental and postural aspects of speech may involve somewhat different mechanisms, as proposed previously by Perkell *et al.* (1992).

ACKNOWLEDGMENTS

This research was supported by Grant No. R01-DC003007 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We are grateful to Dr. Donald Eddington of the Massachusetts Eye and Ear Infirmary and Dr. Daniel Lee of the University of Massachusetts Medical Center for referring the implant users to the study, and to the implant users for their devotion of considerable amounts of their time. We also thank Advanced Bionics, Inc., and the Nucleus Corporation for their generous donations of research implant processors.

Black, J. (**1951**). "The effect of noise-induced temporary deafness upon vocal intensity," Speech Monographs **18**, 74–77.

Blamey, P. J., Dowell, R. C., Brown, A. M., Clark, G. M., and Seligman, P. M. (**1987**). "Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors," J. Acoust. Soc. Am. **82**, 48–57.

Bond, Z. S., Moore, T. J., and Gable, B. (**1989**). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). "Voice *F*0 responses to pitch-shifted auditory feedback: A preliminary study," J. Voice **11**, 202–211.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice *F*0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Clark, J. E., Lubker, J. F., and Hunnicutt, S. (**1987**). "Some preliminary evidence for phonetic adjustment strategies in communication difficulty," in *Language Topics: Essays in Honor of Michael Halliday*, R. Steele and T. Threadgold, eds., pp. 161–180 (Benjamins, Amsterdam).

Cowie, R. I. D., and Douglas-Cowie, E. (**1983**). "Speech production in profound postilingual deafness," in *Hearing Science and Hearing Disorders*, M. Lutman and M. P. Haggard, eds., pp. 183–230. (Academic, London).

Draegert, G. L. (**1951**). "Relationships between voice variables and speech intelligibility in high level noise," Speech Monographs **18**, 272–278.

Dreher, J. J., and O'Neill, J. J. (**1958**). "Effects of ambient noise on speaker intelligibility of words and phrases," Laryngoscope **68**, 539–548.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Hanley, T. D., and Steer, M. D. (**1949**). "Effect of level of distracting noise upon speaking rate, duration and intensity," J. Speech Hear Disord. **14**, 363–368.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. J., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Kawahara, H., and Williams, J. C. (**1996**). "Effects of auditory feedback on voice pitch trajectories: Characteristic responses to pitch perturbations," in *Vocal Fold Physiology*, P. J. Davis and N. H. Fletcher, eds., pp. 263–278 (Singular, San Diego).

Kishon-Rabin, L., Taitelbaum, R., Tobin, Y., and Hildesheimer, M. (**1999**). "The effect of partially restored hearing on speech production of postlingually deafened adults with multichannel cochlear implants," J. Acoust. Soc. Am. **106**, 2843–2857.

Lane, H., Denny, M., Guenther, F. H., Matthies, M., Ménard, L., Perkell, J., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (**2005**). "Effects of bite blocks and hearing status on vowel production," J. Acoust. Soc. Am. **118**, 1636–1646.

Lane, H., Matthies, M., Denny, M., Guenther, F., Perkell, J., Stockmann, E., Tiede, M., Vick, J., and Zandipour, M. (in press, **2007**). "Effects of short-

and long-term changes in auditory feedback on vowel and sibilant contrasts," J. Speech, Lang. Hear. Res.
Lane, H., Matthies, M., Perkell, J., Vick, J., and Zandipour, M. (**2001**). "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation," J. Speech Lang. Hear. Res. **44**, 552–563.
Lane, H., Tranel, B., and Sisson, C. (**1970**). "Regulation of voice communication by sensory dynamics," J. Acoust. Soc. Am. **47**, 618–624.
Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.
Lane, H., and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**, 859–866.
Lindblom, B. E. F. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, pp. 403–439 (Kluwer, Dordrecht).
Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).
Matthies, M. L., Svirsky, M. A., Lane, H. L., and Perkell, J. S. (**1994**). "A preliminary study of the effects of cochlear implants on the production of sibilants," J. Acoust. Soc. Am. **96**, 1367–1373.
Matthies, M. L., Svirsky, M., Perkell, J., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Hear. Res. **39**, 936–946.
McKay, C. M., and McDermott, H. J. (**1993**). "Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP)," Ear Hear. **14**, 350–367.
Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.
Perkell, J. S., Denny, M., Lane, H., Guenther, F. H., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., and Burton, E. (**2007**). "Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users," J. Acoust. Soc. Am. **121**, 505–514.
Perkell, J., Lane, H., Svirsky, M., and Webster, J. (**1992**). "Speech of cochlear implant patients: A longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2978.
Perkell, J., Numa, W., Vick, J., Lane, H., Balkany, T., and Gould, J. (**2001**). "Language-specific, hearing-related changes in vowel spaces: A preliminary study of English- and Spanish-speaking cochlear implant users," Ear Hear. **22**, 461–470.
Peters, R. W. (**1955**). "The effect of filtering of sidetone on speaker intelligibility," J. Speech Hear Disord. **20**, 371–375.
Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R., and Kearney, J. K. (**1989**). "Inhibiting the Lombard effect," J. Acoust. Soc. Am. **85**, 894–900.
Purcell, D. W., and Munhall, K. G. (**2006a**). "Compensation following real-time manipulation of formants in isolated vowels," J. Acoust. Soc. Am. **119**, 2288–2297.
Purcell, D. W. and Munhall, K. G. (**2006b**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966-977.
Stevens, K. N., Nickerson, R. S., and Rollins, A. M. (**1983**). "Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility," in *Speech of the Hearing-Impaired*, pp. 35–51 (University Park Press, Baltimore).
Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.
Tartter, V. C., Gomes, H., and Litwin, E. (**1993**). "Some acoustic effects of listening to noise on speech production," J. Acoust. Soc. Am. **94**, 2437–2440.
Tourville, J. A., Guenther, F. H., Ghosh, S. S., Reilly, K. J., Bohland, J. W., and Nieto-Castanon, A. (**2005**). "Effects of acoustic and articulatory perturbation on cortical activity during speech production," in *11th Annual Meeting of the Organization for Human Brain Mapping*, Toronto, June 12-16, p. S49.
Van Summers, W., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effect of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928.
Villacorta, V., Perkell, J. S., and Guenther, F. H. (**2004**). "Sensorimotor adaptation to acoustic perturbations in vowel formants," J. Acoust. Soc. Am. **115**, 2430(A).
Villacorta, V., Perkell, J. S., and Guenther, F. H. (**2005**). "Relations between speech sensorimotor adaptation and perceptual acuity," J. Acoust. Soc. Am. **117**, 2618–2619(A).
Waldstein, R. S. (**1990**). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.
Wilson, B., Lawson, D., Zerbi, M., Finley, C., and Wolford, R. (**1995**). "New Processing Strategies in Cochlear Implantation," Am. J. Otol. **16**, 669-681.
Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Consonant and vowel confusions in speech-weighted noise[a)]

Sandeep A. Phatak[b)] and Jont B. Allen
ECE, University of Illinois at Urbana-Champaign, Beckman Institute, 405 N. Mathews Avenue, Urbana, Illinois 61801

(Received 7 April 2006; revised 30 October 2006; accepted 20 January 2007)

This paper presents the results of a closed-set recognition task for 64 consonant-vowel sounds (16 C×4 V, spoken by 18 talkers) in speech-weighted noise (−22, −20, −16, −10, −2 [dB]) and in quiet. The confusion matrices were generated using responses of a homogeneous set of ten listeners and the confusions were analyzed using a graphical method. In speech-weighted noise the consonants separate into three sets: a low-scoring set C1 (/f/, /θ/, /v/, /ð/, /b/, /m/), a high-scoring set C2 (/t/, /s/, /z/, /ʃ/, /ʒ/) and set C3 (/n/, /p/, /g/, /k/, /d/) with intermediate scores. The perceptual consonant groups are C1: { /f/-/θ/, /b/-/v/-/ð/, /θ/-/ð/ }, C2: { /s/-/z/, /ʃ/-/ʒ/ }, and C3: /m/-/n/, while the perceptual vowel groups are /ɑ/-/æ/ and /ɛ/-/ɪ/. The exponential articulation index (AI) model for consonant score works for 12 of the 16 consonants, using a refined expression of the AI. Finally, a comparison with past work shows that white noise masks the consonants more uniformly than speech-weighted noise, and shows that the AI, because it can account for the differences in noise spectra, is a better measure than the wideband signal-to-noise ratio for modeling and comparing the scores with different noise maskers.
[DOI: 10.1121/1.2642397]

PACS number(s): 43.71.An, 43.71.Gv, 43.72.Dv [ADP] Pages: 2312–2326

I. INTRODUCTION

When a perceptually relevant acoustic feature of a speech sound is masked by noise, that sound becomes confused with related speech sounds. Such confusions provide vital information about the human speech code, i.e., the perceptual feature representation of speech sounds in the auditory system. When combined with a spectro-temporal analysis of the specific stimuli, this confusion analysis forms a framework for identifying the underlying *perceptual features* or the *events* (Allen, 2005a). Events are defined as the features, extracted by the human auditory system, which form the basis for perception of different speech sounds. It is these events which make human speech recognition highly robust to noise, as compared to machine recognition (Lippman, 1997). Thus, the use of events should increase the noise robustness of a speech recognition system, and should improve the functionality of hearing aids and cochlear implants.

It is our goal to identify these events by directly comparing the sound confusions with the corresponding masked speech stimuli, on an utterance by utterance basis. We wish to identify the acoustic features in speech which become inaudible when a masked speech sound is confused with other sounds.

Towards this goal we have performed a series of perceptual experiments that involve noise masking, time truncation, and filtering of speech. We employed large numbers of talkers and listeners, to take advantage of the large natural variability in speech production and perception. This paper presents the analysis of the confusion data for one of these noise-masking experiments.

We use the confusion matrix (CM), which is an important analytical tool for quantifying the results of closed-set recognition tasks, to characterize the nature of perceptual confusions (Allen, 2005a). Each entry in the CM, denoted $P_{s,h}(SNR)$, is the empirical probability of reporting sound h as heard when sound s was spoken, as a function of the *signal-to-noise ratio* (SNR). A Bayesian average of the diagonal entries $[P_{s,s}(SNR), h=s]$ gives the conventional "Recognition Score" or "Performance Intensity" (PI) measure $P_c(SNR)$. However, such an average obscures the detailed and important information about the nature of the sound confusions, given by the off-diagonal entries.

The confusion matrix was first used for analyzing speech recognition by Campbell (1910). CMs have been used to analyze confusions among vowel sounds in English [Peterson and Barney (1952), Strange *et al.* (1976), Hillenbrand *et al.* (1995)]. Miller and Nicely (1955) used the CM to analyze the consonant confusions for consonant-vowel (CV) sounds with 16 consonants and one vowel, presented at different levels of white masking noise. In 1955, Miller and Nicely (denoted MN55) collected data with five talkers and listeners, at six SNR levels and 11 filtering conditions. This classic confusion analysis experiment inspired many related and important noise-masking studies, such as Wang and Bilger (1973), Dubno and Levitt (1981), Grant and Walden (1996), and Sroka and Braida (2005).

The MN55 study clearly demonstrated that at low SNR, their consonants form three basic clusters of confusable sounds: Unvoiced, Voiced (non-nasals), and Nasals. As the SNR is increased, the first two clusters split into two

[a)]Parts of this analysis were presented at the ARO Midwinter Meeting 2005 (New Orleans), the Aging and Speech Communication 2005 Conference (Bloomington, IN) and the International Conference on Spoken Language Processing 2006 (Pittsburgh, PA).

[b)]Author to whom correspondence should be addressed. Electronic mail: phatak@uiuc.edu

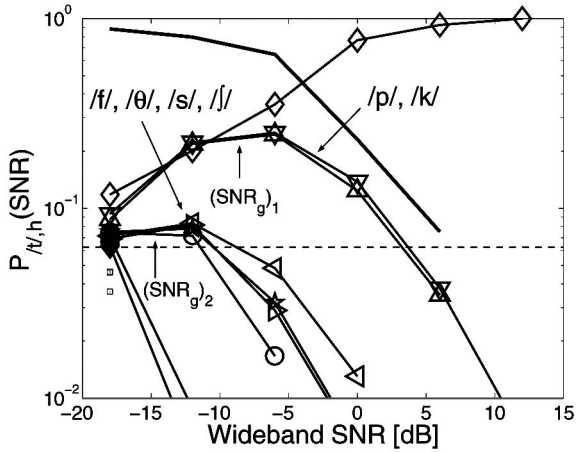
subgroups—plosives and fricatives. Wang and Bilger (1973) extended the CM analysis to more consonants and vowels, but unfortunately their published CM data are pooled over all SNRs, thereby reducing the utility of their database for analyzing perceptual grouping. Dubno and Levitt (1981) compared the acoustic features of syllables with CM data, but they used only two SNR values and did not find common acoustic features that correlate with the confusions at their SNR levels. Grant and Walden (1996) measured the confusions of 18 consonants with auditory and visual cues, but did not provide a confusion analysis, as the primary goal of their study was to investigate the articulation index (AI). Sroka and Braida (2005) measured CMs at several SNRs and filtering conditions for humans and automatic speech recognizers (ASRs), however only one talker was used (either male or female, depending on the syllable), as the primary purpose of their study was to compare the human performance with that of ASRs.

In the consonant CM tables from MN55, the order of the consonants is crucial when viewing or analyzing the formation of such clusters. With a different order of consonants, the perceptual clusters of consonants are not obvious. An alternate clustering method, multi-dimensional scaling, does not depend on the order of consonants but is not stable and does not guarantee a unique solution (Wang and Bilger, 1973). A *confusion pattern* (CP) analysis, defined by a graphical representation of a row (particular value of s) of the CM as a function of SNR, is a simple tool that overcomes all of these difficulties (Allen, 2005b). In this report we use the CPs to further study human speech coding.

A. Confusion patterns

Figure 1 shows the CPs for sound $s=$/tɑ/ from MN55. Each curve corresponds to a particular column entry (h) for the /t/ row, plotted as a function of SNR, namely $P_{/t/,h}(SNR)$. The diagonal entry $P_{/t/,/t/}(SNR)$, denoted by $\diamond$, increases with SNR. As the SNR decreases, confusions of /t/ with /p/ ($\triangle$) and /k/ ($\triangledown$) increase and eventually become equal to the target for SNRs below −8 dB. We say that /t/, /p/ and /k/ form a *confusion group* (or *perceptual group*) at (or near) the *confusion threshold*, indicated by $(SNR_g)_1 \approx -8$ dB, where $(SNR_g)_1$ is the point of local maximum in $P_{/t/,/p/}(SNR)$ and $P_{/t/,/k/}(SNR)$ curves. When the SNR is decreased below $(SNR_g)_2 \approx -15$ dB, consonant group [/f/, /θ/, /s/ and /ʃ/] merges with the [/t/, /p/, /k/] group, forming a super group. Since $(SNR_g)_2 < (SNR_g)_1$, consonants [/p/, /k/] are perceptually closer to /t/, and thereby form a stronger perceptual group with /t/ than the consonants [/f/, /θ/, /s/, /ʃ/]. Thus we use the *confusion threshold* SNR_g as a quantitative measure to characterize the hierarchy in the perceptual confusions.

At very low SNRs, where no speech is audible, all the sounds asymptotically reach the chance performance of 1/16, shown by the dashed line. The remaining nine off-diagonal entries are never confused with the target sound /t/, and as a result never exceed chance (e.g., the small squares).



Consonant	*Marker*	*Consonant*	*Marker*
p	△	b	▲
t	◊	d	♦
k	▽	g	▼
f	☆	v	★
θ	◁	ð	◀
s	○	z	●
ʃ	▷	ʒ	▶
m	✶	n	×

FIG. 1. Confusion patterns (CPs) for $s=$/tɑ/ from MN55. The thick solid line without markers is $1-P_{s,s}(SNR)$, which is the sum of off-diagonal entries. The horizontal dashed line shows the chance level of 1/16. The legend provides the marker style used for consonants. These markers will be used throughout the paper.

B. Experiment UIUCs04

The speech stimuli for the previous CM experiments either do not exist in recorded format, or were not publicly available. Without these speech wave forms, it is not possible to determine the acoustic, and thus the corresponding perceptual features. Thus a number of MN55 related closed-set confusion matrix experiments were conducted at the University of Illinois, using a commercially available database (LDC-2005S22) composed of nonsense sounds having 24 consonants, 15 vowels and 20 talkers. The first of these experiments, reported here and denoted "UIUCs04," used 64 context-free consonant-vowel (CV) sounds (16Cs × 4Vs). Our first goal was to analyze consonant confusions. The purpose of choosing multiple vowels was to analyze the extent of the effect of vowels on the listener's consonant CPs (i.e., the coarticulation effects).

The long-term goal of our data collection exercise is to identify perceptual features by the use of masking noise. Specifically, we wish to determine the acoustic features that are masked near the confusion thresholds. We have also used the natural variability in the confusion thresholds across utterances to identify the acoustic features and events. The analysis in the present paper is limited to consonant and vowel confusions, but not events. We compare our results with past work, and show how the consonant confusions in speech-weighted noise are different from those in white noise. We also show that the observed consonant groups are related to the spectral energy in the consonant above the noise spectrum, and show that the Articulation Index (AI),

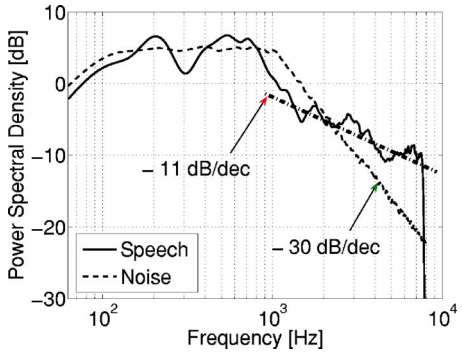


FIG. 2. (Color online) The power spectral densities (PSD) of average speech (solid) and noise (dashed) for UIUCs04 at 0 dB wideband SNR. The PSDs for both speech and noise were calculated using the *pwelch* function in MATLAB, with a hanning window of duration 20 ms (i.e., 320 samples) with an overlap of 10 ms and a fast Fourier transform length of 2048 points.

derived from the speech and noise spectra, is a better metric than the wideband SNR to characterize and compare the consonant scores.

II. METHODS

A. Stimuli

A subset of isolated CV sounds from the LDC-2005S22 corpus (Fousek *et al.*, 2004), recorded at the Linguistic Data Consortium (University of Pennsylvania), was used as the speech database. This subset had 18 talkers speaking CVs composed of one of the 16 consonants (/p/, /t/, /k/, /f/, /θ/, /s/, /ʃ/, /b/, /d/, /g/, /v/, /ð/, /z/, /ʒ/, /m/, /n/) and followed by one of the four vowels (/ɑ/, /ɛ/, /ɪ/, /æ/). The vowels were chosen to have formant frequencies close to each other, with the goal of making them more confusable. All talkers were native speakers of English, but three talkers were bilingual and had a part of their upbringing outside the U.S./Canada. Ten talkers spoke all 64 CVs, while each of the remaining eight talkers spoke different subsets of 32 CVs, such that each CV was spoken by 14 talkers.

MN55 had five female talkers, who also served as the listeners. Because the power spectrum for average speech [Dunn and White (1940); Benson and Hirsh (1953); Cox and Moore (1988)] has a roll-off of about −29 dB/dec (≈−8.7 dB/oct) above 500 Hz, the white noise masks the high frequencies in speech to a greater extent than low frequencies. A noise signal that has a spectrum similar to the average speech spectrum would mask the speech uniformly over frequency. Such a speech-weighted noise, shown in Fig. 2, was used as masker in UIUCs04. The noise power spectrum was constant from 100 Hz to 1 kHz, with a roll-off of 12 dB/dec (≈3.6 dB/oct) and −30 dB/dec (≈−9.0 dB/oct) on the lower and higher sides, respectively. The noise was generated by taking the inverse Fourier transform of the magnitude spectrum, obtained from this power spectrum, combined with a random phase. The rms level of this noise was then adjusted according to the level of the CV sound to achieve the desired SNR. The average spectrum of CV sounds (Fig. 2) was found to have a different roll-off characteristic than the making noise spectrum. The roll-off of the average speech for this experiment was −30 dB/dec between 800 Hz and 1.5 kHz but then it reduced to about −11 db/dec (≈−3.3 dB/oct), resulting in a high-frequency SNR boost. The change in the slope above 2 kHz can also be observed in the speech spectrum from several studies [Byrne *et al.* (1994); Grant and Walden (1996)].

A new random noise with the desired spectral characteristics was generated for each presentation, and the wideband noise rms level was adjusted according to the rms level of the CV sound to be presented, to achieve the precise SNR. While calculating the rms level of a CV utterance, the samples below −40 dB with respect to the largest sample were not considered.

The CV sounds were presented in speech-weighted masking noise at six different signal-to-noise ratios (SNR): $[-22,-20,-16,-10,-2, Q]$ dB, where Q represents the quiet condition. The sum of speech signal and masking noise was filtered with a bandpass filter of 100 Hz–7.5 kHz before presentation. The highest amplitude of the bandpass filtered output (i.e., speech plus noise) was scaled to make full use of the dynamic range of the sound card, without clipping any sample.

B. Testing paradigm

The listening test was automated using a MATLAB code with graphic user interfaces. The listener was seated in a sound booth in front of a computer monitor. The computer running the MATLAB code was placed outside the sound-treated booth to minimize ambient noise. The monitor screen showed 64 buttons, each labeled with one of the 64 CVs. The 64 buttons were arranged in a 16×4 table such that each row had the same consonant while each column had the same vowel. An example of the use of each consonant or vowel in an English word was displayed as the pronunciation key at the left of the rows and at the top of the columns. Listeners heard the stimuli via headphones (Sennheiser, HD-265) and entered the response by clicking on the button labeled with the identified CV. The listener was allowed to replay the CV sound as many times as desired before entering the response. Repeating the sound helped to improve the scores by eliminating the unlikely choices in the large 64-choice closed-set task. Repeating the sound also allows the listener to recover from the distractions during the long experiment. For each repetition, a new noise sample was generated. After entering the response, the next sound was played following a short pause.

In addition to the 64 buttons, the listener had an option of clicking another button, labeled "Noise Only," to be used only when the listener could not hear any part of the masked speech. The listeners were periodically instructed to use this button only when no speech signal was heard, and to guess the CV otherwise. The primary purpose to allow the Noise Only response was to remove the listener biases. The Noise Only responses for a CV were treated as "chance-level" responses and were distributed uniformly over the 64 columns, corresponding to 64 possible options, in the row of that CV.

Each presentation of CV sound was randomized over consonants, vowels, talkers, and SNRs. The total 5376 presentations ($16C\times4V\times14$ talkers$\times6$ SNRs) were random-

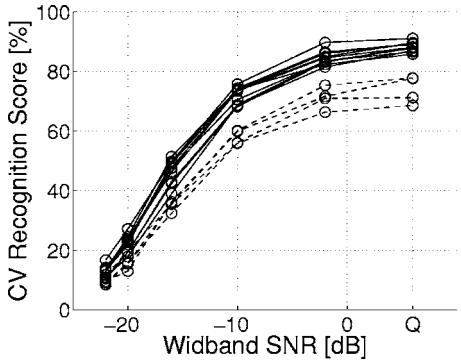


FIG. 3. The CV recognition scores of 14 listeners, as a function of SNR. Dashed lines show the four low performance (LP) listeners. The quiet condition is denoted by "Q."

ized and split into 42 tests, each with 128 sounds. Each listener was trained using one or two practice tests with randomly selected sounds, presented in Quiet, with visual feedback on the correct choice.

C. Listeners

Fourteen L1=English listeners (6M, 8F), ten having American accents and one with Nigerian accent, completed the experiment. All listeners, except one with age of 33 years, were between 19 and 24 years. They had no history of hearing disorder or impairment, and self-reported to have normal hearing. The listeners were verified to be attending to the experimental task, based on their scores, as described in the next section. The average time for completing the experiment was 15 h per listener.

III. RESULTS

Before analyzing the perceptual confusions, it is necessary to verify that the listeners attended to the required task and that the speech database was error free. We select a homogeneous group of listeners, based on their syllable recognition scores. In order to analyze the effect of noise on the perceptual confusions, we must verify that the utterances are heard correctly in the quiet condition, as a control. The mislabeled utterances can contaminate the perceptual confusions in noise. Therefore, based on the syllable errors in quiet, we select the low-error utterances that we use for analyzing perceptual confusion in noise. Following listener and utterance selection, we analyze the confusions of the CV syllables, as well as those of individual consonants and vowels. Finally, we compare our results with the past work from literature.

A. Listener selection

Ten "High Performance" (HP) listeners (i.e., listeners with scores greater than 85% in quiet, and greater than 10% correct at −22 dB SNR), shown by solid lines in Fig. 3, formed a homogeneous group. The scores of these HP listeners (5M, 5F) were comparable to the average score of NH listeners from other confusion matrix studies.[1] Responses of the four "Low Performance" (LP) listeners (dashed lines) were not considered for the subsequent analysis. All HP listeners had American accents (five Midwest, one New York, and four unspecified).

To investigate the sources of low scores for the LP listeners, their errors in the Quiet condition were further analyzed. All four LP listeners had 10–21% vowel errors, while three of the four LP listeners had 14–15% consonant errors in quiet. The average consonant and vowel errors for the HP listeners, in quiet, were 8% and 4%, respectively. For LP listeners, 61–72% of the consonant errors were for consonants /θ/, /v/ and /ð/, while the vowel errors were consistently high for vowel /æ/. The vowel sound /æ/ was mainly confused with /ɑ/. For the remaining consonants and vowels, scores of all 14 listeners were comparable. The pronunciation keys "/TH/ as in *THick*" and "/th/ as in *that*" and the labels TH and th were used for consonant sounds /θ/ and /ð/, respectively. It is possible that the four LP listeners confused the labels of these two consonants, which have the same spelling. However, the LP listeners confused /θ/ more with /f/ than with /ð/. Also, /ð/ was confused equally with /θ/ and /v/. Therefore, the most likely reason for the bad performance of the LP listeners is their inability to distinguish among consonants /f/, /θ/, /v/, /ð/, and between vowels /æ/ and /ɑ/.

B. Utterance selection

The syllable error e_n for each of the 896 utterances ($1 \leqslant n \leqslant 896$) was estimated from listener responses in the quiet condition. A syllable error occurs when a listener reports an incorrect consonant or an incorrect vowel, or both. These errors can be estimated from the CM as $e_n = 1 - P_{s,s}(n,\text{quiet}) = \Sigma_h^{s \neq h} P_{s,h}(n,\text{quiet})$, where $P_{s,s}(n,\text{quiet})$ is the diagonal element of the CM, representing correct recognition for utterance n. The syllable errors were calculated for all listeners, as well as for the 10 HP listeners. When responses of the 10 HP listeners were pooled, 59% of the total 896 utterances had no errors in quiet. These are the well-formed or "good" utterances. However, some utterances had very high errors; ten utterances had 100% error. Some of these high error utterances, which had $e_n > 80\%$, were consistently reported as another CV sound and therefore are better described as mislabeled. The responses to the other high error utterances were mostly incorrect and inconsistent. Such high error sounds, which were unaffected by listener selection, are inherent to the database.

Since the errors in the speech database (i.e., the high error sounds) could be misinterpreted as the listener confusions while analyzing the listener responses, they would contaminate the perceptual confusions in noise. Therefore, 146 utterances (≈16% of the 896 utterances), which had more than 20% errors, are defined as "confusable" utterances, and the responses to these utterances were removed before generating CMs. Removal of the confusable utterances improved the consonant recognition scores by greater margin ($91.6\% \rightarrow 98.0\%$) than the vowel recognition scores ($96.0\% \rightarrow 98.2\%$). The utterances with $0 < e_n < 20\%$ are defined as the "marginally confusable" utterances and were considered for analysis. Figure 4 shows the distribution of

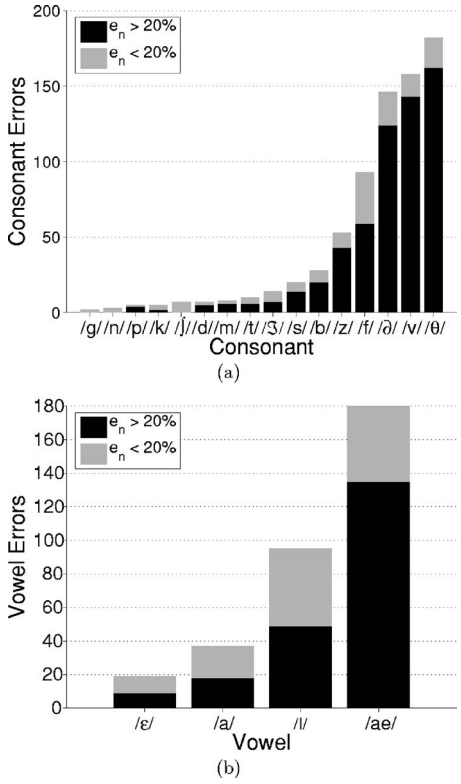


FIG. 4. Histograms of the incorrect recognition of (a) consonants and (b) vowels, in the "confusable" ($e_n > 20\%$) and "marginally confusable" ($0 < e_n < 20\%$) utterances, where e_n is the syllable error for that utterance in the quiet. There are 560 responses for each consonant (4 vowels $\times$ 14 talkers $\times$ 10 HP listeners) while there are 2240 responses for each vowel (16 consonants $\times$ 14 talkers $\times$ 10 HP listeners) in the quiet condition.

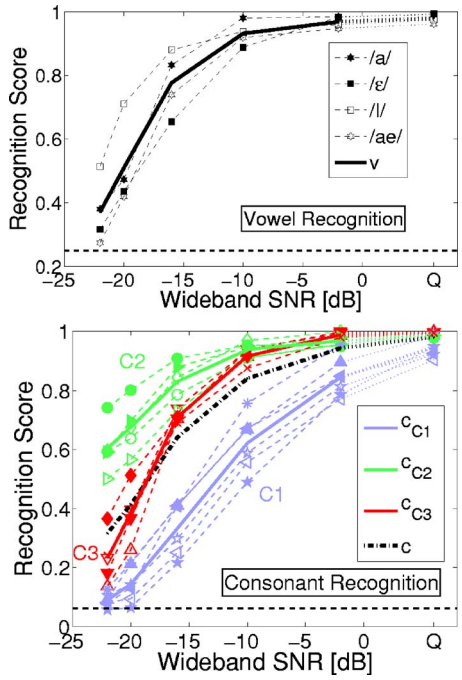


FIG. 5. (Color online) Recognition scores for vowels (*top*) and consonants (*bottom*), as a function of wideband SNR. In the top plot, the thick solid line is the average vowel recognition score (v), while in the bottom plot, the three solid lines represent the average scores for the three consonant sets and the dash-dotted line represents the average consonant score (c). The chance levels, 1/4 for vowels and 1/16 for consonants, are shown by the horizontal dashed black lines.

the total number of errors per consonant (left panel) and per vowel (right panel), in the confusable utterances ($e_n > 20\%$) as well as for the marginally confusable utterances ($e_n < 20\%$). Most of these errors occurred for the five consonants /θ/, /v/, /ð/, /f/, /z/, and for the two vowels /æ/, /ɪ/ (Fig. 4). These consonants and vowels were also the ones for which the LP listeners performed very poorly relatively to the HP listeners (Sec. III A, last paragraph). This suggests that the reason for poor performance of the LP listeners was their inability to recognize the confusable sounds. It is possible that the LP listeners perform as well as HP listeners for the marginally confusable and good utterances, in which case, LP listener data would be useful. However, the score of LP listeners for marginally confusable utterances was found to be even lower than that of HP listeners. The /θ/ → /f/ and /ð/ → /v/ confusions of the LP listeners decreased significantly after removing the confusable utterances, but /θ/ ↔ /ð/ and /æ/ → /ɑ/ confusions did not show a similar decrease. The recognition scores of the LP listeners for /θ/, /ð/ and /æ/ increased, but still remained lower than those of the HP listeners. The LP listeners had 4–8% consonant error and 7–19% vowel error after the utterance selection, as compared to the HP listeners, who had less than 2% consonant and vowel errors. All subsequent analysis uses 10 HP listener responses to the marginally confusable and good utterances.

C. Recognition scores

The top panel in Fig. 5 shows the recognition scores for the four vowels, as well as the average vowel score (thick solid line), as a function of SNR. Except for the slightly higher scores of vowel /ɪ/ in presence of noise, the vowel scores are approximately equal. The previous studies show that /ɪ/ scores are relatively greater in a masking noise with speech-like spectrum, possibly due to a higher F_2 value (Gordon-Salant, 1985). The speech-weighted noise, therefore, seems to uniformly mask the four vowels.

One of the most interesting observations in this study is that the recognition scores of consonants show three groups (Fig. 5, bottom). One set of curves, shown in blue color, has relatively low scores, approaching the chance level of 1/16 below −20 dB. This set, which we call C1, contains consonants /f/, /θ/, /v/, /ð/, /b/ and /m/. In contrast, the consonants /t/, /s/, /z/, /ʃ/ and /ʒ/, which form set C2 (green lines) are high-scoring consonant and have scores greater than 50% even at −22 dB SNR. The remaining consonants (/n/, /p/, /g/, /k/ and /d/), grouped into set C3 (red lines), have relatively high scores, close to set C2 scores, above −10 dB SNR. However, the scores of C3 consonants drop sharply below −10 dB SNR, approaching the C1 scores at −22 dB.

The separation of the three sets of consonant curves is more evident in the vowel-to-consonant recognition ratio ($\lambda \equiv v/c$) plots shown in Fig. 6. The ratio λ was first used by Fletcher and Galt (1950) to compare the consonant and vowel performances. Figure 6 shows the values of $\lambda_i = v/c_i$, where v is the average vowel score and c_i represents the scores of individual consonants. The dash-dot line shows the

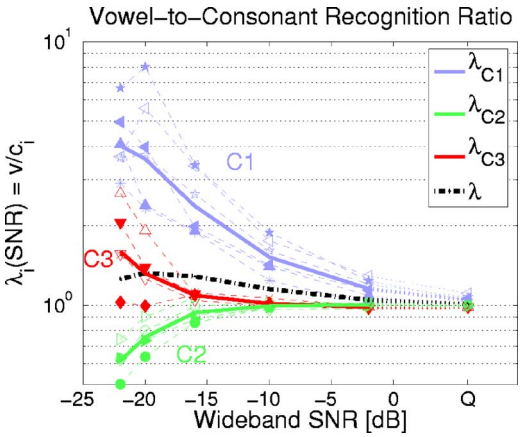


FIG. 6. (Color online) Vowel-to-consonant recognition ratio (on a log scale) as a function of SNR [$\lambda(SNR)$], for each consonant. The color and the markers denote the same information as that in the bottom panel of Fig. 5. The average $\lambda(SNR)$ for the consonant sets are shown by the thick solid lines, while that for average consonant score is shown by the thick, dash-dotted line.

average value of λ, which was just above unity. The C1 consonants have $\lambda_i(SNR)$ curves that well above 1 even for small amounts of noise, while those for set C3 stay close to unity for wideband SNRs $\geqslant -16$ dB, but rise sharply below that. The below-unity values of $\lambda_i(SNR)$ for C2 consonants in speech-weighted noise contradict the traditional assumption that the vowels are always better recognized in noise than consonants.

These consonant groups are also observed in the Grant and Walden (1996) data, which were collected in a speech-weighted noise masker, but not observed in the confusion data of Miller and Nicely (1955). Thus, while the white noise masks the 16 consonants almost uniformly, the speech-weighted noise has a nonuniform masking effect for consonants, masking set C1 more than set C2. This is further discussed in Sec. III G.

1. Articulation index (AI)

Allen (2005b) showed that the MN55 recognition scores for 11 of the 16 consonants, as well as the average consonant scores, can be modeled as

$$P_C(AI) = 1 - e_{\text{chance}} e_{\text{min}}^{AI}, \tag{1}$$

where AI is the articulation index, $e_{\text{min}} = 1 - P_C(AI=1)$ is the recognition error at $AI=1$ and $e_{\text{chance}} = 1 - 1/16$ is the error at chance ($AI=0$). Based on this relation, the log-error $\log(1-P_c(AI)) = AI \log(e_{\text{min}}) + \log(e_{\text{chance}})$ is a linear function of the AI. The AI, which is based on the SNRs in articulation bands, accounts for the shapes of signal and noise spectra (French and Steinberg, 1947). The articulation bands were estimated to contribute equally to the recognition context-free speech sounds (Fletcher, 1995). Allen (2005b) refined the AI formula to be

$$AI = \frac{1}{K}\sum_{k=1}^{K} \min\left[\frac{1}{3}\log_{10}(1 + r^2 snr_k^2), 1\right], \tag{2}$$

where snr_k is the SNR (in linear units, not in dB) in kth articulation band, $K=20$ is the number of articulation bands, and r is a factor that accounts for the peak-to-rms ratio for the speech.[2] The peak-to-rms ratios for the CV sounds used in UIUCs04, estimated using the method described in Appendix A, were found to vary over articulation bands. Therefore, a frequency-dependent value of r, denoted as r_k, was used for estimating AI. The resulting expression for the AI becomes

$$AI = \frac{1}{K}\sum_{k=1}^{K} \min\left[\frac{1}{3}\log_{10}(1 + r_k^2 snr_k^2), 1\right], \tag{3}$$

where r_k values are directly estimated from the speech stimuli (Appendix A).

The AI values were calculated for all SNRs, except the quiet condition, using the same 20 articulation bands ($K=20$) as those specified by Fletcher (1995) and used by Allen (2005b). The AI for the quiet condition cannot be directly estimated, as the actual SNR for that condition is not known.

Figure 7(a) shows the individual consonant recognition scores of UIUCs04 data, plotted as a function of AI. The average recognition scores (c, dash-dotted line) match very closely with the predictions of the AI model $1-e_{\text{min}}^{AI}$, with $e_{\text{min}}=0.003$ (black solid curve). Following the transformation from the wideband SNR to the AI scale, the recognition

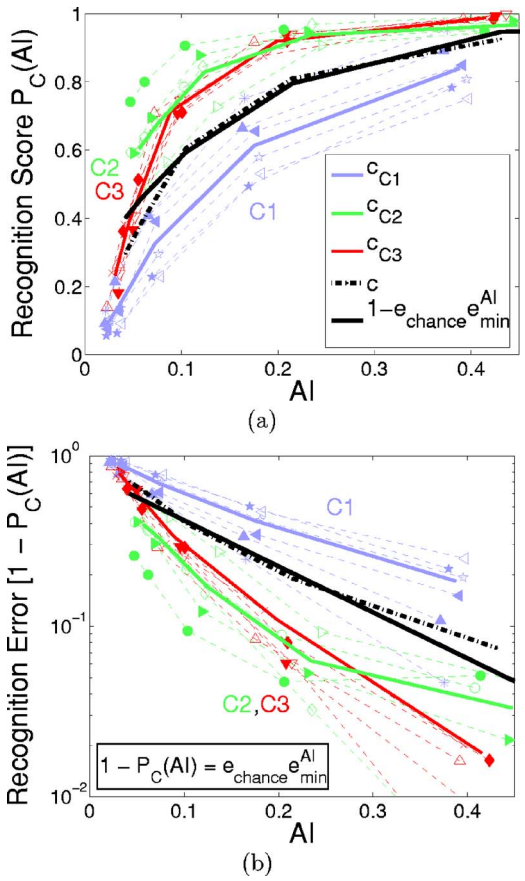


FIG. 7. (Color online) (a) Consonant recognition scores $P_C(AI)$, and (b) consonant recognition error $1-P_C(AI)$ (Log scale), plotted as a function of AI. The dashed lines represent individual consonants, while the three colored solid lines represent average values for the three consonant sets. The average consonant score (thick dash-dotted line) is very close to that predicted by the AI model $P_C(AI) = 1 - e_{\text{chance}} e_{\text{min}}^{AI}$ (thick solid line). The data for the Quiet condition are not shown.

scores for sets C2 and C3 nearly overlap. This is because the AI accounts for the spectral differences between sets C2 and C3 (see Sec. III D 1). However, the curve for C1 scores remains lower than the other two sets. We therefore conclude that, in addition to the spectral differences, there are other differences between sets C1 and C2, which cannot be accounted by Eq. (3).

Figure 7(b) shows that the log errors for all consonants, with the exception of four C2 consonants (green lines), are linear functions of the AI with different slopes. The slopes are given by the $\log(e_{\min})$ for each consonant. The C1 consonants, with the exception of /m/ (∗), have significantly higher $e_{\min}$ values than the remaining consonants. The approximate $e_{\min}$ values for sets C1, C2, and C3 are 0.01, 2×10^{-5}, and 3×10^{-5}, respectively. This explains why the curves of C1 consonants do not overlap with those of C2 and C3 consonants. Note that C1 consonants were the most frequent among the confusable utterances (Fig. 4). Thus, the $e_{\min}$ for the C1 consonants would be even higher without utterance selection. High $e_{\min}$ values for the C1 consonants are also observed in our analysis of the Grant and Walden (1996) data (see Sec. III G).

The total recognition error $1-P_C$ (black dash-dotted line), which is the average of errors for the three sets (colored solid lines), can be expressed as

$$1-P_C(AI)=\frac{1}{3}[e^{AI}_{\min,C1}+e^{AI}_{\min,C2}+e^{AI}_{\min,C3}]e_{\text{chance}} \tag{4}$$

$$=\frac{1}{3}[(0.01)^{AI}+(2\times10^{-5})^{AI}+(3\times10^{-5})^{AI}]e_{\text{chance}}. \tag{5}$$

Since the total error is a sum of exponentials with different bases, it need not be an exponential. However, in this case, the exponential model $e_{\text{chance}}e^{AI}_{\min}$ with $e_{\min}=0.003$ (solid black line) fits very closely to the average error $1-P_C(AI)$.

D. Confusion analysis

In this section, we analyze the individual confusions. Figure 8 shows the 64×64 row-normalized syllable CM at four different SNRs, displayed as gray-scale images. The intensity is proportional to the log of the value of each entry in the row-normalized CM, with black representing a value of unity and white representing the chance-level probability of 1/64. The rows and columns of the CM are arranged such that four CVs having the same consonant are consecutively placed with vowels /ɑ/, /ɛ/, /ɪ/, and /æ/, in that order. The consonants are stacked according to the sets C1, C3, and C2, separated by dashed lines.

For SNR$\leqslant-16$ dB, $\lambda<1$ for the C2 consonants implies that the syllables with C2 consonants have more vowel confusions than the consonant confusions. This shows up in the CM images as the blocks around the diagonal for the CVs with C2 consonants. On the other hand, $\lambda>1$ for the C1 consonants results in lines parallel to the diagonal in the C1 part of the CM images. The parallel-line structure is prominent at SNRs>-20 dB, however as the SNR decreases, vowel confusions appear, smearing the parallel lines.

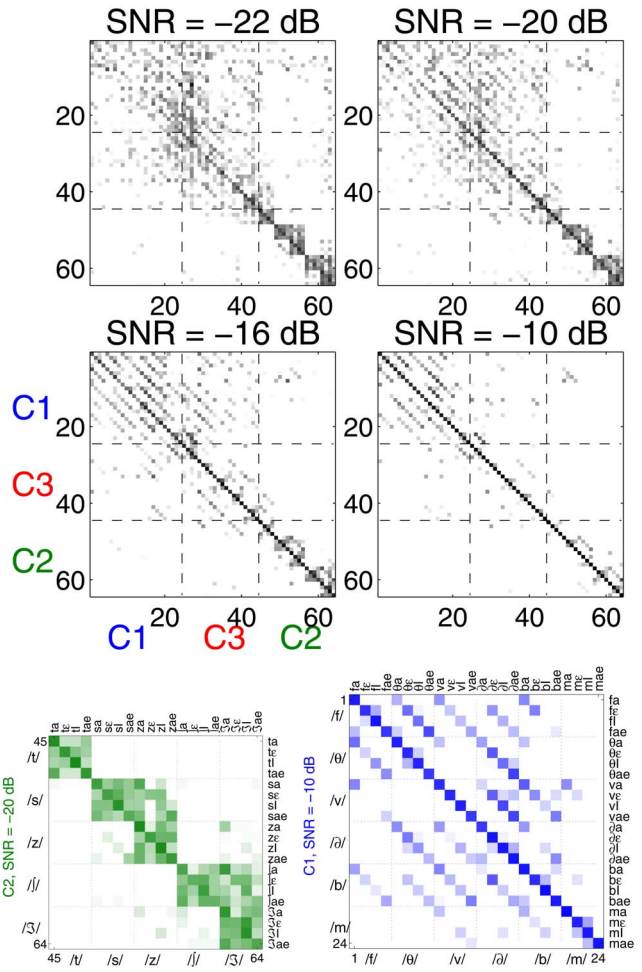


FIG. 8. (Color online) The four (2×2) small panels show the gray-scale images of the CMs at four SNR values. The gray-scale intensity is proportional to the log of the value of each entry in the row-normalized CM, with black color representing unity and white color representing the chance performance (1/64). Dashed lines separate sets C1 (Nos. 1–24), C3 (Nos. 25–44), and C2 (Nos. 45–64), in that order, from left to right and top to bottom. The two enlarged color panels at the bottom show set C2 at −20 dB SNR and set C1 at −10 dB SNR.

Sets C1 and C3 are confused with each other, while set C2 has negligible confusions with the other two sets. This correlates with the spectral powers of the consonants in the three sets (see Sec. III D 1). The asymmetry in the C1–C3 confusions, especially at −20 dB SNR, can be easily explained based on the recognition performances of C1 and C3 consonants. At −20 dB SNR, the C1 consonants are very close to chance level, while C3 consonants have scores between 20% and 50%. Thus, C1 consonants are confused with C3 consonants but not vice versa, which gives rise to the asymmetric confusions between sets C1 and C3.

Within set C2, there are asymmetric confusions between /s/-/z/ and /ʃ/-/ʒ/. This asymmetry is further investigated in Sec. III E. A few vertical lines can be observed in the CM images, suggesting some kind of bias towards certain CVs, however there is no consistent trend in terms of consonants or vowels in these lines.

1. Consonant PSD analysis

The nature of the confusions among the three consonant sets correlate with the SNR spectrum of the consonants. The

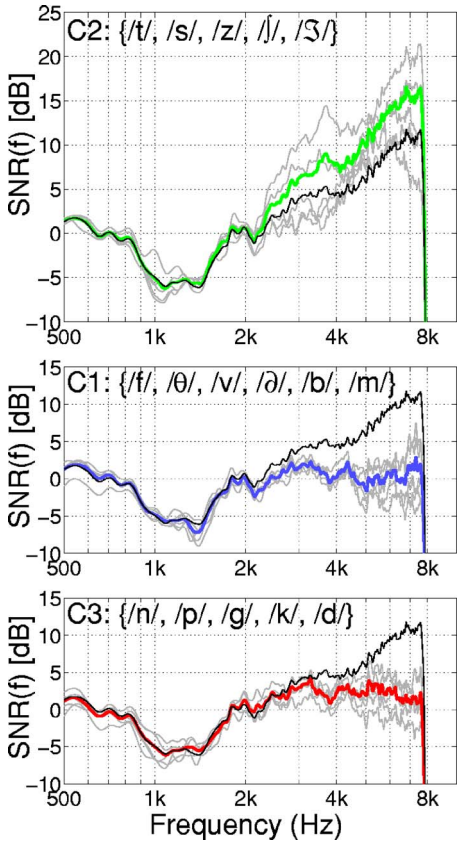


FIG. 9. (Color online) The SNR spectra [SNR(f)] for consonants in set C2 (*top*), C1 (*center*), and C3 (*bottom*). The thin gray lines show the SNR spectra for individual consonants while the thick colored line in each panel shows the average SNR spectrum for that set. Each panel contains the SNR spectrum for average speech (thin black line), estimated using the speech and noise spectra shown in Fig. 2.

SNR spectrum for a consonant is defined here as the ratio of power spectral density (PSD) of that consonant to the PSD of the noise. To estimate the PSD of a consonant, the PSD of all CV utterances with the given consonant were averaged. Such an average would practically average out the spectral variations due to the four different vowels and enhance the consonant spectrum.

Figure 9 shows the SNR spectra for consonants (thin gray lines) in sets C2 (top), C1 (center), and C3 (bottom) at 0 dB wideband SNR. Each panel shows the average SNR spectra for that set (thick colored line), as well as the SNR spectra for the average speech (thin black line). The average speech PSD starts to roll-off at 800 Hz, while the noise PSD is flat up to 1 kHz (Fig. 2). The speech PSD crosses over the noise PSD at about 2 kHz. Correspondingly, the SNR spectra for average speech has a valley between 500 Hz and 2 kHz. Above 2 kHz, speech dominates the noise, resulting in the high-frequency boost in the SNR spectrum that is more than 10 dB above 6 kHz.

The C2 consonants have rising SNR spectra at high frequencies, while those of C1 and C3 either remain flat or slightly drop at higher frequencies, in spite of the high-frequency boost. The high-frequency energy makes the SNR spectra of C2 consonants significantly different from the other consonants, resulting in high scores and very few confusions for the C2 consonants. The SNR spectra of C1 consonants are indistinguishable from those of C3 consonants, which explains why C1 and C3 consonants are confused with each other, but not with C2 consonants.

2. Confusion matrices

There are 64 curves in each CV confusion pattern for the 64×64 CM, which makes it very difficult to analyze the confusions. Also, the row sums for the 64×64 CM are not large enough to obtain smooth curves in the confusion patterns. Therefore, we analyze the consonant and vowel confusions separately. We will also analyze the interdependence of consonant and vowel confusions.

To analyze the consonant confusions, the responses were scored for consonants only. This resulted in a 64×16, syllable-dependent consonant CM $P(C_h|C_sV_s)$. Averaging the rows of this CM over the spoken vowel gives a 16×16 vowel-independent consonant CM $P(C_h|C_s)$. Similar CMs can be generated to analyze vowel confusions. Five such CMs are listed in Table I, including the two CMs that are generated for vowel analysis, which will be discussed in Sec. III F.

E. Consonant confusions

The perceptually significant consonant confusions (i.e., those with a well-defined SNR_g) observed in UIUCs04 are /m/-/n/ (set C3), /f/-/θ/, /b/-/v/-/ð/, /θ/-/ð/ (set C1), /s/-/z/ and /ʃ/-/ʒ/ (set C2). Note that each of these confusion groups is within one of the three sets. At −2 dB SNR, more than 84% of the consonant confusions are within the three consonant sets. The consonant confusions across the three sets increase with decrease in SNR, but are mostly between sets C1 and C3.

The consonant confusions for C2 consonants do not depend on the following vowel, as $\lambda<1$ for set C2. When C2

TABLE I. Mathematical expressions, sizes, and descriptions of the five basic types of CM used in this study. C_s and V_s indicate the spoken consonant and vowel, while C_h and V_h represent the consonant and vowel reported by the listener.

CM description	Size	Expression
Syllable (CV) confusions	64×64	$P(C_hV_h\|C_sV_s)=P_{s,h}(SNR)$
CVs scored on consonants	64×16	$P(C_h\|C_sV_s)=\Sigma_{V_h}P(C_hV_h\|C_sV_s)$
Consonant confusions	16×16	$P(C_h\|C_s)=\Sigma_{V_s}P(C_h\|C_sV_s)$
CVs scored on vowels	64×4	$P(V_h\|C_sV_s)=\Sigma_{C_h}P(C_hV_h\|C_sV_s)$
Vowel confusions	4×4	$P(V_h\|V_s)=\Sigma_{C_s}P(V_h\|C_sV_s)$

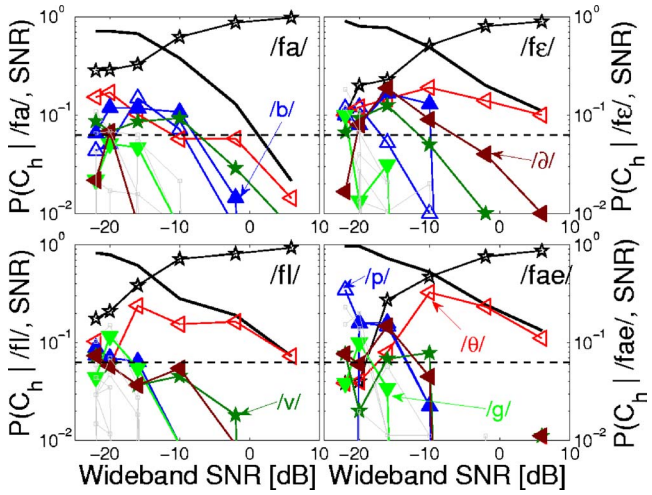


FIG. 10. (Color online) Consonant CPs $P(C_h|C_sV_s,SNR)$ (64×16) for C_s = /fɑ/ (top left), /fɛ/ (top right), /fɪ/ (bottom left), and /fæ/ (bottom right). $P(C_h|/fV_s/,SNR)$ are four rows of the 64×16 consonant CM that correspond to presentation of consonant /f/ and vowel V_h at a given SNR. The gray thin lines with square symbols in the CP figures represent the sounds that are not confused with the diagonal sound and hence do not cross above the chance level. In all CP figures, the quiet condition is plotted at +6 dB SNR for convenience.

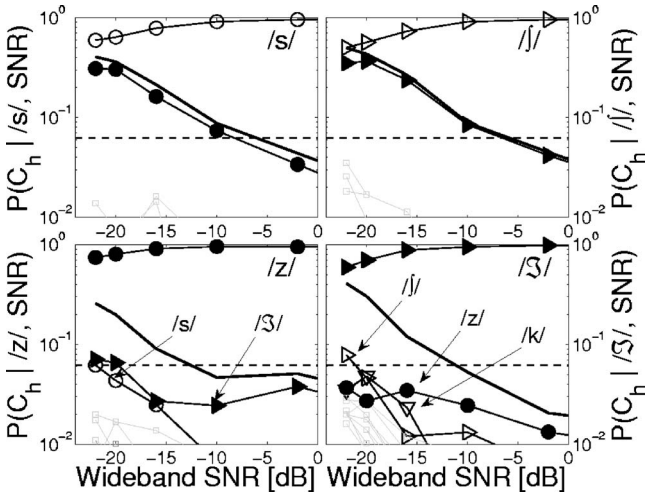


FIG. 11. Consonant CPs $P(C_h|C_s,SNR)$ (16×16) for consonants /s/ (top left), /ʃ/ (top right), /z/ (bottom left), and /ʒ/ (bottom right). The unvoiced consonants (top panels) have only one strong competitor which accounts for the total recognition error (thick solid line), while the voiced consonants have multiple competitors that contribute to the total error.

consonants start to get confused (SNR ≤ −20 dB), the vowels are hardly recognizable (Fig. 5) and are very close to being inaudible. The vowels can affect the consonant confusions only if they have high recognition when the consonants are being confused. Thus, only for consonants with λ > 1 (sets C1 and C3), the CPs can depend on the following vowel.

1. Vowel-dependent consonant confusions

The vowel-dependent 64×16 consonant CM $P(C_h|C_sV_s)$ showed that the CPs for some consonants depend on the spoken vowel V_s. Figure 10 shows the CPs for the four CV sounds with consonant /f/. The strongest competitor /θ/ (symbol ◁) stood out from the other competitors for the sounds /fɪ/ and /fæ/; it was closely accompanied by the secondary competitors (/b/, /ð/, and /v/) in case of /fɛ/, while it was buried as a secondary competitor of /fɑ/. Identical trends were observed for consonants /θ/, /v/, /ð/, and /m/ (all in C1). For /b/ and some C3 consonants, the CPs varied with V_s, but the variations had no specific identifiable trend.

2. Vowel-independent consonant confusions

Since the CPs for four C2 consonant /s/, /ʃ/, /z/, and /ʒ/ are independent of the spoken vowel they could be averaged across V_s. Figure 11 shows the corresponding four rows of the vowel-independent 16×16 consonant CM, $P(C_h|C_s)$. The /s/-/z/ and /ʃ/-/ʒ/ confusions are highly asymmetric (Fig. 8, set C2). The total error in recognizing unvoiced consonants /s/ and /ʃ/ can be accounted by the confusions with the voiced consonants /z/ and /ʒ/, respectively, whereas /z/ and /ʒ/ have multiple competitors that contribute to the total error. Thus the asymmetry is biased towards the voiced consonants /z/ and /ʒ/, i.e., these two are the preferred choices in /s/-/z/ and /ʃ/-/ʒ/ confusions in speech-weighted noise. The asymmetric parts of the confusion probability are as high as 0.13 and 0.14 for /s/-/z/ and /ʃ/-/ʒ/ confusions, respectively. This asymmetry is slightly greater than the largest asymmetry found in the consonant CM of MN55, which was 0.1, but for a different set consonant confusion pair (Allen, 2005b).

F. Vowel confusions

The vowel confusions were analyzed using the 64×4 consonant-dependent vowel CM $P(V_h|C_sV_s)$ (Table I) and were found to be independent of the preceding consonant C_s. Therefore, the vowel confusions were averaged over C_s, giving the 4×4 vowel CM $P(V_h|V_s)$.

Figure 12 shows these consonant-independent vowel CPs $P(V_h|V_s)$. At very low SNR values, all entries in the 4×4 vowel CM converge to the chance level performance for recognizing the vowels (1/4). The recognition score for each of the four vowels was greater than 30% at −22 dB SNR, not low enough to see clear groupings having a well-formed SNR_g, with the exception of $P(/ɪ/|/ɛ/)$ (top right panel, Fig. 12). However, the off-diagonal entries show some interesting

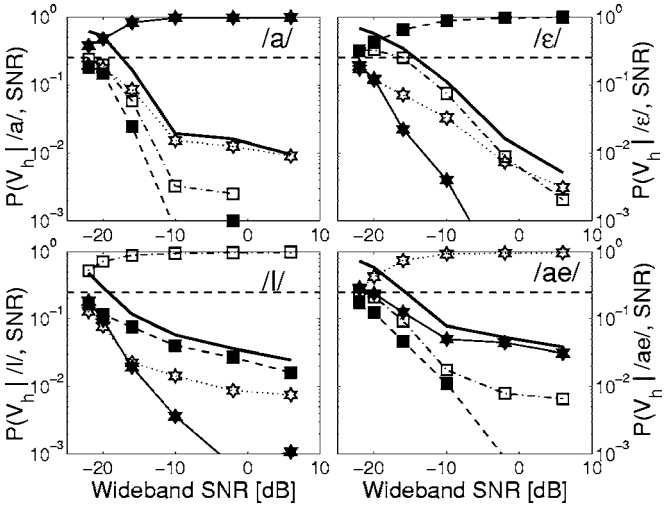


FIG. 12. Consonant-independent 4×4 vowel CPs $P(V_h|V_s,SNR)$ The legend for vowel symbols is given in Fig. 5, top panel.

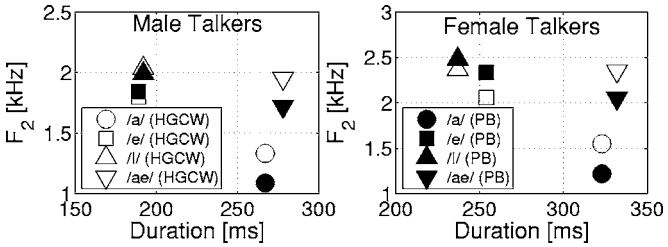


FIG. 13. Plots of the average values of the second formant frequency (F_2) of vowels vs the vowel durations for male (left panel) and female talkers (right panel). The values of the duration are from Hillenbrand *et al.* (HGCW), while the values of F_2 are from HGCW (hollow symbols) as well as Peterson and Barney (PB) (filled symbols), estimated using isolated /hVd/ syllables.

behavior, at scores that are an order of magnitude smaller than the chance level. Due to the very large row sums (1700–2000 responses), the data variability is relatively small, resulting in well-defined curves, even at such low values.

At very low SNR, each vowel seems to be equally confused with the other three vowels, except for /ɛ/, which clearly formed a group with /ɪ/ ($SNR_g \approx -20$ dB, top right panel of Fig. 12). But as the SNR is increased, /ɛ/ becomes equally confused with /ɪ/ and /æ/, though the total number of confusions decrease. For the other three vowels, the curves of the off-diagonal entries separate, showing a clear rank ordering in the confusability. Above −10 dB, /æ/ and /ɑ/ emerge to be the strongest competitors of each other (top left and bottom right panels), with /ɪ/ being the next stronger competitor and /ɛ/ being the weakest competitor for both vowels. The vowel /ɛ/ is the strongest competitor of /ɪ/ above −20 dB (bottom left panel), with /æ/ as the second strongest competitor.

Thus, the four vowels seem to fall into two perceptual groups: {/ɑ/, /æ/} and {/ɛ/, /ɪ/}. These two groups correlate with the vowel durations (Hillenbrand *et al.*, 1995), i.e., /ɑ/-/æ/ are long, stressed vowels, while /ɛ/-/ɪ/ are short and unstressed. Vowel /æ/ is a stronger competitor than /ɑ/ for the short vowels /ɛ/ and /ɪ/ at SNR ⩾−16 dB. This relates to the second formant frequencies of the vowels [Peterson and Barney (1952); Hillenbrand *et al.* (1995)], which would be audible at higher SNRs. Figure 13 shows the vowel durations measured by Hillenbrand *et al.* (HGCW) versus the second formant frequencies measured by HGCW and Peterson and Barney (PB) for our four vowels, categorized by the talker gender. The vowel group /ɛ/-/ɪ/, which was the only vowel group with a clear SNR_g, is much more compact than the /ɑ/-/æ/ group in the Duration-F_2 plane.

1. Vowel clustering

A principal component analysis was performed on the 4×4 vowel CM $[P(V_h|V_s)]$ to analyze the grouping of vowels. The four dimensions of the eigenvectors were rank ordered from 1 to 4 in the decreasing order of the corresponding eigenvalues. The highest eigenvalue was always unity since the vowel CM was row normalized and the coordinates of the four vowels along the corresponding dimension (i.e., Dimension 1) were identical (Allen, 2005a). Figure 14(a) shows the four vowels in the vector space of the remaining

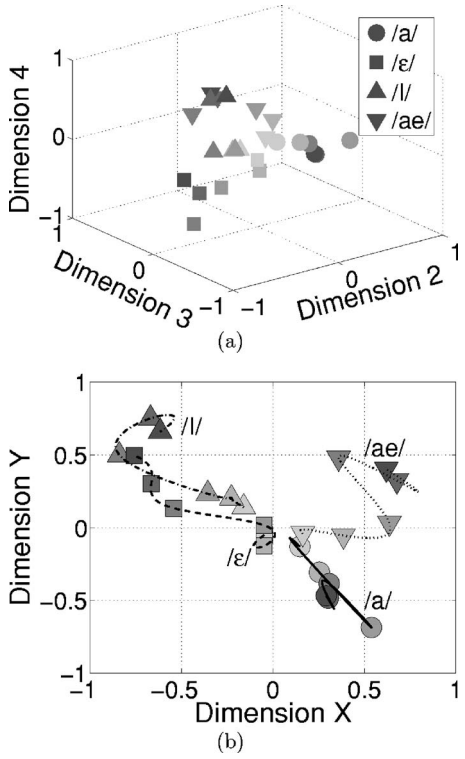


FIG. 14. (a) Vowel clustering in 3D eigenspace (dimensions 2–4) of the 4×4 vowel CM $[P(V_h|V_s)]$. The gray-scale intensity of the symbols corresponds to the six SNR levels (i.e., the lightest≡−22 dB SNR and the darkest≡Quiet). (b) Two-dimensional projection of the vowel clusters. The projection matches the clean speech clustering with the vowel distribution in the left panel of Fig. 13. The lines indicate paths traced by the vowels in the 2D plane of projection, as the SNR decreases.

three dimensions. The gray-scale intensities of the symbols show the six SNR levels, with the lightest corresponding to −22 dB SNR and the darkest corresponding to the quiet condition. The clustering of the vowels in the three-dimensional (3D) eigenspace, when projected on a specific plane in the eigenspace, is very close to the graph of vowel duration versus the second formant frequencies (Fig. 13). The procedure used for obtaining the two-dimensional (2D) projection [Fig. 14(b)] is described in Appendix B. The dimensions of the 2D projection are abstract and hence the axes are labeled *Dimension X* and *Dimension Y*. However, the dimensions *X* and *Y* are closely related to the vowel duration and the second formant frequency, respectively. The projection coefficients indicate that dimension *X* is almost identical to Dimension 2 (see Appendix B), which is associated with the largest eigenvalue of the 3D subspace. This suggests the vowel duration was the most dominant acoustic cue for the perceptual grouping of the four vowels.

The addition of masking noise reduces the perceptual distance among the vowels and draws them closer in the eigenspace. The vowel /ɛ/ is perceptually closer to /ɪ/ for an SNR as low as −16 dB. However, below −16 dB, /ɛ/ makes a large shift towards /ɑ/ and /æ/, and becomes equally close to the three vowels in the eigenspace. This is consistent with the vowel CPs for /ɛ/ (Fig. 12, bottom left panel). The vowel /ɪ/ is the most remote in the presence of noise, which is consistent with its highest scores (Fig. 5).

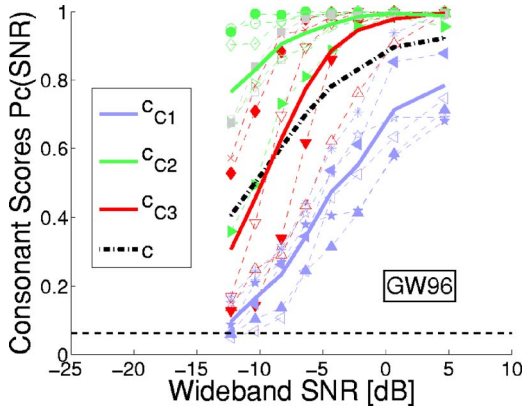


FIG. 15. (Color online) Consonant recognition scores for the 18 consonants used by Grant and Walden (1996). The hollow square and the opaque square represent the consonants /tʃ/ and /dʒ/, respectively.

G. Comparison with the past work

1. Grant and Walden (1996)

The Grant and Walden (1996) [GW96] consonant recognition scores (Fig. 15), measured in speech-weighted noise, also show consonants grouped into the same three sets. However, set C3 is not as tightly formed in GW96 as in UIUCs04. At 50% score the SNR spread of C3 consonants in UIUCs04 is about 2 dB while that in GW96 is about 7 dB. The average consonant recognition in GW96 is smaller than UIUCs04, primarily because of the low scores of C1 consonants in GW96. Note that our utterance selection (Sec. III B) removed mostly C1 consonant syllables. Without utterance selection, the C1 scores in the two experiments are closer in quiet. However, in the presence of noise, the C1 scores and hence the average consonant recognition in UIUCs04 still remain significantly greater than that in GW96. There could be several reasons for these differences. For example, the average speech spectrum in GW96, which was for a single female talker, had a greater roll-off than that in UIUCs04, which had 18 talkers. Also the noise spectrum in GW96 was a better match to the average speech spectrum than in UIUCs04. In spite of these differences, the consonant confusions in GW96 (not shown) are very similar to those in UIUCs04.

2. Miller and Nicely (1955)

Unlike MN55, the consonant groupings observed in UIUCs04 and in GW96 were not correlated with the *production* or the *articulatory* features (sometimes known as the *distinctive* features) such as voicing or nasality. In fact, a large number of voicing confusions such as /s/-/z/ and /ʃ/-/ʒ/ were observed in UIUCs04. Furthermore, the stop plosives { /p/, /t/, /k/ } and { /b/, /d/, /g/ } did not form perceptual groups in speech-weighted noise, as they do in MN55 data.

If a noise masker masks the consonants uniformly, then the consonant scores should be almost identical at a given SNR, with very little spread. The maximum spread in the UIUCs04 consonant scores is at −20 dB SNR (Fig. 5, bottom panel), with consonant /v/ (★) at 6% and consonant /z/ (●) at almost 80%. The consonant scores GW96 show even greater spread at −10 dB SNR, with a similar distribution of

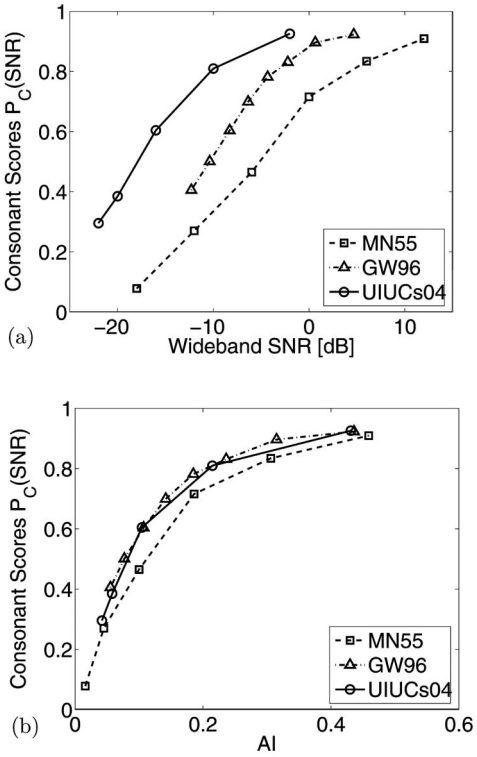


FIG. 16. A comparison of the consonant recognition scores for the current experiment [UIUCs04], Grant and Walden (1996) [GW96], and Miller and Nicely (1955) [MN55] as functions of (a) SNR and (b) AI.

individual consonant scores. In comparison, the highest spread of the consonant scores in white noise [i.e., MN55 data, shown in Fig. 6 of Allen (2005b)] is 50–90% at 0 dB SNR, which is almost half of the maximum spread observed in UIUCs04. Nasals have very high scores in MN55 data. Unlike UIUCs04 and GW96, the consonant scores in MN55 form a continuum, rather than distinct sets. Thus, white noise masks consonants more uniformly relative to the speech-weighted noise, which implies that the events for the consonant sounds are distributed uniformly over the bandwidth of speech. The events important for recognizing the C2 set consonants are at the higher frequencies that are relatively less masked by the speech-weighted noise. The events for nasals are located at low frequencies, which are masked more by speech-weighted noise than white noise.

3. AI

At a given wideband SNR, the recognition score of a consonant depends on the spectrum of the noise masker. Therefore, the wideband SNR is not a good parameter for modeling recognition scores and a parameter that accounts for the spectral distribution of speech and noise energy is required. As the AI accounts for the speech and noise spectra, it is a better measure for characterizing and comparing the scores across experiments.

Figure 16(a) shows the consonant scores $P_C(SNR)$ from UIUCs04, MN55, and GW96, as functions of wideband SNR. At 50% score, the SNR of GW96 is about 8 dB higher than UIUCs04, while the MN55 SNR is about 5 dB higher

than GW96. The foremost reason for this SNR difference is the different noise spectra in the three experiments.

However, these consonant scores overlap when replotted on an AI scale [$P_C(AI)$, Fig. 16(b)]. The AI values for GW96 were calculated using the spectra and the peak-to-rms ratios (r_k) estimated from the GW96 stimuli. The original stimuli for MN55 are not available and therefore the AI values from Allen (2005a), which were estimated using Dunn and White spectrum and $r=2$, were used. This could be one reason why the $P_C(AI)$ curve for MN55 does not match as closely as the other two curves.

IV. DISCUSSION

In this research we have explored the perception of CV sounds in speech-weighted noise. Our analysis tried to control for the many possible sources of variability. For example, we do not want the bad or incorrectly labeled utterances to be misinterpreted as the perceptual confusions in noise. There is no gold standard for a correct utterance. We used the listener responses in the quiet condition as a measure to select the good utterances. It was therefore necessary to make sure that the listeners are performing the given task accurately, in quiet. Hence the responses of four LP listeners, having significantly lower scores than the ten HP listeners, were removed before the utterance selection.

The HP listener responses showed that although 59% of the utterances had zero error in the quiet, there were few utterances which had more than 80% error and very small response entropy (not shown). A low response entropy for a high-error utterance indicates a clear case of mislabeling, i.e., the utterance was consistently perceived, but the perception of the CV was different from its label. Such utterances must be either removed or relabeled. Other high-error utterances had high response entropy, which indicates that the listeners were unsure about these utterances. The syllable error threshold for separating the good utterances from the high-error utterances should be set according to the experimental design and aims. For our purpose, we selected a conservative threshold of 20%. Also, the results were not significantly different for a 50% threshold. The listener selection and the utterance selection are interdependent. However, we verified that the HP-LP listener classification was unaffected by the utterance selection.

Another source of variability is the primary language of the listeners and the talkers. In addition to the 14 L1 = English listeners, there were 6 L1 ≠ English listeners who completed the experiment. Three of the L1 ≠ English listeners had scores worse than the LP listeners while only one L1 ≠ English had scores comparable to that of the HP listeners. Since it has been shown that the primary language affects the consonant and vowel confusions [Singh and Black (1965); Fox *et al.* (1995)], the analysis in this paper was limited to only L1 = English listeners. All the talkers from LDC2005-S22 database were native speakers of English and three of those were bilingual. The syllable errors were not different for bilingual and monolingual talkers, and therefore, all talkers were used.

Once the listeners and the utterances were selected, it was possible to reliably study the effects of noise. The speech-weighted noise masks the vowels uniformly, but has a nonuniform masking effect on the consonants, dividing them into three sets: low-scoring C1 consonants, high-scoring C2 consonants, and the remaining consonants, clubbed together as C3, and having intermediate scores. The predominant sets C1 and C2 are also observed in GW96. However, in case of MN55, no distinct consonant sets are observed and the spread in the consonant scores is much smaller in white noise (i.e., MN55) relative to the speech-weighted noise (UIUCs04 and GW96).

Analysis of the 64×64 CM (Sec. III D) shows two well-defined structures that relate to the consonant sets. The syllables with C1 consonants ($\lambda>1$) show the parallel-line structure (i.e., consonant confusion but correct vowel) while those with C2 consonants ($\lambda<1$) show the diagonal blocks (i.e., vowel confusion but correct consonant). Thus the vowel-to-consonant recognition ratio λ quantifies the qualitative analysis of CM images. It is also correlated with the vowel dependence of the consonant confusions. The CPs for consonants with $\lambda>1$ (sets C1 and C3) are more likely to be affected by the following vowel than those for consonants with $\lambda<1$ (i.e., set C2).

The consonant PSDs and therefore the SNR spectra (Fig. 9) are dominated by the vowels at low frequencies, but a clear difference can be observed at the high frequencies. The high SNR at high frequencies distinguishes C2 consonants from the other two sets. The PSDs of C3 consonants are indistinguishable from the C1 PSDs, which explains why C1 and C3 consonants are confused with each other, but not with the C2 consonants. The C2 scores are higher than C1 and C3 scores at a given SNR (Fig. 5) due to the high SNRs at high frequencies. This spectral difference is accounted by the AI, which makes the $P_C(AI)$ curves for C2 and C3 overlap on the AI scale [Fig. 7(a)]. However, the $P_C(AI)$ curves for C1 do not overlap with the C2 and C3 curves, due to higher $e_{\min}$ values. There may also be spectral differences between C1 and the remaining consonants at lower frequencies, which are dominated by the vowel energy. In such a case, the spectral differences are not detectable in the SNR spectra and therefore cannot be accounted for by the AI. Also, note that since most of the removed utterances had C1 consonants, these consonants are not only hard to perceive, but are also difficult to pronounce clearly.

Confusions within the C2 consonants are highly asymmetric and are biased in favor of the voiced consonants (Fig. 11). These asymmetric confusions are not observed in MN55. Therefore, it is possible that the speech-weighted noise, which has more energy at low frequencies, introduces a percept of voicing. Another explanation for the asymmetry is that the speech-weighted noise masks the voicing information (i.e., either presence or absence) at the low frequencies and in absence of this information, human auditory system assumes the voicing to be present, by default. Specific experiments would be required to test these hypotheses.

In several cases, there is a noticeable variation in the consonant confusions for different utterances of the same CV

(not shown). This variation is obscured after pooling the responses to all utterances of a given CV. Some utterances show interesting phenomenon that we call *consonant morphing*, i.e., when confusion of a consonant with another consonant is significantly greater than its own recognition. The confusion threshold of an utterance depends on the intensities of various features in that utterance. This natural variability in speech could be used to locate the perceptual features. For that matter, the confusable sounds with high response entropy could be a blessing in disguise. Comparing spectro-temporal properties of such sounds with that of the nonconfusable sounds will provide vital information about the perceptual features.

The SNRs used in this study were not low enough to get clear perceptual grouping of vowels, as defined by SNR_g, in spite of having close formant frequencies. The four vowels are uniformly masked by the speech-weighted noise, resulting in practically overlapping recognition scores $P_C(SNR)$. However, based on the hierarchy of the competitors in the vowel CPs, vowels formed two groups—the long, stressed vowels (/ɑ/-/æ/) and the short, unstressed vowels (/ɛ/-/ɪ/). The eigenspace clustering of the vowels is strikingly similar to that in the Duration-F_2 space, with the Duration relating to the strongest eigenspace dimension. The vowel confusions were found to be independent of the preceding consonant. However, these observations should be verified with a larger set of vowels before generalizing.

Finally, we compare the consonant scores from UIUCs04 with the Grant and Walden (1996) and Miller and Nicely (1955) scores (Fig. 16). The $P_C(SNR)$ curves for the three experiments are neither close nor parallel to each other on the wideband SNR scale, due to different noise spectra. However, the $P_C(AI)$ curves practically overlap. Thus, we have shown that, in spite of different experimental conditions, the AI can consistently characterize and predict the consonant scores, for any speech and noise spectra.

V. CONCLUSIONS

The important observations/implications from this study can be briefly summarized as follows.

1. Unlike the white noise, the speech-weighted noise nonuniformly masks the consonants, resulting in a larger spread in the consonant recognition scores. The C1 consonants (/f/, /θ/, /v/, /ð/, /b/, /m/) have the lowest scores while consonants C2 (/s/, /ʃ;/, /z/, /ʒ/, /t/) have the highest scores (Fig. 5, bottom). The remaining consonants have scores between the C1 and C2 scores and are grouped together as set C3.
2. Sets C1 and C3 are confused with each other with some degree of asymmetry, but set C2 is not confused with the other two groups (Fig. 8). This is consistent with the spectral power of the consonants above the noise spectrum (i.e., the SNR spectra, Fig. 9). The asymmetric confusions between sets C1 and C3 can be explained by the difference in their recognition scores.
3. The consonant confusion groups in speech-weighted noise are C1: { /f/-/θ/, /b/-/v/-/ð/, /θ/-/ð/ }, C2: { /s/-/z/, /ʃ/-/ʒ/ }, and C3: /m/-/n/ (Sec. III E). There is no across-set consonant group. Unlike the white-noise case (MN55), there are very high voicing confusions in the speech-weighted noise. The perceptual groups /s/-/z/ and /ʃ/-/ʒ/ are highly asymmetric, biased in favor of the voiced consonant in the presence of noise (Fig. 11).
4. The vowel-to-consonant recognition ratio λ is a quantitative measure of the confusions observed in the CM images, i.e., $\lambda > 1 \Rightarrow$ consonant confusions dominate, resulting in the parallel lines, while $\lambda < 1 \Rightarrow$ vowel confusions dominate, resulting in the diagonal blocks in CM images.
5. The confusions for set C1 ($\lambda > 1$) depend on the vowels, while those for set C2 ($\lambda < 1$) are independent of vowel. (Sec. III E.)
6. Vowels are uniformly masked by the speech-weighted noise (Fig. 5, top) and form two confusion groups, viz. /ɑ/-/æ/ and /ɛ/-/ɪ/. The eigenspace clustering of the vowels (Fig. 14) relates to the duration and the second format frequencies of the vowels (Fig. 13).
7. The recognition errors for 12 of the 16 consonants (dashed lines, Fig. 7) used in this study, as well as the average error (dash-dotted line) can be modeled with the exponential AI model [Eq. (1)] proposed by Allen (2005b). However, the model works better with a frequency-dependent peak-to-rms ratio r_k [Eq. (3)], than the frequency-independent ratio (Allen, 2005b).
8. The Articulation Index accounts for the spectral differences in the speech and noise spectra and is a better parameter than the wideband SNR for characterizing and comparing the consonant scores across experiments (Fig. 16).

ACKNOWLEDGMENTS

We thank all members of the HSR group at the Beckman Institute, UIUC for their help. We thank Andrew Lovitt for writing a user-friendly code that made data collection easier and faster. Bryce Lobdell's input was crucial in revising the manuscript. We are grateful to Kenneth Grant for sharing his confusion data, stimuli, and his expertise.

APPENDIX A

Traditionally, the peak level of speech is measured using the volume-unit (VU) meter. The peak level is given by the mean value of peak deflections on the VU meter ("dBA fast" setting) for the given speech sample (Steeneken and Houtgast, 2002). The peak deflections in the VU meter correspond to the peaks of the speech envelope, estimated in $T = 1/8$ s intervals [Lobdell and Allen (2006), French and Steinberg (1947)].

The speech signal filtered through each of the K articulation bands has the same bandwidth B_k as that of the articulation band. Therefore, for estimation of the envelope with optimum sampling, according to the Nyquist criterion, the duration of intervals is selected to be

$$T_k = \frac{1}{2B_k} = \frac{1}{2(f_{U_k} - f_{L_k})}, \tag{A1}$$

where f_{L_k} and f_{U_k} are, respectively, the lower and the upper cutoff frequencies of the kth articulation band. The value of r_k is then calculated as

$$r_k = \frac{p_k}{\sigma_k}, \tag{A2}$$

where σ_k is the rms value of the speech signal filtered through kth articulation band and p_k is the envelope peak for the same filtered speech. The value of r_k increases with the center frequency of the articulation band, ranging from 3.3 ($\approx$10.4 dB, for /n/) in the lowest articulation band (200–260 Hz) to 11.2 ($\approx$21.0 dB, for /d/) in the highest articulation band (6750–7300 Hz). For the GW96 stimuli, the r_k values ranged from 1.2 ($\approx$1.56 dB, for /ʃ/) in the 200–260 Hz band, to 8.98 ($\approx$19.1 dB, for /d/) in the 6370–6750 Hz band.

APPENDIX B

The PCA or the eigenvalue decomposition of the 4×4 vowel CM $P(V_h|V_s)$ can be represented in matrix form as $P(V_h|V_s)=EDE^{-1}$, where

$$D = \begin{bmatrix} D_1 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 \\ 0 & 0 & D_3 & 0 \\ 0 & 0 & 0 & D_4 \end{bmatrix}$$

is the rank-ordered eigenvalue (singularity) matrix, with D_1 being the largest eigenvalue and D_4 being the smallest eigenvalue, and

$$E = [\mathbf{E}_1\ \mathbf{E}_2\ \mathbf{E}_3\ \mathbf{E}_4] = \begin{bmatrix} e_{11} & e_{21} & e_{31} & e_{41} \\ e_{12} & e_{22} & e_{32} & e_{42} \\ e_{13} & e_{23} & e_{33} & e_{43} \\ e_{14} & e_{24} & e_{34} & e_{44} \end{bmatrix}$$

is the eigenvector matrix. Each eigenvector represents a dimension in the eigenspace. Because $P(V_h|V_s)$ is row normalized, D_1 is unity and the coordinates along the first dimension are identical (e_{1i}=0.5). Therefore, the vowel clustering in the eigenspace is plotted only along the dimensions 2–4 [Fig. 14(a)]. The 4×3 coordinate matrix for the four vowels along the three dimensions is

$$C = (E\sqrt{D})_{2-4} = \begin{bmatrix} \sqrt{D_2}e_{21} & \sqrt{D_3}e_{31} & \sqrt{D_4}e_{41} \\ \sqrt{D_2}e_{22} & \sqrt{D_3}e_{32} & \sqrt{D_4}e_{42} \\ \sqrt{D_2}e_{23} & \sqrt{D_3}e_{33} & \sqrt{D_4}e_{43} \\ \sqrt{D_2}e_{24} & \sqrt{D_3}e_{34} & \sqrt{D_4}e_{44} \end{bmatrix}.$$

Let

$$F = \begin{bmatrix} f_1 & d_1 \\ f_2 & d_2 \\ f_3 & d_3 \\ f_4 & d_4 \end{bmatrix}$$

be the feature matrix of the four vowels, which contains the values of the second format frequencies F_2 (f_i) and the vowel durations (d_i). The feature matrix F is normalized to have zero mean and unit variance along both dimensions f and d. The 2D projection of C that matches the F can be obtained by the linear transform

$$F = CA, \tag{B1}$$

where A is a 3×2 matrix that rotates C about the origin and orthogonally projects it on a 2D plane. The closed form solution for the minimum mean square estimate for A is

$$\hat{A} = [C^T C]^{-1} C^T F. \tag{B2}$$

The 2D projection in Fig. 14(b) was obtained by matching the eigenvectors for quiet condition to the normalized version of the 2D clustering in Fig. 13, left panel. The feature matrix F was obtained using the average of the second format frequencies (f) and the vowel durations (d) for male talkers. The matrix was then normalized by subtracting the means and dividing by the standard deviations along the f and d dimensions. The projection matrix in this case was

$$\hat{A} = \begin{bmatrix} 0.9932 & -0.7860 \\ 0.5512 & 0.4203 \\ 0.4363 & -0.0261 \end{bmatrix}.$$

The value of coefficient a_{11}, which is the projection of Dimension X on Dimension 2, is almost unity. This suggests that Dimension X is almost the same as Dimension 2, since the angle between the two dimensions is very close to zero $[\cos^{-1}(0.9932)=6.68°]$.

[1]The "listener scores" are the CV syllable recognition scores, i.e., the scores of recognizing both consonant and vowel correctly. The average consonant scores (c) are equal to the average vowel scores (v) in quiet condition (see Sec. III C). Therefore, a threshold of 85% for CV recognition corresponds to a threshold of 92.2% for phone recognition.

[2]The relative levels of speech and noise spectra are set according to the wide-band SNR, which is calculated from the rms levels. The contribution of the articulation bands to the speech intelligibility is proportional to the peaks in the articulation band-filtered speech signal, that are above the noise floor, and therefore a correction for the peak-to-rms ratio of speech is necessary (French and Steinberg, 1947). French and Steinberg (1947) suggested a correction of 12 dB, for all articulation bands, which is consistent with the measured peak-to-rms ratios for speech (Steeneken and Houtgast, 2002), and approximately corresponds to r=4.

Allen, J. B. (**2005a**). *Articulation and Intelligibility*, Synthesis Lectures in Speech and Audio Processing, series editor B. H. Juang (Morgan and Claypool).

Allen, J. B. (**2005b**). "Consonant recognition and the articulation index," J. Acoust. Soc. Am. **117**, 2212–2223.

Benson, R. W., and Hirsh, I. J. (**1953**). "Some variables in audio spectrometry," J. Acoust. Soc. Am. **25**, 499–505.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An internation comparison of long-

term average speech spectra," J. Acoust. Soc. Am. **96**, 2106–2120.
Campbell, G. A. (**1910**). "Telephonic intelligibility," Philos. Mag. **19**, 152–159.
Cox, R. M., and Moore, J. N. (**1988**). "Composite speech spectrum for hearing aid gain prescriptions," J. Speech Hear. Res. **31**, 102–107.
Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.
Dunn, H. K., and White, S. D. (**1940**). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. **11**, 278–287.
Fletcher, H. (**1995**). *The ASA Edition of Speech and Hearing in Comunication*, edited by Jont B. Allen (Acoustical Society of America, New York).
Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.
Fousek, P., Svojanovsky, P., Grezl, F., and Hermansky, H. (**2004**). "New nonsense syllables database—analyses and preliminary ASR experiments," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, October 4–8, Jeju, South Korea, http://www.isca-speech.org/archive/interspeech_2004. Viewed 3/26/07.
Fox, R. A., Flege, J. E., and Munro, M. J. (**1995**). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," J. Acoust. Soc. Am. **97**, 2540–2551.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Gordon-Salant, S. (**1985**). "Some perceptual properties of consonants in multitalker babble," Percept. Psychophys. **38**, 81–90.
Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415–2424, URL http://www.wramc.amedd.army.mil/departments/aasc/avlab/datasets.htm. Viewed 3/26/06.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Lippman, R. P. (**1997**). "Speech recognition by machines and humans," Speech Commun. **22**, 1–15.
Lobdell, B., and Allen, J. B. (**2007**). "Modeling and using the vu-meter (volume unit meter) with comparisons to root-mean-square speech levels," J. Acoust. Soc. Am. **121**, 279–285.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.
Singh, S., and Black, J. W. (**1965**). "Study of twenty-six intervocalic consonants as spoken and recognized by four language groups," J. Acoust. Soc. Am. **39**, 372–387.
Sroka, J., and Braida, L. D. (**2005**). "Human and machine consonant recognition," Speech Commun. **45**, 401–423.
Steeneken, H. J. M., and Houtgast, T. (**2002**). "Basics of STI measuring methods," *Past, Present and Future of the Speech Transmission Index*, edited by S. J. van Wijngaarden (TNO Human Factors, Soesterberg, The Netherlands).
Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **60**, 213–224.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.

Speaker-independent factors affecting the perception of foreign accent in a second language[a)]

Susannah V. Levi,[b)] Stephen J. Winters,[c)] and David B. Pisoni
Speech Research Laboratory, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405

(Received 7 February 2006; revised 9 January 2007; accepted 12 January 2007)

Previous research on foreign accent perception has largely focused on speaker-dependent factors such as age of learning and length of residence. Factors that are independent of a speaker's language learning history have also been shown to affect perception of second language speech. The present study examined the effects of two such factors—listening context and lexical frequency—on the perception of foreign-accented speech. Listeners rated foreign accent in two listening contexts: auditory-only, where listeners only heard the target stimuli, and auditory+orthography, where listeners were presented with both an auditory signal and an orthographic display of the target word. Results revealed that higher frequency words were consistently rated as less accented than lower frequency words. The effect of the listening context emerged in two interactions: the auditory +orthography context reduced the effects of lexical frequency, but increased the perceived differences between native and non-native speakers. Acoustic measurements revealed some production differences for words of different levels of lexical frequency, though these differences could not account for all of the observed interactions from the perceptual experiment. These results suggest that factors independent of the speakers' actual speech articulations can influence the perception of degree of foreign accent. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2537345]

PACS number(s): 43.71.Bp, 43.71.Hw [ARB] Pages: 2327–2338

I. INTRODUCTION

The ability to speak a second language fluently depends in large part on how well a speaker has been able to acquire the second language (L2) phonology and to accurately realize the intended phonetic targets. The perceived degree of foreign accent of a speaker, however, is not based exclusively on the amount of acoustic and articulatory mismatches between non-native and native productions. Degree of foreign accent also reflects a *listener's* perception of the L2 speech. Many of the factors known to affect the perception of foreign-accented speech are speaker-specific factors that are inherent to a particular speaker. We will refer to these factors as *speaker dependent* since they are dependent upon a particular speaker's language learning history and cannot be directly changed or manipulated by an experimenter. Speaker-dependent factors have received considerable attention in the L2 literature. They include age of learning (the age at which a speaker begins learning a second language), length of residence in an L2 environment, the first language of the speaker, and his/her motivation to attain unaccented or less-accented speech [see Piske *et al.* (2001) for a review].

Additional factors that are not inherent to a particular speaker and are not part of the speaker's language learning history can also affect the perception of degree of foreign accent. These factors can be manipulated or controlled by the researcher and often reflect the specific methodology involved in obtaining measures of degree of foreign accent. We will refer to these as *speaker-independent* factors. For example, Southwood and Flege (1999) suggest that different rating scales may affect participants' judgments of perceived degree of foreign accent. They point out that scales with fewer intervals may produce ceiling effects and therefore not be sensitive enough to differentiate L2 speakers.

Different types of elicitation techniques can also affect the degree of perceived foreign accent. Studies investigating the perception of foreign accent have used a variety of techniques to produce their stimulus materials; these techniques vary in whether the L2 speakers spontaneously generate speech, read printed text (words, sentences, or paragraphs), or repeat samples of speech after hearing the intended target produced by a native speaker. Oyama (1976) and Thompson (1991) have found that read speech is judged as more accented than spontaneous speech.

Studies also differ in whether native speaker controls are included in the rating task. Native controls serve to confirm that listeners are correctly performing the task by testing that they can distinguish native from non-native speech. Using native controls also ensures that listeners use a wider range of the rating scale (Flege and Fletcher, 1992). The proportion of native speakers included in a rating set also affects the perception of foreign accent. Flege and Fletcher (1992) found that increasing the proportion of native speakers in the stimulus set caused non-native speakers to be rated as more accented. Characteristics of the listener can affect the per-

[a)]Portions of this work were presented at the 1st ASA workshop on Second Language Speech Learning held on 14–15 May 2005 in Vancouver, B.C., Canada.
[b)]Electronic mail: svlevi@indiana.edu
[c)]Now at the Department of Linguistics, University of Illinois, Urbana, Illinois 61801.

ceived degree of foreign accent, as well. Studies have varied whether naive listeners (e.g., Flege and Fletcher, 1992; Flege *et al.*, 1995) or experienced listeners such as linguists (e.g., Fathman, 1975) and ESL teachers (e.g., Piper and Cansin, 1988) serve as raters. Thompson (1991) found that naive listeners tended to perceive a greater degree of foreign accent than experienced listeners, although Bongaerts *et al.* (1997) did not find a significant difference. Flege and Fletcher (1992) also found that if listeners are familiarized with the target sentences, then non-native speakers are rated as more accented. An additional speaker-independent factor can be speaking rate, when the change in rate is caused by experimental manipulation. Munro and Derwing (2001) used speech compression-expansion software to increase and decrease speaking rate by 10%. By using this software they were able to ensure that other properties of the speech (e.g., number of segmental substitutions) remained unchanged. They found that fast stimuli were rated as less accented than stimuli presented at normal and slowed rates. Taken together, these studies show that speaker-independent factors, in addition to speaker-dependent factors, can also affect the perceived degree of foreign accent.

The current study investigated the effects of two additional speaker-independent factors—lexical frequency and listening context—on the perception of degree of foreign accent using an accent rating task. These two factors were chosen because they have been shown to affect speech perception and language processing of native speech. This study extends these two factors to the perception of foreign-accented speech.

Lexical frequency has been found to play an integral role in language processing and may therefore be expected to affect the perception of degree of foreign accent. Lexical frequency affects spoken word recognition (Howes, 1957; Savin, 1963; Luce and Pisoni, 1998), the recognition of words in a gating paradigm (Grosjean, 1980), and word shadowing (Goldinger, 1997). In a word identification task, Howes (1957) mixed words of varying frequency with multiple signal-to-noise ratios. High frequency words exhibited greater intelligibility by being perceived at less favorable signal-to-noise ratios than were low frequency words. In a similar study, Savin (1963) examined listeners' response errors and found that incorrect responses tended to be words of higher frequency than the target word. In a lexical decision task, Luce and Pisoni (1998) found that listeners responded more quickly and more accurately to high frequency words than to low frequency words.

Goldinger (1997) showed that listeners rely more heavily on the acoustic-phonetic information in the speech signal when they perceive low frequency words than when they perceive high frequency words. Using a word shadowing task, Goldinger presented listeners with both high and low frequency words from a variety of speakers and asked them to repeat the words as quickly as possible. Goldinger predicted that subjects would change their productions to match the different speakers using "spontaneous vocal imitation." The amount of vocal imitation was quantified by comparing how well the response utterances matched the stimulus in fundamental frequency and duration. Goldinger found that low frequency words resulted in higher rates of spontaneous imitation than high frequency words, suggesting that subjects were more sensitive to surface acoustic-phonetic details in low frequency words than in high frequency words.

Goldinger explained these findings within the framework of Hintzman's (1986, 1988) MINERVA2 model, an exemplar-based model of memory (see also Johnson, 1997; Kirchner, 1999, 2004; Pierrehumbert, 2001, 2002). The MINERVA2 model, like other exemplar models, assumes that every exposure to a stimulus creates a memory trace that includes all perceptual details. When a new token (the probe) is heard, it activates an aggregate of all traces in memory, called the *echo*. This *echo* forms the listener's percept. The intensity of the echo depends upon both the similarity of the traces to the probe and the number of these traces. Thus, for speech and language processing, high frequency words induce "generic" echoes because they have many existing traces in memory and are therefore less influenced by any particular probe that enters the perceptual system. Low frequency words, on the other hand, have many fewer existing traces in memory. Any incoming probe will therefore have a greater influence on the subsequent percept. For the low frequency words in Goldinger's word shadowing task, speakers based their repetitions more heavily on the incoming instance-specific information than on traces in memory for low frequency words. Their subsequent productions of low frequency words were therefore affected more by specific properties of the stimulus than were high frequency words.

Working within the framework of exemplar models of speech perception, we hypothesized that the degree to which a speaker is perceived to have a foreign accent will be directly related to the amount of acoustic-phonetic mismatch between the signal and its resulting echo. In a nativeness rating task, we expected listeners' perception of L2 speech to rely more heavily on the acoustic-phonetic features of an incoming speech token for low frequency words. Listeners have fewer exemplars of low frequency words in memory and will thus generate less generic echoes in response to productions of those words. Potential acoustic-phonetic mismatches between productions of those words and their corresponding exemplars in memory should therefore be larger for low frequency words, which should in turn be rated as more accented than high frequency words. An alternative hypothesis is that the lexical frequency of the target word will have no effect on the perception of foreign accent because accent rating does not require accessing the lexicon and therefore can be based solely on the phonetic and phonological properties of the stimulus.

The second speaker-independent factor investigated in this study was the listening context. Spoken words were either presented to participants in the auditory modality alone (auditory-only) or with the addition of a simultaneous orthographic display (auditory+orthography). Previous work has shown that knowledge of the intended target, as in the auditory+orthography context, facilitates the perception of degraded speech stimuli (Davis *et al.*, 2005). Since non-native speech can be considered a type of degraded stimuli, the same type of facilitation should be found. Davis *et al.* use

the term "pop-out" to refer to a phenomenon where a degraded speech stimulus immediately becomes comprehensible after it is played to listeners in its original, undegraded form. Davis *et al.* tested the effects of pop-out on a type of noise-vocoded speech that simulates the signal heard by cochlear implant users.[1] In one experiment, they found that listeners were able to correctly report more words from a noise-vocoded target sentence after hearing the sentence in the clear. In another experiment, they found that listeners showed the same advantage, or pop-out effect, after seeing the written version of a noise-vocoded sentence presented on a computer screen. This combination of effects demonstrates that top-down processing can influence the perception of severely degraded, noise-vocoded speech regardless of the modality in which the original sentences are presented. As Davis *et al.* concluded, "pop-out must be at a non-acoustic, phonological level or higher" (p. 230).

The effects of this type of pop-out on the perception of degree of foreign accent are unclear, however. One possibility is that simultaneously presenting the auditory and orthographic representations of the target word together will cause non-native speech samples to be rated as less accented. If a non-native production of a target word is ambiguous or difficult to understand, presenting the target word in orthography on the screen may promote a type of pop-out effect to occur where the "degraded," non-native production immediately becomes more intelligible. Once the listener knows the intended utterance, possible ambiguities or confusions about which lexical item the listener should retrieve are lost. In this case, the perception of a high degree of foreign accent may also be significantly attenuated.

A second possibility is that simultaneously presenting auditory and orthographic representations of the target word will cause non-native speech samples to be rated as more accented. This outcome might occur because knowledge of the target word may serve as a perceptual benchmark and therefore highlight the amount of mismatch between the target and its corresponding exemplars in memory. An actual example from our data serves to illustrate this point. Several of the L2 speakers in the current study consistently produced word final target /s/ as [z]. For these speakers, the target word "noose" [nus] was produced as [nuz] (identical to "news," which was not one of the target words). Hearing [nuz] while seeing "noose" could focus listeners' attention on mismatches between the expected and observed productions. It might be expected that listeners would rate these speakers as having more of a foreign accent when they hear the word [nuz] in conjunction with seeing "noose" on the screen than when they simply hear [nuz] alone and could freely conclude that they had heard an accurate production of "news."

To summarize, the current study examined the effects of lexical frequency and listening context on the perceived degree of foreign accent of native and non-native speakers of English. We predicted that higher frequency words would be rated as less accented than lower frequency words. In terms of the listening context, two competing hypotheses were assessed. The addition of orthographic displays may induce pop-out effects, making the stimuli more intelligible, resulting in their being rated as less accented. Alternatively, the presentation of the target word may cause listeners to focus their attention on mismatches between the target utterance and the actual stimuli, resulting in the stimuli being rated as more accented. Positive results with these two factors will provide additional evidence for the contribution of speaker-independent factors in the perception of foreign accent.

II. EXPERIMENT: PERCEPTION OF FOREIGN ACCENT

A. Methodology

1. Materials

All speakers were recorded in a sound-attenuated IAC booth in the Speech Research Laboratory at Indiana University. Speech samples were recorded using a SHURE SM98 headmounted unidirectional (cardioid) condenser microphone with a flat frequency response from 40 to 20 000 Hz. Utterances were digitized into 16-bit stereo recordings via Tucker-Davis Technologies System II hardware at 22 050 Hz and saved directly to a PC. A single repetition of 360 English words was produced by each speaker. Each word was of the form consonant-vowel-consonant (CVC) and was selected from the CELEX database (Baayen *et al.*, 1995). Speakers read each word in random order as it was presented to them on a computer monitor in the recording booth. Before each presentation, an asterisk appeared on the screen for 500 ms, signaling to the speaker that the next trial was about to begin. This was followed by a blank screen for 500 ms. After this delay, a recording period began that lasted for 2000 ms. The target word was presented on the screen for the first 1500 ms of this recording period. After the conclusion of the recording period, the screen went blank for 1500 ms, and then another asterisk appeared to signal the beginning of the next recording cycle. Items that were produced incorrectly or too loudly were noted and re-recorded in the same manner following the recording session. The total recording time was approximately one hour.

This process yielded recordings that were uniformly 2000 ms long. Since the actual productions of the stimulus words were always shorter than 2000 ms, the silent portions in the recording before and after each production were manually removed using Praat sound editing software (Boersma and Weenink, 2004). All edited tokens were then normalized to have a uniform RMS amplitude of 66.4 dB.

Words were selected to represent a range of lexical frequencies based on counts from the CELEX database. For the purposes of analysis, words were divided into three equal groups of varying frequency. The 120 lowest frequency words all had a CELEX frequency count of less than or equal to 96, while the 120 highest frequency words all had a frequency of greater than or equal to 586. The remaining 120 words thus all had frequency counts between 96 and 586. The frequency count of homophones (e.g., rite, write, right) was taken to be the frequency count of the most frequent homophone; this homophone was also the word that was presented orthographically to the speakers during the recording sessions.

TABLE I. Demographic variables for the bilingual speakers. "Years of English" refers to the number of years speakers have been learning/using English (current age-age of acquisition). "Fluency" is a self-reported measure of English proficiency (1=poor, 5=fluent). The final column provides each speaker's mean *z*-score accent rating. Larger *z*-scores reflect a higher degree of foreign accent.

Speaker	Age of acquisition	Years of English	Length of residence	Fluency	Accent rating
f2	12	9	1	4	0.00
f3	10	14	1	5	0.22
f4	13	13	3	4.5	0.02
f7	9	12	1	5	0.33
f8	13	16	4	5	−0.27
f9	9	16	2	4	0.13
f11	⋯	⋯	2	4.5	0.94
m2	12	18	3	5	0.02
m3	10	13	1	4	0.31
m4	11	18	1	4	0.39
m6	13	13	1	5	0.69
m9	12	20	2	4	0.95
m10	10	14	1	5	0.57
Mean	11.2	14.7	1.77	4.54	0.33
SD	1.53	3.06	1.01	0.48	0.37

2. Speakers

Twelve female and ten male German L1/English L2 speakers were recorded. Of the 22 speakers, nine speakers were eliminated due to dialect differences (Austrian German: $N=3$, Southern German: $N=2$, Romanian-German: $N=1$), reported speech or hearing disorders ($N=2$), or only completing part of the recordings ($N=1$). Recordings from the remaining seven female and six male speakers were used in this study. All speakers were paid $10/h for their time. Demographic variables for the remaining bilingual speakers are given in Table I.

Thirteen native speakers (six male, seven female) of American English were also recorded producing only the list of English words under the same conditions as the bilingual speakers. These speakers were from various dialect areas of American English (Midland: $N=7$, West: $N=1$, South: $N=1$, North: $N=1$, more than one dialect area: $N=3$). [See Labov *et al.* (2006) for descriptions of these dialect labels.] Productions from two of the female speakers were not included in the study due to problems these speakers had with completing the task accurately. Productions from the remaining six male and five female native speakers were included in the study. All of these speakers received partial course credit for their participation.

3. Listeners

A total of 87 listeners between the ages of 18 and 25 participated in this experiment; 42 were assigned to the auditory-only context and 45 to the auditory+orthography context. Twenty-seven listeners were eliminated (polylingual/non-native speakers of English: $N=6$, L2 German: $N=8$, machine malfunction: $N=9$, non-American English dialect: $N=1$, speech/hearing disorder: $N=2$, not completing: $N=1$), resulting in 30 native English listeners in each listening context. None of the remaining listeners had studied German, and only 6 reported having German acquaintances (friend: $N=3$, teaching assistant: $N=2$, professor: $N=1$). All remaining listeners reported no history of a speech or hearing disorder at the time of testing. Each listener participated in only one of the two listening contexts. All listeners were students enrolled in introductory psychology courses at Indiana University and received partial course credit for their participation.

4. Procedure

The experiment was implemented on Macintosh G3 computers running a customized SuperCard (version 4.1.1) stack. Listeners were seated in front of these computers in a quiet testing room while wearing Beyerdynamic DT-100 headphones. Stimuli were presented at a comfortable listening level (approximately 65 dB SPL) to all subjects. The SuperCard stack played productions of individual words to listeners and then presented them with the on-screen question, "How much of a foreign accent did that speaker have?" Participants answered this question by clicking the appropriate button in a seven-point rating scale ranging from 0 (="no foreign accent—native speaker of English") to 6 ("most foreign accent") presented on-screen. All listeners were informed that some of the speakers they would hear were native speakers of English and some were non-native speakers. All listener ratings were converted to normalized *z*-scores per listener prior to statistical analysis.

The auditory tokens of each word were presented to listeners in one of two different ways. Listeners in the auditory-only context heard each word prior to making a judgment of how accented the spoken stimulus was. Listeners in the auditory+orthography context, however, saw the orthographic representation of each word on the computer screen for 500 ms before hearing an auditory production of that word. The orthographic representation of the word remained on screen until the conclusion of the auditory stimulus, after which the listener rated its accentedness.

The experiment was divided into two blocks. In each block, 12 words were randomly selected for presentation from each of the 11 monolingual and 13 bilingual speakers, yielding a total of 288 tokens per block. Listeners thus heard a total of 576 words over the duration of the entire experiment. Each block of words was rated by two different listeners. The experiment was self-paced and took approximately one hour for most listeners to complete.

Participants had the option of listening to stimuli again after the initial presentation. In the auditory+orthography context, participants listened to 89.5% of the tokens only once and to 10.5% more than once prior to making their responses. Participants in the auditory-only context listened to 78.5% of the tokens once and to 21.5% two or more times. An ANOVA of the percentage of trials that listeners elected to hear more than once was conducted with listening context (auditory-only versus auditory+orthography) as a between-subjects factor and with native language of the speaker (L1 English versus L2 English) as a within-subjects factor. The ANOVA revealed main effects of both listening context [$F(1,58)=5.688$, $p=0.020$] and native language [$F(1,58)=5.617$, $p=0.021$], but no interaction. Listeners replayed

stimuli more often in the auditory-only context than in the auditory+orthography context (means: 21.5% versus 10.5%, respectively). Furthermore, listeners replayed words from native speakers of English more often than from non-native speakers of English (16.6% versus 15.3%).

To assess the overall consistency and reliability of raters, the intraclass correlation coefficient was calculated for raters in each of the two listening contexts (McGraw and Wong, 1996). The intraclass correlation coefficient was high for both groups (auditory-only, 0.734; auditory+orthography, 0.798), indicating a strong degree of agreement among the raters. The reliability for all judges averaged together, sometimes referred to as the interrater reliability coefficient (MacLennan, 1993) was 0.988 for the auditory-only listeners and 0.991 for the auditory+orthography listeners.

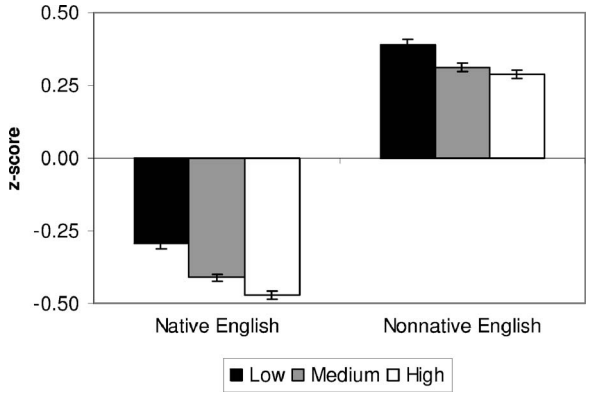


FIG. 1. Mean z-score ratings with standard errors for native and non-native speakers for each of the three levels of lexical frequency. Positive values indicate more foreign accent while negative values indicate less foreign accent.

B. Results

Before analysis, responses were pooled across speakers and across stimulus items, yielding an average measure of degree of foreign accent for each of the three levels of lexical frequency and the two levels of speakers (native versus non-native). An ANOVA with lexical frequency (low, medium, high) and native language of the speaker (native, non-native) as within-subjects variables and listening context (auditory-only, auditory+orthography) as a between-subjects variable was conducted on the z-scores of the nativeness ratings for all listeners. In the presentation of the results, larger z-score ratings indicate a greater degree of foreign accent.

The ANOVA revealed a significant main effect of lexical frequency $[F(2,116)=44.8,\ p<0.001]$. Paired-samples t tests revealed significant pairwise differences between all three levels of lexical frequency (all $p\leqslant 0.002$). The direction of this effect indicated that lower frequency words were rated as more accented than words of higher frequency. A main effect of native language of the speaker was also found $[F(1,58)=1214.8,\ p<0.001]$ where native speakers were rated as having less foreign accent overall than non-native speakers. The main effect for listening context was not significant.

The analysis also revealed three significant interactions. The interaction between lexical frequency and native language of the speaker $[F(2,116)=6.51,\ p=0.002]$ is shown in Fig. 1. Paired-samples t tests revealed that ratings for the medium and high frequency words differed between the two groups of speakers. For native speakers of English, low frequency words were rated as having more of a foreign accent than medium frequency words, which were in turn rated as more accented than high frequency words (all $p\leqslant 0.001$). For non-native speakers, low frequency words were rated as more accented than both medium and high frequency words ($p<0.001$). No significant difference between the medium and high frequency words ($p=0.213$) was found for the non-native speakers, although the trend was in the same direction as native English speakers with medium frequency words rated as more accented than high frequency words.

To further investigate this interaction, we examined the differences between the perception of native and non-native words at each level of lexical frequency. The mean difference between native and non-native productions for low frequency words was 0.679, for medium frequency words 0.722, and for high frequency 0.761. These means indicated that the difference between natives and non-natives was enhanced for high frequency words (i.e., a greater difference) relative to lower frequency words. Paired samples t tests revealed that the difference between these two groups of speakers was indeed increased in high frequency words when compared to low frequency words ($p=0.003$) but not when compared to medium frequency words. The difference between native and non-native speakers was smaller for low frequency words than medium frequency words ($p=0.021$). The interaction between lexical frequency and native language of the speaker, then, is primarily due to the increase in differences between native and non-native speakers for words of high frequency and a decrease in differences for low frequency words.

Figure 2 shows the interaction between lexical frequency and listening context $[F(2,116)=13.81,\ p<0.001]$. Visual inspection of Fig. 2 reveals a decrease of the overall differences between the three levels of lexical frequency in the auditory+orthography context. A one-way ANOVA of difference scores (|high frequency-low frequency|) revealed a significant difference between the two listening contexts $[F(1,59)=25.95,\ p<0.001]$, where differences in perceptual

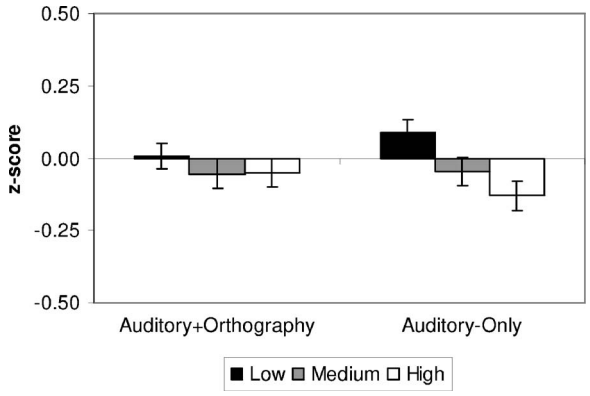


FIG. 2. Mean z-score ratings with standard errors for auditory+orthography and auditory-only contexts for each of the three levels of lexical frequency.

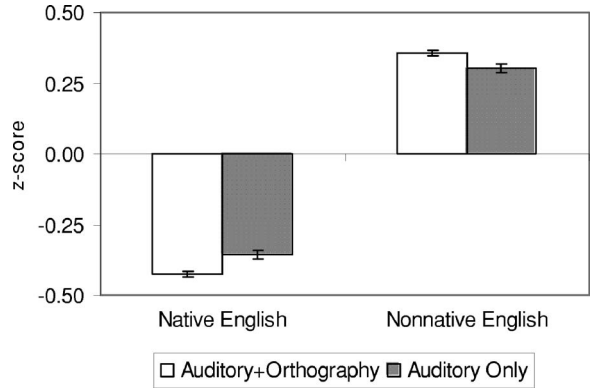


FIG. 3. Mean z-score ratings with standard errors for native and non-native speakers for each of the two listening contexts.

ratings between high and low frequency words were smaller in the auditory+orthography context (mean difference =0.094) than in the auditory-only context (mean difference =0.235). Thus, the interaction between lexical frequency and listening context reflects an attenuation of frequency effects in the auditory+orthography context.

The cross-over interaction between native language of the speaker and listening context [$F(1,58)=8.76$, $p=0.004$] is presented in Fig. 3. *Posthoc t* tests of this interaction revealed significant differences between the two listening contexts for both speaker groups. Native speakers were rated as less accented in the auditory+orthography context than in the auditory-only context ($p=0.004$), whereas non-native speakers were rated as *more* accented in the auditory +orthography context ($p=0.004$).

C. Discussion

The results of this perceptual ratings study demonstrate that two speaker-independent factors—lexical frequency and listening context—affect the perception of foreign accent in spoken words. First, high frequency words were rated as less accented than low frequency words. This result is consistent with predictions based on exemplar models of speech perception. The more frequently a word occurs in the language, the more often a listener will hear it being spoken, which will in turn lead to encoding more exemplars of the word in memory. Thus, high frequency words are more likely to match (or approximate) an exemplar of a high frequency word in memory and therefore sound comparatively less accented to a native listener of English. Low frequency words, on the other hand, will be experienced less, have many fewer exemplars in memory, and therefore sound more accented.

The effect of lexical frequency also interacted with the native language of the speaker. One analysis of this interaction revealed that differences between native and non-native speakers are increased for high frequency words. In other words, perceptual differences between the two groups of speakers were increased for words of higher frequency. This has implications for future studies of foreign accent perception. If the goal is to enhance differences between native and non-native speakers, words of higher frequency should be used. If, on the other hand, differences between speakers are to be minimized, words of lower frequency should be used. Acoustic measures in the following section indicate that these differences are not due to production differences, at least as measured by vowel format frequencies. As will be shown, production differences between native and non-native speakers do not vary based on lexical frequency. Thus, the differences found in the ratings are due to listener effects.

The other analysis of the native speaker by lexical frequency interaction revealed that lexical frequency had a stepwise effect on accent ratings for natively produced tokens: high frequency words were rated as less accented than medium frequency words, which were in turn rated as less accented than low frequency words. The effect of frequency was slightly attenuated, however, for the non-native speech where medium and high frequency words did not significantly differ from one another.

The attenuation of the lexical frequency effect for the non-native tokens may have been caused by the relationship between incoming acoustic stimuli and their stored exemplars. If the degree of perceived foreign accent is dependent upon the number of exemplars in memory that are acoustically similar to the input signals, then stimulus tokens that are acoustically similar to many exemplars in memory will be rated as less accented than those that are acoustically similar to only a few exemplars, as was observed for native tokens. Non-native tokens, however, are likely to have fewer acoustically similar stored exemplars than native tokens, especially for naive listeners who have little if any experience with non-native speech. Because the non-native tokens lie in sparsely populated areas of the exemplar space, differences between high and medium frequency words may be eliminated. Indeed, acoustic analyses presented in the next section show that high frequency non-native words are not better approximations of their native targets than are medium or low frequency words, implying that all non-native tokens, regardless of lexical frequency, are equally deviant from their native targets.

The lack of the expected frequency effects for non-natives may, however, be due to the way the three levels of frequency were created. No *a priori* notion of high, medium, or low frequency was assumed. Instead, the 360 lexical items were simply ranked by lexical frequency and then divided into three equal groups. Since the difference between high and medium frequency was arbitrary and continuous between the levels, differences between the two highest levels of frequency may have been too small.

The production of non-native speech may also be a cause of differences in the perception of native and non-native speech. It is possible that productions of words at these three levels of frequency are different for native speakers, but that productions for non-native speakers are not. Previous work has shown that lexical factors can affect speech production. In a study that explicitly manipulated a combination of lexical frequency and neighborhood density, Wright (2003) found that speakers differed in the degree of vowel reduction/centralization as a result of changes in these lexical factors. In particular, he found that vowels in lexically "easy" words (i.e., high frequency words from sparse lexical

neighborhoods) exhibited greater centralization than lexically "hard" words (i.e., low frequency words from dense lexical neighborhoods).

Though differences in the perception of foreign accent differed slightly for native and non-native speakers, globally the results were the same; higher frequency words were rated as less accented than low frequency words. An important point emerges from this observation; lexical effects are found in a task that does not, on the surface, require accessing lexical information. Rating the degree of foreign accent could be done by simply accessing knowledge of the native language phonology and does not inherently require interacting with lexical information. The presence of lexical frequency effects demonstrates that listeners do not bypass the lexicon, but instead use lexical information when making a judgment of foreign accent. It is also important to point out that these effects were found for highly proficient non-native speakers and may not be observed for less fluent bilinguals.[2]

Although the main effect of listening context did not reach significance, it did have an effect on perceived degree of foreign accent through interactions with both lexical frequency and native language of the speaker. The interaction between listening context and lexical frequency demonstrated that presenting a visual display of the target word on the screen attenuated the effect of lexical frequency. The range of ratings in the auditory+orthography context (between the two extreme values of lexical frequency) was greatly reduced when compared to the auditory-only context. The difference between the two listening contexts with respect to frequency is most likely the result of different processing requirements in the two contexts. In the auditory-only context, listeners must perform both a word recognition task and a nativeness rating task after hearing a stimulus. Listeners must evaluate the stimulus, compare it with stored exemplars in memory, determine the identity of the word, and then evaluate its degree of foreign accent. In the auditory+orthography context, the processes of auditory word recognition and lexical access can be bypassed because the correct word is displayed visually on the computer screen.

The attenuation of frequency effects on the perception of foreign accent in the auditory+orthography context is consistent with numerous studies showing that effects of lexical frequency that are observed in open-set word recognition tasks disappear in analogous closed-set tasks (Pollack *et al.*, 1959; Sommers *et al.*, 1997; Clopper *et al.*, 2006b). Since the auditory+orthography listening context eliminates the need for auditory word recognition, the perceived accentedness of a target word in this listening context could be based solely on acoustic-phonetic or phonological differences between the incoming stimulus and existing exemplars, and not on knowledge of the lexical properties of the items. However, some lexical effects, though attenuated, still persist in this context, suggesting that properties of the lexicon influence speech perception even in tasks that do not require direct contact with the lexicon.

Presentation context also influenced the degree of perceived accentedness through an interaction with the native language of the speaker. Native speakers were rated as less accented in the auditory+orthography context than in the auditory-only context. The pattern of results was reversed, however, for non-native speakers who were rated as more accented in the auditory+orthography context than in the auditory-only context. This crossover interaction may reflect differences in the relevant task demands placed on the listener. As discussed above, the auditory+orthography context allows listeners to bypass word recognition because the orthographic presentation serves to limit the possible response alternatives. The auditory+orthography context requires listeners to only judge the accentedness of a stimulus based on acoustic-phonetic similarity with existing exemplars of a particular word type, effectively ignoring phonetically similar words.

In their classic study of speech intelligibility, Miller *et al.* (1951) showed that fewer response alternatives in a word-recognition task leads to higher levels of speech intelligibility at the same signal-to-noise ratio. These findings illustrate that more noise may be added to stimuli when there are fewer response alternatives while maintaining the same level of intelligibility. Miller *et al.* argued that speech intelligibility is not determined by the stimulus item alone, but also by the context in which it is perceived. Likewise, in the present study, the intelligibility of a particular stimulus is increased in the auditory+orthography context because there is essentially only a single response alternative. Though intelligibility and accentedness are not equivalent dimensions, they are correlated (Derwing and Munro, 1997). The availability of context may account for why native speakers are judged as less accented in the auditory+orthography context than in the auditory-only context. In other words, the reduction of response alternatives increases the intelligibility of the individual stimuli and thus the decrease in perceived foreign accent.

This explanation does not, however, account for the ratings of non-native speakers in the two listening contexts. Non-native speakers were instead rated as *more* accented in the auditory+orthography context than in the auditory-only context. Since the process of word recognition can be bypassed in the auditory+orthography listening context, accent ratings will be based solely on the acoustic-phonetic or phonological mismatch between a stimulus item and stored exemplars. Presenting the target word to listeners orthographically in this context may highlight how poorly a non-native production of that word matches its stored exemplars. Hence, non-native productions of words may sound more accented when listeners are informed of the word's identity. In some cases, the auditory percept may even conflict with the orthographic target (e.g., [nuz] with "noose") and therefore result in a significantly higher rating of perceived foreign accent than if the auditory stimulus were presented without its orthographic representation. Data from the number of times listeners chose to repeat stimuli provide converging evidence that context modulates a listener's judgment. Participants in the auditory-only context listened to stimuli more often than in the auditory+orthography context.

The observed interaction of listening context and native language of the speaker in this study has an important impli-

cation for future nativeness rating studies. Presenting words to listeners in an auditory+orthography context results in non-native speakers sounding *more accented* and native speakers *less accented* than in the auditory-only context. The auditory+orthography listening context thus makes the accent ratings for the two groups of speakers diverge in the appropriate directions: Native speakers are rated as less accented and non-native speakers as more accented.

While the listening context effects must be rooted in perceptual processes unrelated to speakers' productions, an account of the results surrounding lexical frequency as rooted entirely in perceptual processes is less plausible. We have argued in this section that differences in accent ratings for high, medium, and low frequency words reflect differences in the perceptions of these words based on the number of stored exemplars and their similarity to an input utterance. As was alluded to above, however, these results may be confounded by concomitant differences in the productions of high, medium, and low frequency words. This hypothesis is further investigated in the next section where acoustic measures of vowels at the three levels of lexical frequency are compared.

III. ACOUSTIC ANALYSIS

A. Methodology

To test the hypothesis that differences found in the perception of high, medium, and low frequency words reflect differences in perception and not production, we examined one measure of speech production, namely, vowel formant frequencies. Vowel formants were selected as a possible locus of production differences between words with different lexical characteristics because other lexical factors have been shown to affect the production of vowels [e.g., Wright (2003) for "easy" versus "hard" words]. Similarly, languages with the "same" vowels have been shown to have different acoustic targets [e.g., Bradlow (1995) for English, Spanish, and Greek vowels]. Furthermore, the exact location of vowel targets in American English is a salient sociolinguistic marker of dialect affiliation (e.g., Labov *et al.* 2006; Clopper *et al.*, 2006a). Thus, listeners in the perceptual rating task are likely to have had some experience discriminating differences in vowel formant frequencies. If production differences in vowel formants exist, they could be used by listeners when performing the accent rating task.

Each target word was segmented into three parts: C1 (onset consonant), V (vocalic nucleus), and C2 (coda consonant). First and second formant measures of the vowel portion were made using Praat sound editing software (Boersma and Weenick, 2004). The formant values reported here come from measures taken at the temporal midpoint of the vowel. The window length was set to 50 ms with a step size of 10 ms. Aberrant formant measures were checked by the first author using a wideband spectrogram and corrected as needed. A total of 91 out of 4080 (2.2%) tokens were hand-corrected in this manner.

Only 170 of the original 360 English words were used in the formant analysis. All words with sonorant coda consonants were eliminated from analysis due to difficulty with determining the boundary between the vowel and the final consonant. All words containing the diphthongs [aj], [oj], and [aw] were also eliminated. Only vowel categories that contained at least three examples at each of the three frequency levels were included in the analysis, eliminating [a], [o], [ɔ], and [ʊ] from analysis. The remaining vowels [æ], [ʌ], [ɛ], [e], [ɪ], [i], and [u] were used in the analysis.

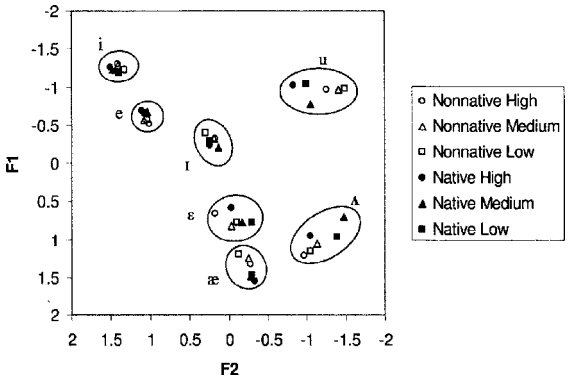


FIG. 4. Mean z-transformed formant values for native and non-native speakers at each level of lexical frequency.

All formant measures were transformed to z-scores for each person for each formant separately using eq. (1) in (Adank *et al.* 2004; Clopper *et al.* 2006a). This procedure centers each speaker's vowel space at the origin and reduces gender differences while maintaining phonemic differences. In (1), f refers to the formant frequency in hertz (F1 or F2), μ to the overall mean frequency of that formant for a specific speaker, and σ to the standard deviation of that formant for a specific speaker. A plot of the transformed z-scores for native and non-native speakers at each of the three levels of lexical frequency is given in Fig. 4.

$$z = (f - \mu)/\sigma. \tag{1}$$

B. Results and Discussion

Two ANOVAs with vowel (æ, ʌ, ɛ, e, ɪ, i, u) and lexical frequency (high, medium, low) as within-subjects factors and gender (male, female) and native language (native, non-native speakers) as between-subjects factors were calculated for the first formant and second formant z-scores separately. For both F1 and F2, a main effect of vowel was found [$F(6,120)=914.8$, $p<0.001$ for F1 and $F(6,120)=298.7$, $p<0.001$ for F2]. These results merely indicate that vowels have different F1 and F2 targets and will not be further analyzed. A main effect of lexical frequency was found only for F2 [$F(2,40)=14.2$, $p<0.001$]. *Posthoc* paired samples t tests revealed that words with high lexical frequency were more fronted (higher F2) than words of medium and low lexical frequency [$t(167)=4.58$, $p<0.001$ and $t(167)=4.37$, $p<0.001$, respectively]. Both F1 and F2 analyses also revealed significant vowel by lexical frequency interactions [$F(12,240)=8.02$, $p<0.001$ for F1, $F(12,240)=7.51$, $p<0.001$ for F2] and vowel by native language interactions [$F(6,120)=6.56$, $p<0.001$ for F1, $F(6,120)=3.54$, $p=0.003$ for F2]. In addition, the gender by vowel interaction

TABLE II. The p values of *posthoc* analyses of the vowel by lexical frequency interaction, collapsed across native and non-native speakers. Comparisons that are significant for either F1 or F2 are bold.

Vowel	Lexical frequency comparisons	F1	F2
æ	**high vs. low**	0.006	0.005
	medium vs. low		
	high vs. medium		
ʌ	high vs. low		
	medium vs. low	0.003	0.005
	high vs. medium	0.009	0.004
ɛ	**high vs. low**	0.003	0.000
	medium vs. low		
	high vs. medium	0.000	0.000
e	high vs. low		
	medium vs. low		
	high vs. medium		
I	**high vs. low**	0.001	
	medium vs. low	0.000	0.000
	high vs. medium		
i	**high vs. low**	0.003	0.000
	medium vs. low		0.000
	high vs. medium		
u	**high vs. low**		0.002
	medium vs. low	0.003	
	high vs. medium	0.008	

was significant for F1 [$F(6,120)=4.32$, $p=0.001$]. No other main effects or interactions were significant.

To further investigate these interactions, *posthoc* analyses were conducted for F1 and F2 separately. The α level was set to 0.01 due to the large number of *posthoc* analyses. Paired-samples t tests on the vowel by lexical frequency interaction revealed several significant differences between vowels in words of different levels of lexical frequency, as shown in Table II. Five of the seven vowels show some difference between high and low frequency words. This interaction, as well as the overall main effect of lexical frequency found for F2, indicates that phonetic differences between words of different levels of lexical frequency exist. Nonetheless, later analyses will reveal that phonetic differences in formant measures cannot explain all the details of the perceptual interactions involving lexical frequency.

The vowel by native language interaction indicates, not surprisingly, that natives and non-natives produce vowels differently. *Posthoc* one-way ANOVAs of this interaction revealed that non-native productions of [ʌ] were more fronted (higher F2) than native productions [$F(1,23)=7.987$, $p=0.010$]. Additionally, productions of [ɛ] and [u] exhibited differences in F2 that approached significance ($p=0.033$ and $p=0.041$). As is apparent in Fig. 4, native productions of [u] exhibited more fronting while non-native productions of [ɛ] exhibited fronting. None of the *posthoc* analyses of this interaction for F1 reached significance, although non-native productions of [æ] were marginally higher than native productions ($p=0.016$), but lower for [ʌ] and [e] ($p=0.013$ and $p=0.048$). The observed differences in formant frequencies between native and non-native productions showed that the non-native speakers, while globally producing target American English vowels correctly, were not perfect in their productions of these vowels. Differences found for [ʌ] and [æ] may be a result of the absence of these vowels in the native system for the L1-German bilingual speakers. Finally, *posthoc* analyses of the gender by vowel interaction revealed that male productions of [i] were higher (lower F1) than female productions [$F(1,23)=9.538$, $p=0.005$].

To further investigate the role that lexical frequency plays in vowel productions of native and non-native speakers, we computed the Euclidean distance in the F1×F2 plane between different levels of lexical frequency. For each speaker, the distance between high and medium lexical frequency, between medium and low lexical frequency, and between high and low lexical frequency for each vowel category was computed. This measure was designed to explore the interaction between native language and lexical frequency that was found in the perceptual ratings data. In the accent rating task, a perceptual difference between all three levels of lexical frequency was found for the native speakers: For non-native speakers the difference between high and medium frequency words was attenuated. To rule out production differences as the source of these perceptual differences, we examined the distance between vowels at different levels of lexical frequency. In particular, we were interested in whether native speakers produced a greater difference between vowels of high and medium frequency words than did non-native speakers. If such a production difference were to exist for native speakers but not for non-native speakers, it could explain the lack of a perceptual difference between high and medium frequency words that was found for the non-native speakers.

ANOVAs were conducted on the distance between high and medium frequency productions, between medium and low frequency productions, and between high and low frequency productions with vowel as the within-subjects factor and native language as the between-subjects factor. Critically, none of these ANOVAs revealed significant differences between the two groups of speakers or in the vowel by native language interaction (all $p>0.19$). The lack of a main effect of native language, especially in the comparison of high to medium frequency words, suggests that native speakers do not produce larger differences between high and medium frequency words than do non-native speakers. If such a production difference had been found, it could have been the root of the perceptual differences. These acoustic data can therefore partially rule out production (here measured as differences in vowel formants) as the cause of the perceptual difference between the two groups of speakers since native and non-native speakers did not differ in the amount of difference between words of different frequencies.

One additional analysis was conducted on the acoustic data. For each vowel category, Euclidean distances were computed between each non-native speaker's average high frequency production and the overall average native high frequency production. Corresponding distances were calculated for medium and low frequency words. This measure was designed to examine two effects found in the perceptual ratings data: (1) the general trend for high frequency non-native tokens to be rated as less accented than lower frequency tokens and (2) the native language by lexical fre-

quency interaction, which revealed that perceptual differences between native and non-native speakers were increased in words of higher frequency when compared to words of lower frequency. If the former effect were rooted in production differences, then non-native speakers should produce high frequency words *better* (i.e., closer to the native target) than words of lower frequency and therefore be rated as less accented. In other words, the Euclidean distances between native and non-native productions should be smaller for high frequency words and greater for low frequency words. The second perceptual effect mentioned above showed native and non-native speakers diverging in terms of perception for high frequency words and converging for low frequency words. If this perceptual effect arose from production differences, then non-native speakers should produce high frequency words *worse* (i.e., further from the native target) than low frequency words. In other words, the Euclidean distances between native and non-native productions should be greater for high frequency words and smaller for low frequency words.

Paired-samples *t* tests comparing high to medium, high to low, and medium to low productions revealed no significant differences, indicating that high frequency non-native productions were neither closer to nor further from native targets than medium or low frequency words. Thus, the acoustic analysis of non-native vowel productions from words in the different lexical frequency groups allows us to at least partially rule out the possibility that non-native tokens of high frequency words are acoustically better (or worse) productions. Despite broad differences found in the productions of vowels in words of different levels of lexical frequency, vowel formant production differences do not account for the specific interactions found in the perceptual data. Of course, measurements of other phonetic characteristics (e.g., VOT, degree of obstruent-final devoicing, or vowel formant trajectories) could reveal systematic production differences between levels of lexical frequency. Nonetheless, the lack of consistent effects attributable to an acoustic measure known to be affected by frequency lends support to the claim that the frequency results found in the perception experiment are due primarily to perceptual factors.

IV. GENERAL DISCUSSION AND CONCLUSIONS

The results of the present study demonstrate that two speaker-independent factors—lexical frequency and listening context—affect the perception of degree of foreign accent in isolated spoken words. Listeners consistently perceived high frequency words as less accented than low frequency words. Simultaneously presenting a target word to listeners both auditorily and orthographically attenuated the effect of frequency, however. Furthermore, the addition of orthographic information in the auditory+orthography context caused native speakers of English to be rated as less accented and non-native speakers of English to be rated as more accented than in the auditory-only context. Evidence from acoustic analyses of vowel formant frequencies for native and non-native productions demonstrated that production differences do exist between native and non-native tokens. These analyses further showed that production differences between words at different levels of lexical frequency are found for some, but not all, vowels. However, production differences do not seem to account for some of the perceptual interactions, especially the native language by frequency interaction, thus confirming the role of perception in the ratings of words of different levels of lexical frequency.

The explanation of the effects of lexical frequency and listening context was framed in an exemplar model of speech perception. In exemplar models, words with higher lexical frequency have more stored exemplars in memory. We hypothesized that words of higher frequency would be rated as less accented because the greater number of stored exemplars makes it more likely that an incoming stimulus will be similar to existing targets, resulting in their being rated as less accented. This prediction was supported by our data.

A second effect that can be accounted for within this framework was the attenuation of lexical frequency effects in the auditory+orthography condition. In this listening context, the process of auditory word recognition can be bypassed. A pop-out effect was induced for native tokens presented in the auditory+orthography context. Tokens become more intelligible and are rated as less accented than in the auditory-only context. This effect can be accounted for in an exemplar framework by assuming that in the auditory-only context, some words of the wrong lexical category (e.g., competing "bet" for target "bat") may cause a decrease in the goodness-of-fit between an incoming token and the target lexical item, increasing the amount of perceived accent. In the auditory+orthography context, on the other hand, all competing targets and their exemplars are eliminated, causing the incoming token to be rated as less accented. This effect of context on perceived accentedness was reversed for non-natives. Non-native tokens may sometimes be so deviant as to completely switch lexical categories, something less likely to occur among native tokens. In these cases, tokens will be perceived as extremely aberrant and receive very high ratings of accentedness.

Our results may be consistent with other models of speech perception and spoken word recognition. The TRACE model (McClelland and Elman, 1986; McClelland *et al.*, 2006) allows for both bottom-up and top-down processing of speech. Because top-down information can affect processing of ambiguous or degraded sounds, TRACE correctly predicts a decrease in perceived accent in the auditory+orthography listening context. This model alone may not be able to account for the variation of these effects, however. Norris *et al.* (2000) point out that top-down processing in TRACE acts to correct erroneous information in the input and therefore prevents a listener from reliably detecting and perceiving mispronunciations. One might expect that mispronunciations in non-native input would be similarly overlooked, resulting in non-native speakers also being rated as less accented in the auditory+orthography context. However, results from the perceptual ratings indicate that this is not the case. A full examination of our data within the TRACE model is beyond the scope of the current article, though some aspects appear promising and merit further examination.

The findings of the current study have several implications for future research on accent perception. First, the present results demonstrate that researchers should consider the role that lexical frequency plays in perception of degree of foreign accent. If the effects of frequency are to be avoided or minimized, an orthographic representation of the target word can be used to attenuate these effects. Second, presenting target words to listeners both auditorily and orthographically yields different measures of perceived foreign accent; in the auditory+orthography context, native speakers were rated as less accented while non-native speakers were rated as more accented. The auditory+orthography context thus mitigates the effects of lexical frequency on accent ratings and also helps listeners better distinguish speech samples of native and non-native speakers.

The results of this study also have several broader theoretical implications. Our findings show that perceived degree of foreign accent is not just related to a speaker's language learning history or to how well a speaker is able to phonetically approximate native speech, but also depends on the context in which that voice is perceived. We have shown here that perception of foreign accent is partially dependent upon non-acoustic properties of the signal such as listening context and possibly lexical frequency. Previous work has shown that the intelligibility of L2 speakers reflects not only the actual acoustic realization of non-native productions but also the prior experience and history of the listener. For example, Bent and Bradlow (2003) found that the intelligibility of several groups of non-native speakers depended on the language background of the listeners. Non-native listeners in both a matched and mismatched native language background performed equally well in a sentence intelligibility task with proficient non-native and native speakers. Native listeners, on the other hand, found all non-native speakers to be less intelligible than native speakers. Accent perception is therefore shaped to a large extent by a *listener's* past experiences and developmental history. A speaker may therefore only have an "accent" within a specific perceptual framework and listening context. The perception of a foreign accent thus reflects not only properties of the speaker, but also prior experience of the listener and factors that reflect the attunement between speaker and listener.

The influences of two speaker-independent factors—lexical frequency and listening context—on the perception of foreign accent in this study can be accounted for in large part by casting the process of accent perception more broadly within the framework of exemplar models of speech perception and spoken word recognition. We have assumed that the perception of foreign accent reflects the degree to which there is an acoustic-phonetic mismatch between a stimulus token and the stored exemplars in the listener's memory. The validity and robustness of this theoretical framework can be tested in future research. One possible way to test this framework is to manipulate the amount of experience that listeners have in listening to L2 speech. It has been noted above that highly experienced listeners (e.g., linguists and ESL teachers) sometimes rate the accents of non-native speakers more leniently (i.e., as less accented) than do naive listeners. This result may occur because experienced listeners have more exposure to L2 speech and therefore more L2 speech exemplars in memory than naive listeners. Thus, experienced listeners are more likely to find an acoustic match to incoming tokens of L2 speech and rate these tokens as less accented than naive listeners. Increasing the amount of experience that naive listeners have with L2 speech by presenting them with many tokens of L2 speech should therefore make them more tolerant perceivers of foreign accent in speech produced by unfamiliar L2 speakers. Experimental studies such as these should help increase our understanding of the process by which foreign accents are perceived and also provide us with a more complete picture of what it means for a speaker to "have a foreign accent."

ACKNOWLEDGMENTS

Preparation of this manuscript was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD T32 Training Grant No. DC-00012 and NIH-NIDCD Research Grant No. R01 DC-00111). We wish to thank Jennifer Karpicke and Christina Fonte for help with data collection and Adam Buchwald and three anonymous reviewers for useful comments on this and previous versions of this manuscript.

[1]This type of speech is created by filtering the original signal into six logarithmically spaced frequency bands.

[2]We wish to thank an anonymous reviewer for bringing this point to our attention.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (**1995**) "The CELEX Lexical Database (Release 2) [CD-ROM]," (Linguistic Data Consortium, University of Pennsylvania [Distributor], Philadelphia, PA).

Bent, T., and Bradlow, A. R. (**2003**). "The interlanguage speech intelligibility benefit," J. Acoust. Soc. Am. **114**, 1600–1610.

Boersma, P., and Weenink, D. (**2004**). "Praat: Doing phonetics by computer (Version 4.2.12)" [Computer program], http://www.praat.org/ (accessed 3/19/07).

Bongaerts, T., van Summeren, C., Planken, B., and Schils, E. (**1997**). "Age and ultimate attainment in the pronunciation of a foreign language," Stud. Second Lang. Acquis. **19**, 447–465.

Bradlow, A. R. (**1995**). "A comparative acoustic study of English and Spanish vowels," J. Acoust. Soc. Am. **97**, 1916–1924.

Clopper, C. G., Pisoni, D. B., and de Jong, K. (**2006a**). "Acoustic characteristics of the vowel system of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Clopper, C. G., Pisoni, D. B., and Tierney, A. T. (**2006b**). "Effects of open-set and closed-set task demands on spoken word recognition," J. Am. Acad. Audiol **17**, 331–349.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Derwing, T. M., and Munro, M. J. (**1997**). "Accent, intelligibility, and comprehensibility: Evidence for four L1s," Stud. Second Lang. Acquis. **20**, 1–16.

Fathman, A. (**1975**). "The relationship between age and second language productive ability," Lang. Learn. **25**, 245–253.

Flege, J. E., and Fletcher, K. L. (**1992**). "Talker and listener effects on degree of perceived foreign accent," J. Acoust. Soc. Am. **91**, 370–389.

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995**). "Factors affecting strength of perceived foreign accent in a second language," J. Acoust. Soc. Am. **97**, 3125–3134.

Goldinger, S. (**1997**). "Words and voices: Perception and production in an episodic lexicon," in *Talker Variability in Speech Processing*, edited by K.

Johnson and J. Mullennix (Academic, San Diego), pp. 33–66.
Grosjean, F. (**1980**). "Spoken word recognition processes and the gating paradigm," Percept. Psychophys. **28**, 267–283.
Hintzman, D. (**1986**). "Schema abstraction in a multiple-trace memory model," Psychol. Rev. **93**, 411–428.
Hintzman, D. (**1988**). "Judgments of frequency and recognition memory in a multiple-trace memory model," Psychol. Rev. **95**, 528–551.
Howes, D. (**1957**). "On the relation between the intelligibility and frequency of occurrence of English words," J. Acoust. Soc. Am. **29**, 296–305.
Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. Mullennix (Academic, San Diego), pp. 145–165.
Kirchner, R. (**1999**). "Preliminary thoughts on "phonologisation" within an exemplar-based speech processing system," UCLA Work. Papers Ling. Vol. 1 (Papers Phonol. 2), pp. 207–231.
Kirchner, R. (**2004**). "Exemplar-based phonology and the time problem: A new representational technique," poster presented at LabPhon 9 Conference, 28 June.
Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (Mouton de Gruyter, Berlin), www.linguistics.uiuc.edu/labphon9/Abstract_PDF/kirchner.pdf Viewed 3/27/07.
Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing Spoken Words: The neighborhood activation model," Ear Hear. **19**, 1–36.
MacLennan, R. N. (**1993**). "Interrater reliability with SPSS for Windows 5.0," Am. Stat. **47**, 292–296.
McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.
McClelland, J. L., Mirman, D., and Holt, L. L. (**2006**). "Are there interactive processes in speech perception?" Trends Cogn. Sci. **10**, 363–369.
McGraw, K. O., and Wong, S. P. (**1996**). "Forming inferences about some intraclass correlation coefficients," Psychol. Methods **1**, 30–46.
Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.
Munro, M. J., and Derwing, T. M. (**2001**). "Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate," Stud. Second Lang. Acquis. **23**, 451–468.
Norris, D., McQueen, J. M., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–370.
Oyama, S. (**1976**). "A sensitive period for the acquisition of a nonnative phonological system," J. Psycholinguist. Res. **5**, 261–283.
Pierrehumbert, J. (**2001**). "Exemplar dynamics: Word frequency, lenition, and contrast," in *Frequency Effects and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (John Benjamins, Amsterdam), pp. 137–157.
Pierrehumbert, J. (**2002**). "Word-specific phonetics," in *Laboratory Phonology VII*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 101–140.
Piper, T., and Cansin, D. (**1988**). "Factors influencing the foreign accent," Can. Mod. Lang. Rev. **44**, 334–342.
Piske, T., MacKay, I. R. A., and Flege, J. E. (**2001**). "Factors affecting degree of foreign accent in an L2: Areview," J. Phonetics **29**, 191–215.
Pollack, I., Rubenstein, H., and Decker, L. (**1959**). "Intelligibility of known and unknown message sets," J. Acoust. Soc. Am. **31**, 273–279.
Savin, H. B. (**1963**). "Word-frequency effect and errors in the perception of speech," J. Acoust. Soc. Am. **35**, 200–206.
Sommers, M. S., Kirk, K. I., and Pisoni, D. B. (**1997**). "Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format," Ear Hear. **18**, 89–99.
Southwood, M. H., and Flege, J. E. (**1999**). "Scaling foreign accent: direct magnitude estimation versus interval scaling," Clin. Linguist. Phonetics **13**, 335–349.
Thompson, I. (**1991**). "Foreign accents revisited: The English pronunciation of Russian immigrants," Lang. Learn. **41**, 177–204.
Wright, R. (**2003**). "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology, VI: Phonetic Interpretation*, edited by J. Local, R. Ogden, and R. Temple (Cambridge U. P., Cambridge, UK), pp. 75–87.

Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners

Ann R. Bradlow[a] and Jennifer A. Alexander[b]
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

(Received 25 July 2006; revised 11 January 2007; accepted 17 January 2007)

Previous research has shown that speech recognition differences between native and proficient non-native listeners emerge under suboptimal conditions. Current evidence has suggested that the key deficit that underlies this disproportionate effect of unfavorable listening conditions for non-native listeners is their less effective use of compensatory information at higher levels of processing to recover from information loss at the phoneme identification level. The present study investigated whether this non-native disadvantage could be overcome if enhancements at various levels of processing were presented in combination. Native and non-native listeners were presented with English sentences in which the final word varied in predictability and which were produced in either plain or clear speech. Results showed that, relative to the low-predictability-plain-speech baseline condition, non-native listener final word recognition improved only when both semantic and acoustic enhancements were available (high-predictability-clear-speech). In contrast, the native listeners benefited from each source of enhancement separately and in combination. These results suggests that native and non-native listeners apply similar strategies for speech-in-noise perception: The crucial difference is in the signal clarity required for contextual information to be effective, rather than in an inability of non-native listeners to take advantage of this contextual information per se. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2642103]

PACS number(s): 43.71.Es, 43.71.Hw, 43.71.Gv [PEI] Pages: 2339–2349

I. INTRODUCTION

Speech recognition differences between native and highly proficient non-native listeners tend to emerge under suboptimal listening conditions. For example, several studies have shown that in the presence of background noise or reverberation, non-native listeners who had attained a very high level of proficiency in English were less accurate at speech recognition than native listeners. In contrast, under more favorable listening conditions, speech recognition accuracy by these highly proficient non-native listeners was comparable to that of native listeners (e.g., Nábelek and Donahue, 1984; Takata and Nábelek, 1990; Mayo *et al.*, 1997; Meador *et al.*, 2000; Rogers *et al.*, 2006). Thus, the general pattern of experimental findings, and the common experience of non-native listeners, is that the detrimental effects of environmental signal distortion are greater for non-native than native language speech perception.

One possible explanation for this phenomenon is that the primary source of the sharper decline for non-native than native listener speech-in-noise perception is at the segmental level. According to this explanation, the masking effects of noise on the acoustic cues to phoneme identification are more detrimental for non-native listeners than for native listeners, presumably due to their reduced experience with the full range of cues for any given phoneme. Specifically, native listeners are likely to have developed highly effective and efficient strategies for compensating for the masking effects of background noise and reverberation by focusing their attention on segmental cues that are less vulnerable to noise-related distortion [e.g., see Parikh and Loizou (2005), and Jiang *et al.* (2006) for some examples of noise-induced cue-weighting shifts in English]. In contrast, based on learned patterns of perception from the native language, non-native listeners may attend primarily to cues that, while relatively reliable in the non-native target language in quiet, are largely obscured by background noise.

An alternative explanation for the extra difficulty of non-native listeners under suboptimal listening conditions is that noise and/or reverberation have detrimental effects at all levels of processing with the result that overall non-native levels of performance on sentence- or word-in-noise recognition tasks reflect cumulative effects of noise throughout the processing system. For example, in addition to the dramatic effects of noise on acoustic cues to segment identity, the presence of background noise can dramatically affect listener access to prosodic boundary cues, such as silent pauses or sudden rises and falls in fundamental frequency. Native listeners may be able to compensate for this loss of phonetic information for phrase and discourse structure by drawing on their highly practiced sentence processing mechanisms. In contrast, due to their relatively poorly developed syntactic, semantic, and pragmatic processing skills in the target language, non-native listeners cannot readily draw on higher-level linguistic structural and contextual information in order to recover from losses at the perceptual level.

In a recent direct test of these alternatives, Cutler *et al.* (2004) examined native and non-native listener phoneme perception in the context of meaningless CV and VC syllables. These syllable-sized stimuli ensured that the perception of phoneme-level information was isolated from lexical-

[a]Electronic mail: abradlow@northwestern.edu
[b]Electronic mail: jenalex@northwestern.edu

or sentence-level information. Cutler *et al.* found similar declines in English phoneme identification with a decrease in signal-to-noise ratio for native and for highly proficient Dutch-speaking non-native listeners. This finding indicates that disproportionate difficulties with speech-in-noise perception for non-native listeners such as these probably do not stem from disproportionate effects of noise at the level of phoneme identification where word- and sentence-level factors are irrelevant. Instead, Cutler *et al.* (2004) suggest that the special troubles of non-native speech-in-noise perception result from the compounding of difficulties at lower levels of processing with limitations at higher levels.

Though limited to highly proficient Dutch listeners, the finding reported in Cutler *et al.* (2004) dovetails perfectly with a finding reported in Mayo *et al.* (1997). These authors found that highly proficient late bilinguals (listeners who had acquired the target language after puberty) benefited less from sentence- level contextual information for word recognition than either native listeners or highly proficient early bilinguals (who had acquired the target language as an infant or toddler). Mayo *et al.* (1997) examined sentence-final word recognition across various signal-to-noise ratios and across conditions in which the target word was either highly predictable or not at all predictable from the preceding sentence context. The native and early bilingual listeners tolerated significantly greater noise levels for high than for low predictability sentences. In contrast, the late bilinguals showed no difference in noise tolerance levels across high and low predictability sentences. Furthermore, the slopes of the psychometric functions (accuracy across signal-to-noise ratios) for the native listeners and the early bilinguals were significantly greater for the high predictability sentences than for the low predictability sentences; that is, there was a dramatic effect of noise on final word recognition accuracy in exactly the sentences where context mattered. In contrast, the late bilinguals showed approximately parallel psychometric functions for the high and low predictability sentences. Thus, while the native and early bilingual listeners showed a strong benefit from contextual information at the sentence level, relatively late (yet highly proficient) non-native listeners showed essentially no benefit for final word recognition in high predictability sentences as compared with final word recognition in low predictability sentences.

A remaining question is whether the observed non-native listener deficit in drawing on higher level contextual information to aid in speech recognition persists even under enhanced signal conditions. It is possible that the sentence-in-noise perception patterns across native and non-native listeners are qualitatively similar in showing reduced exploitation of contextual information under degraded signal conditions and significant benefits from context under more favorable listening conditions. Indeed, if extended over a wider range of noise levels in both directions, it is possible that the average high and low predictability psychometric functions presented by Mayo *et al.* (1997) for all listeners groups (native, earlier bilinguals, and later bilinguals) would diverge dramatically at some point on the low-noise end and converge at some point on the high-noise end of the noise level axis, indicating some degree of context dependency for all listeners. Thus, it is possible that non-native speech-in-noise perception is not necessarily doomed by limited abilities to exploit contextual information as a means of recovery from recognition losses at the levels of phoneme identification and lexical access. Instead, if given a signal of sufficient acoustic clarity, non-native listeners (both early and late learners) may exhibit a strong ability to exploit semantic-contextual information at the sentence level.

Accordingly, the purpose of the present study was to investigate the ability of non-native listeners to benefit from a combination of semantic and acoustic enhancements for sentence-in-noise recognition. We tested the hypothesis that for all listeners, including native listeners as well as non-native listeners across the full range of proficiency levels in the target language, speech recognition accuracy is facilitated by the availability of higher-level semantic information to the extent that such information is well specified in the signal. What varies across listeners with different levels of proficiency and experience with the target language is the definition of "well specified." For native listeners with highly practiced skills in phonetic processing of the language, contextual information early on in an utterance may be sufficiently well perceived even in conditions with relatively high levels of signal distortion and/or masking to provide an effective means of recovering from perceptual losses. In contrast, for non-native listeners with less practice and experience with the sound structure of the target language, contextual information that occurs early in an utterance may only be a useful source of information for later-occurring portions of the utterance if it is sufficiently well-perceived. This, in turn, may require a relatively high degree of signal clarity. A specific prediction of this hypothesis that we set out to test is that under conditions where the signal clarity is enhanced substantially due to a clear speaking style on the part of the talker, non-native listeners should exhibit a recognition advantage for words that are highly predictable from the preceding utterance in comparison with words that are not at all predictable from the preceding context. That is, we sought to extend the finding of Mayo *et al.* (1997) to the situation where greater contextual information is made available to non-native listeners by means of signal enhancement through naturally produced clear speech. In contrast to Mayo *et al.* (1997) (and due primarily to practical limitations regarding the available pool of study participants) the present study did not systematically vary non-native listener proficiency. The focus of this study was instead on identifying the conditions under which a group of non-native listeners with varying levels of target language proficiency could make use of contextual information for the processing of an incoming spoken word.

The overall design of this experiment involved the manipulation of two independent factors: semantic cues and acoustic cues. Following numerous previous studies (e.g. Kalikow *et al.*, 1977; Mayo *et al.*, 1997; Fallon *et al.*, 2002 and many others) the semantic-contextual cues were manipulated by using English sentences in which the final word was highly predicted by the preceding words ("high probability sentences") and sentences in which the final word could not be predicted on the basis of the preceding words ("low prob-

ability sentences"). The acoustic-phonetic cues were manipulated by including English sentence recordings in both plain and clear speaking styles (e.g., Picheny *et al.*, 1985; Ferguson and Kewley-Port, 2002; Bradlow and Bent, 2002; Uchanski, 2005; Smiljanic and Bradlow, 2005). Both native and non-native English listeners responded to all four types of sentences (high probability clear speech, high probability plain speech, low probability clear speech, and low probability plain speech), allowing us to assess the separate and combined effects of semantic and acoustic cues to speech-in-noise perception by native and non-native listeners.

In contrast to the terminology used in previous work (e.g., Picheny *et al.*, 1985, 1986, 1989; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Ferguson and Kewley-Port, 2002; Bradlow and Bent, 2002; Uchanski, 2005; Smiljanic and Bradlow, 2005, and several others), in this report we have adopted the term "plain speech" instead of "conversational speech." We propose this change in terminology in order to better reflect the fact that this mode of speech production is distinct from truly conversational speech as presented in, for example, the Buckeye Corpus of Conversational Speech (http://buckeyecorpus.osu.edu/), in which talkers were recorded conversing freely with an interviewer in a small seminar room. For our purposes, the key distinction between "plain" and "clear" speech is with respect to intelligibility. In view of the fact that the plain speech samples are read from a script (i.e., not spontaneous responses to a topic or interviewer's question) and are recorded in a relatively formal laboratory setting (i.e., in a sound-attenuated booth rather than in a more relaxed conversational setting), and since their primary purpose is to serve as a baseline from which to measure the intelligibility advantage of clear speech, we refer to them in the present paper as plain rather than conversational speech recordings.

II. METHOD

A. Materials

Since none of the previously published sets of high and low predictability sentences were designed for use with the population of non-native listeners of interest in this study (i.e., non-native listeners with considerably less experience with spoken English than the participants in the study of Mayo *et al.*, 1997), we followed the general procedures outlined in Fallon *et al.* (2002) to develop a new set of sentences. First, a list of approximately 300 high probability sentences was compiled by combining the sentence lists published in Kalikow *et al.* (1977), Bench and Bamford (1979), Munro and Derwing (1995), and Fallon *et al.* (2002) with some original sentences of our own that followed that same general pattern of those in Kalikow *et al.* (1977) and Fallon *et al.* (2002). The sentences were printed with the final (target) word of each sentence replaced by a dotted line. The sentences were randomly arranged into two lists ("List A" and "List B," respectively); each of the two lists contained exactly half of the sentences.

In order to assess the predictability of the final word in these candidate sentences for our target subject population, 24 non-native English speakers evaluated the sentences in Lists A and B. These non-native English speakers were recruited from the group of participants in the Northwestern University International Summer Institute (ISI) 2004 (as described in the following, this program provides intensive English language instruction and general acculturation for incoming international graduate students). Each participant was given a printout of either List A or List B and was asked to make his or her best guess as to the identity of the final word of each sentence. They were required to work individually and were not allowed to use dictionaries or any other reference resources. The task took between 20 and 50 min to complete. All of these subjects were paid for their participation.

Following this initial test, 113 of the "best" sentences, that is, those that yielded the most consistent responses, were combined into a third list, "List C." We then presented List C with the final (target) word of each sentence replaced by a dotted line to 14 native English speakers. These participants were recruited from the Northwestern University Department of Linguistics subject pool and received course credit for their participation. Their responses were tallied and, again, only those sentences that yielded the most consistent responses were included in a fourth list, "List D." List D, which contained 89 sentences, was then checked for predictability with the same procedure as described earlier by 9 non-native English speakers who were recruited from the Northwestern University English as a Second Language (ESL) program. Finally, J.A.A. created a low predictability match for each high predictability sentence using sentence frames modified slightly from those published in Fallon *et al.* (2002).

The final list (provided in the Appendix) consisted of 120 sentences (60 high and 60 low predictability). In this final set of sentences, 43% (26/60) of the high predictability sentences were completed by the final group of 9 non-native respondents with 100% consistency (i.e., 9/9 respondents filled in the same final words), 30% (18/60) were completed with 89% consistency (i.e., 8/9 respondents filled in the same final words), 23% (14/60) were completed with 78% consistency (i.e., 7/9 respondents filled in the same final words), and 3% (2/60) were completed with 67% consistency (i.e., 6/9 respondents filled in the same final words). For the test with native listeners, 83% of the words (50/60) were completed with 100% consistency (14/14 respondents), 13% (8/60) were completed with 93% consistency (13/14 respondents), and the remaining 3% (2/60) were completed with 86% consistency (12/14 respondents).

One monolingual female talker of American English (aged 30 years, with no known speech or hearing impairment) was recorded producing the full set of 120 sentences in both clear and plain speaking styles (for a total of 240 recorded sentences). The complete set of 120 sentences was read first in the plain speaking style followed by a second recording in the clear speaking style. At the time of the plain style recording the talker was not aware of the fact that a clear speaking style condition would follow. The plain speaking style was elicited by instructing the talker to "read the sentences as if you are talking to someone familiar with your voice and speech patterns." The clear speaking style

was elicited by instructing her to "read the sentences very clearly, as if you are talking to a listener with a hearing loss or to a non-native speaker learning your language."

The high and low predictability sentences were mixed together and printed in random order on sheets of paper in groups of 21 with 2 filler sentences at the top and bottom of each page, giving a total of 25 sentences per page (21 target sentences +4 fillers) over a total of 6 pages. The filler sentences were presented to help the talker avoid "list intonation" around the page boundaries. The recording session was conducted in a double-walled, sound-attenuated booth. The talker wore a head-mounted microphone (AKG C420 Headset Cardioid Condenser) and the speech was recorded directly onto flash card using a Marantz PMD670 Professional Solid-State digital recorder (22.050 kHz sampling rate, 16 bit amplitude resolution).

The recorded sentences were segmented into individual sentence-length files which were subsequently equated for rms amplitude across the whole sentence duration. Each file was then digitally mixed with speech-shaped noise at a signal-to-noise ratio of +2 dB for presentation to the non-native test subjects and at a signal-to-noise ratio of −2 dB for presentation to the native listener control subjects. Each of the final stimulus files consisted of a 400 ms silent leader, followed by 500 ms of noise, followed by the speech-plus-noise file, and ending with a 500 ms noise-only tail. The noise in the 500 ms, noise-only header and tail was always at the same level as the noise in the speech-plus-noise portion of the stimulus file. These signal-to-noise ratios (+2 dB for non-natives and −2 dB for natives) were determined on the basis of our experience with prior speech-in-noise experiments with comparable (but different) sentence materials and listener groups, as well as some (rather limited) pilot testing with the current set of stimuli (all sentence types, i.e., high and low context in plain and clear speech) and a small number of subjects (6 native and 3 non-native listeners).

We selected these signal-to-noise ratios with the goal of eliciting comparable levels of speech recognition accuracy for the native and non-native listeners. The advantage of this approach is that it ensures that all listeners are operating at approximately the same effective level of speech recognition accuracy and therefore group differences in performance relative to the baseline condition (i.e., low context sentences in plain speech) are not subject to confounding influences of starting level differences. There is precedent for adopting this approach in, for example, the literature on speech perception by elderly listeners (e.g., Sommers, 1996, 1997). Here we adopt this approach with the understanding that, while the effect on speech recognition of higher noise levels for native listeners may not be qualitatively identical to the effect of perceptual mistuning of the non-native listeners to the target language, the overall equivalence of performance levels in the baseline condition due to different signal-to-noise ratios facilitated valid comparisons of improvement with contextual and/or phonetic enhancement.

B. Participants

Ninety-three adults participated in this study. Of these participants, 57 were native speakers and 36 were non-native speakers of American English. All of the native English speakers were recruited from the Northwestern University Department of Linguistics subject pool and received course credit for their participation. Of the 57 native speakers of English, 21 were excluded from the final analyses due to experimenter error ($n=8$), a bilingual language background ($n=11$) or a reported speech or hearing impairment ($n=2$). The remaining 36 native English speaking participants (22 females and 14 males) ranged in age from 17 to 30 years.

The non-native listeners were recruited from ISI 2005 (Northwestern's International Summer Institute) and received payment for their participation. The participants in this program had all been accepted into a graduate program at Northwestern University and had therefore demonstrated a relatively high level of proficiency with English communication (as measured by a minimum score of 560 on the pencil-and-paper TOEFL examination or 220 on the computer-based version of the test). Based on the accepting departments' subjective experiences with the spoken English skills of previous students from the students' home countries, the participants had been nominated for participation in this summer program, which provides one month of intensive English language training as well as a general introduction to life in the United States.

Table I provides additional details regarding the non-native participants in this study. As shown in Table I, the non-native participants came from various native language backgrounds with the breakdown as follows: Mandarin Chinese ($n=23$), Italian ($n=3$), Korean ($n=2$), Tamil ($n=2$), and 1 each of French, German, Gujarati, Hindi, Kikuyi, and Telugu. They ranged in age from 21 to 32 years, and had 9–23 years of English language study. At the time of testing, the majority of non-native participants had between 1 and 4 weeks of experience living in an English speaking environment. Three non-native listeners had several more weeks (6, 11, and 35 weeks) and 2 non-natives had 2–3 years worth of experience living in the United States prior to testing. As part of the ISI program orientation, participants were divided into eight groups based roughly on proficiency level as determined by the Speaking Proficiency English Assessment Kit (SPEAK) test. (As a reference point, note that Northwestern University requires a SPEAK score of 50 for a non-native English speaking student to be appointed as a teaching assistant). As shown in Table I and explained further in the following, we attempted to distribute the number of non-native listeners from each proficiency group evenly across four experimental conditions (A–D).

It should be noted that the group of non-native listeners was not balanced in terms of native language background. The group was dominated by native speakers of one language: 23 of the 36 non-native listeners, or 64%, were Mandarin speakers. Moreover, the remaining 13 non-native listeners came from vastly different native language backgrounds, making it impossible to conduct any meaningful comparisons across listeners with different native languages. This distribution of native languages is typical of the Northwestern ISI program participants and is a fairly accurate reflection of the distribution of international graduate students across the university. Since the task in the present

TABLE I. Some measures of the English language experience of the non-native listener participants.

Presentation condition	SPEAK[a] test score	Native language	Age at test	Time in USA (weeks)	English study (years)
A	51.6	Tamil	23	4	18
A	49.2	Mandarin	22	2	12
A	48	Mandarin	24	2	12
A	46	Mandarin	22	157	17
A	42.5	Mandarin	22	1	12
A	40	Mandarin	21	4	11
A	39.5	Mandarin	22	4	10
A	37.9	Mandarin	25	4	12
A	37.0	Mandarin	22	1	10
A	36.7	Mandarin	23	3	12
B	59.1	German	28	35	22
B	56.2	Tamil	21	6	17
B	47.7	Korean	27	4	16
B	46.5	Mandarin	23	1	13
B	41.7	Mandarin	22	2	10
B	40.4	Mandarin	26	2	13
B	38.6	Mandarin	30	4	17
B	37.5	Mandarin	25	4	13
B	30	Mandarin	22	4	10
C	52.0	French	23	116	12
C	50.4	Hindi	23	3.5	18
C	42.9	Mandarin	22	2	10
C	40.5	Mandarin	28	3	16
C	40	Mandarin	22	1	10
C	39.75	Mandarin	22	4	10
C	N/A	Italian	22	2	16
D	56.7	Gujarati	23	11	19
D	49.6	Kikuyu	26	4	23
D	43.7	Mandarin	21	2	10
D	41.8	Italian	24	2	16
D	41.2	Mandarin	24	2	9
D	40	Telugu	31	2	9
D	39.3	Mandarin	22	2	14
D	36.7	Mandarin	21	4	9
D	32.5	Korean	32	3	19
D	N/A	Italian	26	2	16

SPEAK=Speaking English Assessment Kit.

study (recognition of words in simple English sentences) requires processing over many levels of representation rather than focusing on specific phonetic contrasts, we assumed that our data would reveal general patterns of non-native language speech recognition rather than specific patterns of interactions between structural features of the target language and the listeners' native languages. Nevertheless, we acknowledge this limitation of our data.

C. Procedure

Both native and non-native subjects were tested in groups of one to three. The data collection session began with a language background questionnaire which probed the subjects' language learning experiences (both native and foreign languages) as well as their self-reported performance on standardized tests of English language proficiency (non-native subjects only). For the sentence-in-noise recognition test, each participant was seated in front of a computer monitor in a sound-attenuated booth. The audio files were played via the computer sound card over headphones (Sennheiser HD580) at a comfortable listening level, which was set by the experimenter before the start of the experiment. The participant's task was to listen to each sentence stimulus and write just the final word on specially prepared answer sheets. After each trial, the participant pressed a button on a response box to trigger the start of the next trial. Each trial was presented only once, but subjects could take as long as they needed to record their responses.

TABLE II. Distribution of sentences across the four presentation conditions. Sentence numbers correspond to the listing in the Appendix .

Condition	High context plain style	Low context plain style	High context clear style	Low context clear style
A	1–15	31–45	16–30	46–60
B	16–30	46–60	1–15	31–45
C	31–45	1–15	46–60	16–30
D	46–60	16–30	31–45	1–15

In order to familiarize subjects with the task, the test session began with 8 practice items (these items were not included in the subsequent test). After completion of these practice items the experimenter checked that the subject understood the task and was ready to begin the test. Each subject responded to a total of 60 sentences, which included a randomly ordered presentation of 15 high context sentences in plain speech, 15 low context sentences in plain speech, 15 high context sentences in clear speech, and 15 low context sentences in clear speech. Over the course of the entire test session, each subject heard each of the 60 target words only once. In order to guard against any inherent differences in either keyword or sentence intelligibility across the four style-context conditions, four test conditions were compiled such that the 60 target words/sentences were evenly distributed across the conditions (see Table II). For the high probability sentences, the response consistency rates from both the native and non-native respondents in the testing during the sentence development phase were evenly distributed across the four sublists shown in Table II: The four average consistency rates for the native respondents ranged from 96% to 100%; for the non-native respondents consistency rates ranged from 85% to 92%. As shown in Table I, these four presentation conditions were evenly distributed within the non-native participants' proficiency level groups, and equal numbers of native subjects participated in each of the four conditions. The average SPEAK score for participants in conditions A, B, C, and D were 42.9, 44.2, 44.3, and 42.4, respectively, which are very close to the overall average SPEAK score of 43.4.

Each subject received a final word recognition accuracy score out of 15 for each of the four word conditions (high context plain speech, low context plain speech, high context clear speech, and low context clear speech). Cases of responses containing alternate or incorrect spellings were accepted as correct when the scorer felt sure that the participant had correctly identified the target word but simply failed to spell it as expected. The scores were converted to percent

TABLE III. Final word durations in milliseconds for each of the four sentence types.

	High context plain style	Low context plain style	High context clear style	Low context clear style
Mean	452	459	594	613
Median	450	461	596	614
Std. Dev.	62	58	69	74
Std. Error	8	8	9	10
Minimum	242	305	424	461
Maximum	610	572	835	812

correct scores, and then converted to rationalized arcsine transform units (RAU) (Studebaker, 1985) for the statistical analyses. Scores on this scale range from −23 RAU (corresponding to 0% correct) to +123 RAU (corresponding to 100% correct).

III. RESULTS

A. Durational analyses of the sentence stimuli

Before turning to the final word recognition accuracy results, we examined the durations of all final words in the full set of 240 sentence stimuli. A decrease in speaking rate is a well-documented feature of clear speech (e.g., Picheny *et al.*, 1986, 1989; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Bradlow *et al.*, 2003; Krause and Braida, 2002, 2004; Smiljanic and Bradlow, 2005); we therefore expected a highly significant increase in final word duration for clear versus plain speech. Moreover, several studies have indicated a general tendency toward some degree of hyperarticulation for words in low predictability contexts relative to words in high predictability contexts (e.g., Lieberman, 1963; Jurafsky *et al.*, 2001; Wright, 2002; Aylett and Turk, 2004; Munson and Solomon, 2004; Munson, 2007; Scarborough, 2006). Since the purpose of this study was to examine the interaction of higher-level semantic cues and lower-level acoustic cues to final word recognition, it was necessary to determine the extent to which the words in our high versus low predictability sentence contexts differed at the acoustic level too. Here we focus exclusively on word duration as an indicator of hyperarticulation, although it should be noted that the effects of style and predictability are likely to be reflected along multiple acoustic dimensions.

Table III shows some descriptive statistics for final word duration in each of the four sentence types. A two-factor repeated-measures analysis of variance (ANOVA) showed highly significant main effects of style [plain versus clear, $F(1,59)=420.66$, $p<0.0001$] and context [high versus low predictability, $F(1,59)=9.15$, $p=0.004$]. The two-way interaction was not significant. This general pattern of increased duration for clear versus plain speech is typical of this speaking style shift and simply indicates that, as expected, the talker in this study produced a large plain versus clear speaking style difference. As shown in Table III, the increase in word duration from plain to clear speech was 154 ms (33.6%) and 142 ms (31.4%) in the low and high predictability contexts, respectively. The small but reliable decrease in duration for high versus low predictability final words is also consistent with previous work indicating some degree of probability-dependent hypo-articulation (e.g., Scarborough, 2006) such that a given word in a highly predictable context is generally reduced relative to its occurrence in a less predictable context. In summary, this comparison of final word durations across the four sentence types established that the talker produced a large and consistent increase in word duration for clear relative to plain speech, and a small but consistent decrease in duration for words in high versus low predictability contexts.

B. Final word recognition accuracy

Figure 1 and Table IV show the final word recognition accuracy scores for the native and non-native listeners (left

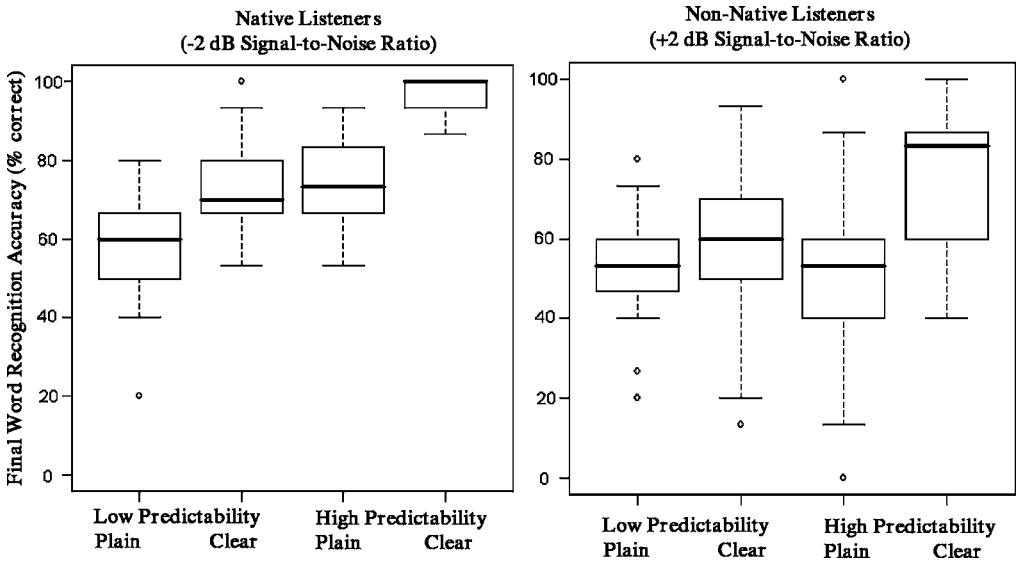


FIG. 1. Final word recognition accuracy (in % correct) for the native (left panel) and non-native listeners (right panel) in both styles of speech (plain and clear) and in both high and low predictability contexts. Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range shown in the boxes.

TABLE IV. Final word recognition accuracy in percent correct for each of the four sentence types. [Note that all statistics reported in the text were conducted on transformed data along the RAU (rationalized arcsine units) scale of the percent correct values reported here].

Listener	Native				Non-native			
Style	Plain		Clear		Plain		Clear	
Context	High	Low	High	Low	High	Low	High	Low
Mean	77	58	98	73	50	51	77	60
Median	73	60	100	70	53	53	83	60
Std. Dev.	10	12	4	11	18	13	17	18
Std. error	1.7	1.9	0.65	1.9	3.1	2.1	2.8	3.0
Minimum	53	20	87	53	0	20	40	13
Maximum	93	80	100	100	100	80	100	93

panel and right panel, respectively) in both speaking styles and with both types of sentences. Recall that the native and non-native listeners were presented with the sentences mixed with noise at a −2 and +2 dB signal-to-noise ratio, respectively. This difference in signal-to-noise ratio was introduced into the procedural design as an attempt to equate the levels of performance across the native and non-native listener groups in the "baseline" condition of low predictability sentences in plain speech. As shown in Fig. 1 and Table IV, the average native and non-native listener final word recognition accuracy scores in the plain speech low predictability context were 58% correct and 51% correct, respectively. While the difference between these scores is significant [unpaired $t(70)=2.2$, $p=0.03$], we have assumed for the purposes of the subsequent analyses that they are close enough to indicate that the two groups of listeners were operating in approximately the same range of accuracy scores and that any observed differences in the ability to take advantage of semantic-contextual cues or to benefit from clear speech enhancements were due to differences in listener-related, native versus non-native language processing strategies rather than in task-related factors that may have differed across the two listener groups.

The overall pattern of word recognition accuracies showed that, relative to the baseline condition of low predictability sentence contexts in plain speech, non-native listener final word recognition accuracy improved by 9 percentage points (from 51% to 60% correct) when presented with clear speech. In contrast, the same non-native listeners showed no benefit for final word recognition from low to high predictability sentence contexts (51% to 50% correct) in the plain speech style. However, they showed a substantial improvement of 26 percentage points from the baseline condition (low predictability sentences in plain speech) when both top-down, semantic-contextual cues and bottom-up, acoustic-phonetic enhancements were available (from 51% correct to 77% correct).

The native listeners in this study benefited from both sources of enhancement whether presented singly or in combination, relative to the baseline condition of low predictability sentence contexts in plain speech. Native listener final word recognition accuracy improved by 15 percentage points (from 58% to 73% correct) when presented with clear speech, by 19 percentage points (from 58% to 77% correct) when presented with high predictability sentences, and by 40 percentage points (from 58% to 98% correct) when presented with high predictability sentences in clear speech.

Separate two-way ANOVAs with style and context as within-subjects factors for the native and non-native listener groups were conducted on the RAU transformed data. For the native listeners, both main effects and the two-way style-context interaction were highly significant at the $p<0.0001$ level [style: $F(1,70)=161.59$, context: $F(1,70)=223.144$, style-context: $F(1,70)=34.04$]. Post hoc comparisons showed significant differences (at the $p<0.0001$ level) for all of the pair-wise comparisons except for low context clear speech versus high context plain speech. This pattern of results indicates that the native listeners derived a significant final word recognition benefit from both semantic-contextual information and acoustic-phonetic enhancements and that these two sources of intelligibility enhancement worked together and were mutually enhancing in the high predictability, clear speech condition.

For the non-native listeners, both main effects and the two-way style-context interaction were highly significant at the $p<0.0001$ level [style: $F(1,70)=60.22$, context: $F(1,70)=27.35$, style-context: $F(1,70)=19.48$]. Post hoc comparisons showed significant differences at the $p<0.0001$ level for all of the pair-wise comparisons except for two cases: The difference between plain and clear speech in the low context condition was significant at the $p<0.001$ level and there was no difference between the high and low context conditions in plain speech. Thus, for these non-native listeners, final word recognition accuracy generally improved from plain to clear speaking styles in both high and low predictability sentence contexts; however, these non-native listeners only benefited from a highly predictive context in the clear speaking style. The lack of a context effect in plain speech is consistent with the finding of Mayo *et al.* (1997) that highly proficient, late bilinguals benefited less from contextual information than native listeners and highly proficient, early bilinguals. In the present study, we extended this finding by demonstrating that non-native speech recognition can be improved by a combination of semantic-contextual and acoustic-phonetic enhancements as shown by the boost in performance in the high predictability clear speech condition relative to the other three conditions.

A major difference between the present study and the Mayo *et al.* (1997) study is that in the present study we did not directly compare performance across groups of non-native listeners with different levels of English proficiency or with different ages of English acquisition onset. Instead, the primary analyses of the present study treated all of the non-native listeners as members of a single, broadly-defined group. Nevertheless, in order to gain some insight into the role of proficiency in determining access to acoustic-phonetic and contextual enhancements for word recognition, we conducted some additional correlational analyses. As shown in Table I, 34 of the 36 non-native listeners reported SPEAK scores. The range of scores on this test was wide enough to permit a correlational analysis between English proficiency (as reflected by these scores) and final word recognition in the present speech-in-noise perception test. For these listeners, SPEAK score was positively correlated with the average final word recognition score, i.e., the overall score averaged across speaking styles and sentence predictability contexts (Pearson correlation=0.722, $p<0.0001$) and with the average high-low predictability difference score, i.e., the size of the context effect averaged across both speaking styles (Pearson correlation=0.346, $p<0.05$). In contrast, the size of the clear speech effect (i.e., the clear-plain speech difference score averaged across predictability contexts) was not significantly correlated with SPEAK score. This pattern of correlations indicates that, while the clear speech effect apparently did not depend strongly on overall proficiency within the range represented by these non-native listeners, the ability to take advantage of higher-level, semantic-contextual cues did improve with increasing general English language proficiency.

IV. GENERAL DISCUSSION

The overall purpose of the present study was to assess whether non-native listener speech-in-noise perception could be improved by the availability of both top-down, semantic-contextual cues and bottom-up, acoustic-phonetic enhancements. The data complement data from previous research by establishing that, if given sufficiently rich information by means of a clear speaking style on the part of the talker, non-native listeners can indeed enhance their word recognition accuracy by taking advantage of sentence-level contextual information. This finding is consistent with the hypothesis that native and non-native listeners are both able to use contextual information to facilitate word recognition provided that the contextual information is well specified in the signal. The data establish further that naturally produced clear speech is an effective means of enhancing access to signal-dependent information for both native and non-native listeners thereby allowing the beneficial effects of contextual information to reveal themselves. A noteworthy implication of this finding is that, while listeners may have to turn up the volume on their radios as they switch from listening in their native language to listening in a non-native language, they may also be able derive dramatic benefit (at all levels of processing) from any signal clarity enhancing device.

With regard to the benefit offered by clear speech, a bottom-up acoustic-phonetic enhancement, the present data are consistent with the finding reported in Bradlow and Bent (2002) that the non-native listener average clear speech benefit was substantially smaller in magnitude than the average native listener clear speech benefit. In the present study, the clear speech benefits in the low predictability context for the native and non-native listeners were 15 and 9 percentage points, respectively. (The data in the high predictability context did not provide for a valid comparison of the magnitude of the clear speech benefit across listener groups since the native listeners reached ceiling levels of performance). In the earlier study, we interpreted this diminished clear speech benefit for non-native listeners as reflecting their reduced experience with the sound structure of the target language relative to native listeners. We reasoned that non-native listeners may not be sensitive to some of the dimensions of contrast that native talkers spontaneously enhance in clear speech production due to their lack of extensive experience with the full range of acoustic-phonetic cues for many of the target language contrasts. Other work has shown that, in addition to language-general features such as a decreased speaking rate and an expanded pitch range, clear speech production involves the enhancement of the acoustic-phonetic distance between phonologically contrastive categories (e.g., Ferguson and Kewley-Port, 2002; Krause and Braida, 2004, Picheny *et al.*, 1986; Smiljanic and Bradlow, 2005, 2007). Therefore, reduced sensitivity to any or all of the language-specific acoustic-phonetic dimensions of contrast and clear speech enhancement would yield a diminished clear speech benefit for non-native listeners. This may appear somewhat surprising given that clear speech production was elicited in our studies by instructing the talkers to speak clearly for the sake of listeners with either a hearing impairment or from a different native language background. However, as discussed further in Bradlow and Bent (2002), the limits of clear speech as a means of enhancing non-native speech perception likely reflect the "mistuning" that characterizes spoken language communication between native and non-native speakers.

A limitation of the Bradlow and Bent (2002) study was that the materials were all meaningful sentences and the listener's task was to write down the full sentences, which were then scored on the basis of a keyword correct count. Thus, since target word predictability was not controlled in the materials of that study, the relatively small clear speech effect for the non-native listeners may have been due (partially or even entirely) to their reduced ability to take advantage of contextual information available in the sentences rather than in their reduced ability to take advantage of the acoustic-phonetic modifications of English clear speech. By directly manipulating final word predictability, the present study addressed this limitation and confirmed that non-native listeners derive a significant, though relatively small benefit from the acoustic-phonetic enhancements of clear speech independently of their reduced ability to take advantage of higher-level semantic-contextual information provided in a sen-

tence. It is thus likely that this smaller non-native clear speech benefit is indeed due to reduced experience with sound structure of the target language.

Two recent studies provide some information regarding the interaction of acoustic- and semantic-level information during native listener spoken language processing that have some bearing on the comparison between native and non-native listeners in the present study. First, in a study of lexical access in English, Aydelott and Bates (2004) showed different reaction time patterns in a lexical decision task depending on whether the speech stimuli were left unaltered or were distorted (by low-pass filtering or time compression). Specifically, they examined lexical decision reaction times to target words presented in a "congruent semantic context" (i.e., in a high predictability sentence context such as "On a windy day it's fun to go out and fly a" for the target word "kite"), in a "neutral semantic context" (i.e., in a low predictability sentence context such as "Its name is" for the target word "kite"), or in an "incongruent semantic context" (i.e., in a sentence context such as "On a windy day it's fun to go out and fly a" for the target word "yert"). The key finding for our purposes was that, when presented with unaltered speech stimuli, the participants showed the expected pattern of increasing reaction times from the congruent to the neutral to the incongruent semantic contexts. In contrast, when presented with distorted speech stimuli, some of these reaction time differences due to semantic context differences were neutralized, indicating that variation in signal clarity can affect the operation of "normal" facilitatory and inhibitory processes at the semantic level. Similarly, in a study of English word segmentation from connected speech, Mattys *et al.* (2005) demonstrated that higher-level, knowledge-driven lexical information interacts with lower-level, signal-derived, sublexical information according to a hierarchical organization of cues with descending weight assignments from lexical to segmental to prosodic cues. Of particular interest with regard to the present study was the finding that these cue weightings were effectively reversed under conditions of signal distortion due to the presence of background (white) noise. Specifically, as signal quality decreased (due to decreasing signal-to-noise ratios), thereby rendering any available contextual information increasingly inaccessible, listeners were forced to rely more heavily on lower-level signal-derived information than on contextual information for word segmentation. Conversely, when presented with intact speech (no added noise) from which contextual information was easily accessible, listeners took advantage of this higher-level information and relied less on lower-level acoustic-phonetic word boundary cues.

While these studies differed from the present study in numerous ways, perhaps most notably by the fact that they examined the effects of acoustic distortion rather than acoustic enhancement, they both demonstrate that, even for native listeners, speech processing strategies that involve higher-level semantic-contextual information can be more or less effective depending on access to the speech signal at the perceptual level. This situation is, of course, highly analogous to the pattern of findings for the non-native listeners in the present study. Like the native listeners in the priming study of Aydelott and Bates (2004) and in the word segmentation study of Mattys *et al.* (2005), the non-native listeners in the present word recognition study were only able to make use of higher-level contextual cues when the lower-level acoustic-phonetic cues were sufficiently clear that the preceding contextual information was indeed easily accessible. When presented with a sufficiently clear signal, the native listeners in the previous priming and segmentation studies and the non-native listeners in the present study all showed processing strategies that involved taking advantage of any available higher-level contextual information.

It is important to note that, in the present study, the signal clarity required to make effective use of contextual information was quite large for the non-native listeners compared to the native listeners, and required both a more favorable signal-to-noise ratio (recall that the non-native and native listeners were presented with stimuli at +2 and −2 dB signal-to-noise ratios, respectively) and a clear speaking style. It remains for future research to determine whether the perceptual patterns indicated in the present study will be obtained with more systematic control over listener proficiency in the target language and across a wider range of signal-to-noise ratios.

When viewed in relation to the work with native listeners presented with degraded signals (Aydelott and Bates, 2004; Mattys *et al.*, 2005), the difference in word recognition patterns across the native and non-native listeners in the present study can be described as a difference in the signal clarity required for semantic-contextual information to be effective, rather than as an inability of non-native listeners to take advantage of contextual information. Thus, as suggested in Sec. I, non-native listener speech-in-noise perception is not necessarily doomed by limited recovery resources at higher levels of processing. Instead, higher level support for information extracted from the acoustic-phonetic level is available to non-native listeners (just as it is for native listeners) albeit in a less efficient mode of operation. An open issue that remains to be investigated further is with regard to the origin and nature of the mechanism that underlies this relative inefficiency of non-native listeners. In all likelihood, the source of this feature of non-native speech processing is multifaceted including factors related to experience-dependent "mistuning" to all levels of linguistic structure of the target language (ranging from the subsegmental, segmental, and prosodic levels of sound structure to the more abstract syntactic, semantic, and pragmatic levels) as well as factors relating to the cognitive demands of managing two (or more) languages. An important challenge for future research is to identify these factors and then to ultimately propose means of overcoming these deficits by either human or technological enhancements. As demonstrated in the present study, clear speech may be a promising avenue to follow toward this goal.

ACKNOWLEDGMENTS

This research was supported by Grant No. NIH-R01-DC005794 from NIH-NIDCD. The authors gratefully ac-

knowledge the assistance of Judy Song and Cynthia Clopper with subject running.

APPENDIX: HIGH AND LOW PREDICTABILITY SENTENCES

High predictability sentences

1. The meat from a pig is called pork.
2. For dessert he had apple pie.
3. Sugar tastes very sweet.
4. The color of a lemon is yellow.
5. My clock was wrong so I got to school late.
6. In spring, the plants are full of green leaves.
7. A bicycle has two wheels.
8. She made the bed with clean sheets.
9. The sport shirt has short sleeves.
10. He washed his hands with soap and water.
11. The child dropped the dish and it broke.
12. The bread was made from whole wheat.
13. The opposite of hot is cold.
14. A wristwatch is used to tell the time.
15. The war plane dropped a bomb.
16. She cut the cake with a knife.
17. A chair has four legs.
18. Cut the meat into small pieces.
19. The team was trained by their coach.
20. The lady wears earrings in her ears.
21. People wear shoes on their feet.
22. When sheep graze in a field, they eat grass.
23. A rose is a type of flower.
24. Football is a dangerous sport.
25. The heavy rains caused a flood.
26. Bob wore a watch on his wrist.
27. Monday is the first day of the week.
28. The pan that was just in the oven is very hot.
29. Rain falls from clouds in the sky.
30. The boy laughed because the joke was very funny.
31. To cool her drink, she added a few cubes of ice.
32. A quarter is worth twenty-five cents.
33. An orange is a type of fruit.
34. People wear scarves around their necks.
35. I wrote my name on a piece of paper.
36. For your birthday I baked a cake.
37. Birds build their nests in trees.
38. My parents, sister and I are a family.
39. The good boy is helping his mother and father.
40. People wear gloves on their hands.
41. A book tells a story.
42. A pigeon is a kind of bird.
43. The sick woman went to see a doctor.
44. The lady uses a hairbrush to brush her hair.
45. At breakfast he drank some orange juice.
46. Last night, they had beef for dinner.
47. A racecar can go very fast.
48. Many people like to start the day with a cup of coffee.
49. He brought the book to school from home.
50. I wear my hat on my head.
51. Red and green are colors.
52. The stars come out at night.
53. February has twenty-eight days.
54. The picture is hung high on the bedroom wall.
55. We heard the ticking of the clock.
56. She laid the meal on the table.
57. She looked at herself in her mirror.
58. Elephants are big animals.
59. After my bath, I dried off with a towel.
60. In the morning it gets light, and in the evening it gets dark.

Low predictability sentences

1. Dad looked at the pork.
2. Mom talked about the pie.
3. We think that it is sweet.
4. Mom thinks that it is yellow.
5. He thinks that it is late.
6. She talked about the leaves.
7. He read about the wheels.
8. Dad talked about the sheets.
9. He looked at the sleeves.
10. We talked about the water.
11. We heard that it broke.
12. Dad pointed at the wheat.
13. She thinks that it is cold.
14. This is her favorite time.
15. Dad talked about the bomb.
16. Mom read about the knife.
17. She looked at her legs.
18. There are many pieces.
19. We read about the coach.
20. She pointed at his ears.
21. Mom looked at her feet.
22. Dad pointed at the grass.
23. She read about the flower.
24. This is her favorite sport.
25. He read about the flood.
26. He looked at her wrist.
27. This is her favorite week.
28. Mom thinks that it is hot.
29. Dad read about the sky.
30. Dad thinks that it is funny.
31. He talked about the ice.
32. He pointed at the cents.
33. He pointed at the fruit.
34. She talked about their necks.
35. We talked about the paper.
36. This is her favorite cake.
37. He read about the trees.
38. We read about the family.
39. Mom pointed at his father.
40. She looked at her hands.
41. We looked at the story.
42. We pointed at the bird.
43. Mom talked about the doctor.
44. He pointed at his hair.
45. Mom looked at the juice.
46. He talked about the dinner.
47. She thinks that it is fast.

48. Mom pointed at the coffee.
49. She pointed at the home.
50. She pointed at her head.
51. Mom read about the colors.
52. This is her favorite night.
53. There are many days.
54. We pointed at the wall.
55. She looked at the clock.
56. Dad read about the table.
57. We looked at the mirror.
58. He pointed at the animals.
59. Dad looked at the towel.
60. Dad thinks that it is dark.

Aydelott, J., and Bates, E. (**2004**). "Effects of acoustic distortion and semantic context on lexical access," Lang. Cognit. Processes **19**, 29–56.

Aylett, M., and Turk, A. (**2004**). "The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang Speech **47**, 31–56.

Bench, J., and J. Bamford, eds. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for learning-impaired children: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Cutler, A., Webber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Fallon, M., Trehub, S. E., and Schneider, B. A. (**2002**). "Children's use of semantic cues in degraded listening environments," J. Acoust. Soc. Am. **111**, 2242–2249.

Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Jiang, J., Chen, M., and Alwan, A. (**2006**). "On the perception of voicing in syllable-inital plosives in noise," J. Acoust. Soc. Am. **119**, 1092–1105.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistics Structure*, edited by J. Bybee and P. Hopper (Benjamins, Amsterdam, The Netherlands), pp. 229–254.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. **134**, 477–500.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Meador, D., Flege, J. E., and MacKay, I. R. (**2000**). "Factors affecting the recognition of words in a second language," Bilingualism: Lang. Cognit. **3**, 55–67.

Munro, M., and Derwing, T. (**1995**). "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners," Lang. Learn. **45**, 73–97.

Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.

Munson, B. (**2007**). "Lexical access, lexical representation and vowel production," *in Papers in Laboratory Phonology IX*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, Berlin).

Nábělek, A. K., and Donohue, A. M. (**1984**). "Perception of consonants in reverberation by native and non-native listeners," J. Acoust. Soc. Am. **75**, 632–634.

Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.

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**, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech," J. Acoust. Soc. Am. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). "Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech Hear. Res. **32**, 600–603.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., and Abrams, H. B. (**2006**). "Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing," Appl. Psycholinguist. **27**, 465–485.

Scarborough, R. (**2006**). "Lexical and contextual predictability: Confluent effects on the production of vowels," presentation at the Tenth Conference on Laboratory Phonology, Paris, France.

Smiljanic, R., and Bradlow, A. R. (**2005**). "Production and perception of clear speech in Croatian and English," J. Acoust. Soc. Am. **118**, 1677–1688.

Smiljanic, R., and Bradlow, A. R. (**2007**). "Stability of temporal contrasts across speaking styles in English and Croatian," J. Phonetics.

Sommers, M. S. (**1996**). "The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition," Psychol. Aging **11**, 333–341.

Sommers, M. S. (**1997**). "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," J. Acoust. Soc. Am. **101**, 2279–2288.

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

Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing, IV. Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.

Uchanski, R. M. (**2005**). "Clear speech," in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Cambridge, MA).

Wright, R. (**2002**). "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 26–50.

Development of the Cantonese speech intelligibility index[a)]

Lena L. N. Wong,[b)] Amy H. S. Ho,[c)] and Elizabeth W. W. Chua[d)]
Division of Speech & Hearing Sciences, University of Hong Kong, Hong Kong, China

Sigfrid D. Soli
House Ear Institute, Los Angeles, California 90057

(Received 4 April 2006; revised 8 December 2006; accepted 11 December 2006)

A Speech Intelligibility Index (SII) for the sentences in the Cantonese version of the Hearing In Noise Test (CHINT) was derived using conventional procedures described previously in studies such as Studebaker and Sherbecoe [J. Speech Hear. Res. **34**, 427–438 (1991)]. Two studies were conducted to determine the signal-to-noise ratios and high- and low-pass filtering conditions that should be used and to measure speech intelligibility in these conditions. Normal hearing subjects listened to the sentences presented in speech-spectrum shaped noise. Compared to other English speech assessment materials such as the English Hearing In Noise Test [Nilsson *et al.*, J. Acoust. Soc. Am. **95**, 1085–1099 (1994)], the frequency importance function of the CHINT suggests that low-frequency information is more important for Cantonese speech understanding. The difference in frequency importance weight in Chinese, compared to English, was attributed to the redundancy of test material, tonal nature of the Cantonese language, or a combination of these factors.
 [DOI: 10.1121/1.2431338]

PACS number(s): 43.71.Gv [ARB] Pages: 2350–2361

I. INTRODUCTION

A. Background

The Articulation Index (AI) or its revised appellation, Speech Intelligibility Index (SII), is a quantitative measure that accounts for the contribution of audible speech cues in given frequency bands to speech intelligibility (Amlani *et al.*, 2002). It is a useful tool for estimating speech understanding ability under specified listening situations. The AI has been suggested for clinical applications such as prediction of speech recognition performance with various configurations of hearing loss (Macrae and Brigden, 1973; Pavlovic, 1984; Kamm *et al.*, 1985; Killion and Christensen, 1998), estimation of unaided and aided speech intelligibility to determine the potential benefits of hearing aids (Mueller and Killion, 1990; Killion and Christensen, 1998; Stelmachowicz *et al.*, 2002), and prescription of hearing aid gain (Rankovic, 1991). Amendments to the original calculations of AI were made for over a decade before the ANSI-S3.5 (1969) standard was adopted. Since then, efforts were made to simplify the calculation of AI for clinical applications (e.g., Pavlovic, 1984; Eisenberg *et al.*, 1998). The term, Speech Intelligibility Index (SII) was later adopted in the ANSI-S3.5 (1997) standard to account for spread of masking and level distortion effects (Amlani *et al.*, 2002).

To establish the SII, it is necessary to gain a thorough understanding of the frequency-importance function (FIF) of a specific test material because the relative importance of various frequency bands to speech intelligibility is a key component of the basic SII equation:

$$\sum_{i=1}^{n} IiAi, \tag{1}$$

where Ii is the importance of a frequency band (i) and is expressed as a weighted factor from 0.0 to 1.0; Ai is the audibility function, representing the amount of speech energy available in the ith frequency band that contributes to the overall intelligibility (French and Steinberg, 1947; Amlani *et al.*, 2002). It is assumed that the speech signals in the adjoining frequency bands that comprise the audible spectrum will *independently* contribute to the articulation score, and speech intelligibility is an *additive* measure of *weighted importance* contributed by different frequency regions (Rankovic, 1995).

The dynamic range (DR) of the long-term average speech spectrum (LTASS), which Byrne *et al.* (1994) found is similar across languages, affects the calculation of the audibility function. Conventionally, the effective DR is assumed to be 30 dB at all frequency bands for English materials (e.g., ANSI-S3.5, 1969; Studebaker and Sherbecoe, 1991; Eisenberg *et al.*, 1998). Studebaker *et al.* (1999) argued that there is "credible evidence" for a larger value, and they experimentally proved that a DR of 40 dB yielded better prediction of speech recognition using NU-6 (Tillman and Carhart, 1966) on normal hearing and hearing-impaired participants under different listening conditions.

The SII can be used to predict speech intelligibility via a transfer function (S) such as the one derived by Fletcher and Galt (1950):

[a)]Portions of this work were presented in a paper "Cantonese Speech Intelligibility Index," Proceedings of International Congress of Audiology, Phoenix, Arizona, September 2004.
[b)]Electronic mail: LLNWONG@hku.hk
[c)]Currently associated with St. Teresa's Hospital Hearing and Speech Centre, Hong Kong.
[d)]Currently associated with Starkey (HK) Hearing and Speech Centre Ltd.

TABLE I. Crossover frequencies of various speech materials.

Study	Speech stimulus	Crossover frequency
Studebaker *et al.* (1987)	Continuous discourse	1189 Hz
Studebaker and Sherbecoe (1991)	W-22	1314 Hz
Eisenberg *et al.* (1998)	HINT sentences	1550 Hz
Sherbecoe and Studebaker (2002)	Connected Speech Test	1599 Hz
ANSI (S3.5-1969)	Nonsense syllables	1660 Hz
French and Steinberg (1947)	Nonsense syllables	About 1900 Hz

$$S = (1 - 10^{-AP/Q})^N, \quad (2)$$

where S is the percent correct intelligibility score, A is the SII value, P stands for a proficiency factor that accounts for talker's and listener's competence and practice effect, and both Q and N are fitting constants depending on the speech stimulus' characteristics (Fletcher and Galt, 1950). More specifically, Q is a correction factor "to compensate for changes in proficiency" to the test stimuli in an experiment; N represents "the number of independent sounds in a test item" or a constant "that controls the shape of the line [S]" (Studebaker and Sherbecoe, 1991, pp. 431 and 433).

B. SII for specific speech materials

Studebaker and Sherbecoe (1993) reported that FIFs vary with speech stimuli so that given the same SII, predicted speech intelligibility varies with speech materials. The original AI calculation was based on CVC nonsense syllables (French and Steinberg, 1947). Other types of speech test materials have been used in subsequent research. These include the Central Institute for the Deaf (CID) W-22 word lists (Studebaker and Sherbecoe, 1991), NU-6 word test (Studebaker *et al.*, 1993), Hearing In Noise Test (HINT) sentence materials (Eisenberg *et al.*, 1998), Consonant-vowel Nucleus-Consonant (CNC) monosyllabic word test (Henry *et al.*, 1998), and Connected Speech Test (CST) passages (Sherbecoe and Studebaker, 2002). DePaolis *et al.* (1996) found statistically different one-third octave band FIFs for PB-50 monosyllabic words, the SPIN test and continuous discourse. Distinct crossover frequencies, i.e., the frequency that divides a speech spectrum into two equally important parts, varied from 1189 to 1900 Hz for various materials (see Table I). With the exception of the W-22 word lists, crossover frequencies shift to lower values as the redundancy of the speech materials increases (Studebaker *et al.*, 1987; Studebaker and Sherbecoe, 1991)—continuous discourse has the lowest values and nonsense syllables have the highest values. The crossover frequency may differ across languages. For example, while French and English did not show much difference in crossover frequencies (about 1500 Hz), Finnish disyllabic words had a significantly lower crossover frequency at about 1000 Hz (Studebaker and Sherbecoe, 1993). The FIF or crossover frequency has never been established for tonal languages such as Cantonese.

C. Cantonese

Cantonese is a tonal language spoken by more than 16 million people in the world. Cantonese is a regional dialect in South-Eastern China (Ramsey, 1987) and one of the main dialects in China (Li, 1989). It is commonly spoken among Chinese immigrants in North America, South Asia, Australia, and Great Britain (Lau and So, 1988; Matthews and Yip, 1994). Among Chinese dialects, its influence is second to that of Mandarin (Matthews and Yip, 1994).

Cantonese morphemes are monosyllabic and monosyllables are combined to form polysyllabic words. Cantonese syllables take the form of optional initial consonant, mandatory vowel, and optional final consonant or (C)V(C). Cantonese has the same long-term average speech spectrum (LTASS) as many other languages including English (Byrne *et al.*, 1994), but Cantonese phonology is very different from English phonology (So and Dodd, 1995). For example, Cantonese speakers would be concerned with discrimination of aspirated and unaspirated consonants and not between voiced and voiceless consonants. Cantonese has fewer consonants and more vowels than English, and tones carry lexical meaning (Dodd and So, 1994). There are nine lexical tones (Browning, 1974; Fok Chan, 1974; Dodd and So, 1994), as listed in Table II. Browning (1974) suggested that the three entering tones (high, mid, and low) of Cantonese are not contrastive as their registers are comparable to tones 1 (high level), 3 (mid level), and 6 (low level). Pitch variations due to changes in fundamental frequency (F0) provide the main cues for tone perception (Fok Chan, 1974; Gandour, 1981; Cheung, 1992). Cheung (1992) found that tones are more resistant to the masking effect of noise than consonants. Thus, it is possible that low-frequency information carries more weight for Cantonese speech understanding than English. In fact, compared to English speakers with the same amount of hearing loss, Cantonese speakers with good low-frequency hearing experience less self-reported difficulty in speech understanding, despite a significant loss at higher frequency (Doyle and Wong, 1996; Doyle *et al.*, 2002; Wong *et al.*, 2004).

TABLE II. Description and examples of each Cantonese tone.

Number	Classification	Example	Transcription
1	High level	Poem	si_1
2	High rising	History	si_2
3	Mid level	Examination	si_3
4	Low falling	Time	si_4
5	Low rising	Market	si_5
6	Low level	Matter	si_6
7	High entering	Color	$sɪk_7$
8	Mid entering	Kiss	$sɪk_8$
9	Low entering	Eat	$sɪk_9$

D. Aim of the study

This study was aimed at deriving a Speech Intelligibility Index for Cantonese (SIIC) using the materials from the Cantonese version of the Hearing in Noise Test (CHINT) (Wong and Soli, 2005). The CHINT is the only standardized Cantonese sentence speech reception test. Deriving a SII based on the CHINT would result in a better understanding of Cantonese speech perception. In particular, cochlear implant coding strategies are based on work to optimize speech understanding in native English speakers, but Cantonese users fail to recognize tones (Ciocca *et al.*, 2002; Wong and Wong, 2004). It is hoped that knowledge of Cantonese SII may result in a better understanding of cochlear implant strategies to help preserve tonal information. How hearing aids should be best prescribed for Cantonese speakers also requires a thorough understanding of how audibility at various frequencies contributes to intelligibility.

Procedures described by Studebaker and Sherbecoe (1991) were used as a basis for Cantonese SII derivation. With Cantonese being a tonal language, it was expected that the crossover frequency for a given type of Cantonese material would be lower than the English equivalent and the FIF would be different from English or French materials. As the effective DR for CHINT has not been determined, results based on the work by Byrne and colleagues (1994) were used. That is, the DR of CHINT was assumed to be 30 dB, but a 40-dB DR was also evaluated.

II. METHOD

A. Participants

Six normal-hearing native Cantonese speakers participated in the pilot study. Seventy-eight (34 male, 44 female) other young normal-hearing native Cantonese speakers participated in the actual study. As participants were recruited in Hong Kong where some individuals are exposed to two dialects (e.g., Cantonese and Mandarin) since birth, first language was difficult to determine. Therefore, participants speaking Cantonese as their primary language were recruited. None of the participants spoke Cantonese with a dialectal accent. Mean age of participants in the actual experiment was 23 years for male (s.d. 4.5) and 22 years for female (s.d. 2.5), with a range from 18 to 34 years. All participants had bilateral hearing thresholds of 20 dB HL or better at the octave frequencies from 250 to 8000 Hz. In the actual experiment, participants' pure-tone hearing thresholds averaged at 500, 1000, and 2000 Hz in the right ear was 9.9 dB HL (s.d. 3.8) and in the left ear was 6.9 dB HL (s.d. 4.3). None of the participants reported histories of noise exposure or middle ear pathology. All of them had normal middle ear function confirmed by tympanometry prior to the experiment. All participants were paid to take part in the study.

B. Materials

Sentences from the CHINT (Wong and Soli, 2005) were used in the present study because it is the only well-standardized material for assessing Cantonese speech intelligibility. The CHINT comprises 24 sets of 10 sentences each, with sentences in each set balanced for the level of difficulty and phonemic characteristics. The Cantonese HINT sentences have 10 syllables represented by 10 Chinese characters; this contrasts with the English HINT sentences that contain four to seven syllables (Nilsson *et al.*, 1994). The CHINT can be used to assess speech intelligibility in quiet and in noise with noise simulated to originate from 0°, 90°, and 270° azimuths. In this study, speech and noise were presented in noise only from 0° azimuth. The noise used was matched to the long-term average speech spectrum of the talker.

C. Equipment

The CHINT sentences and the speech-spectrum shaped noise were presented via the Hearing In Noise Test (HINT) program (version 5.0.3) using a SoundBlaster soundcard. Both speech signal and speech-spectrum shaped noise were mixed before they were delivered to a Tucker-Davis Technologies (TDT) System 3 digital filter. The filter was controlled by a computer program, Realtime Processor Visual Design Studio (RPvds) (version 4.0) and provided a rejection slope of 96 dB/octave at the desired cutoff frequencies. The filtered signals were routed to a GSI 16 audiometer and presented diotically to the participants using TDH-50P headphones. The output of headphones was calibrated to 65 dB A in a 6-cc coupler using the speech-spectrum shaped noise low-pass filtered at 12 000 Hz.

D. Procedures

For the wide-band condition, the noise was fixed at 65 dB A, and the level of speech signal was varied according to the desired signal-to-noise ratio (SNR) in each test condition. Prior to testing, participants listened to two practice lists, one presented in quiet and another in noise to familiarize them with the stimuli and test procedures. Reception thresholds of sentences (RTSs) were obtained adaptively (Nilsson *et al.*, 1994) in quiet with test stimuli low-pass (LP) filtered at 12 000 Hz. Individual RTSs served as reference levels for obtaining speech intelligibility scores in the filtering conditions.

1. Pilot study for selecting filtering conditions

A pilot study was conducted to determine the signal-to-noise ratios (SNRs) and the cutoff frequencies that should be used in the actual experiment. Six participants took part in the pilot study. RTSs were obtained in noise and served as the reference level for determining the speech level at which percent correct intelligibility was measured in various filtering/SNR conditions. Participants were instructed to repeat as much of each sentence as possible. According to the HINT protocol, only small variations in response that did not change the meaning of the sentences were allowed (e.g., mommy instead of mama).

Percent correct intelligibility was then obtained in various filtering/SNR conditions. As there were 10 sentences in each list, every sentence repeated correctly contributed to 10% of the score. To determine the filtering conditions to be

TABLE III. Mean percent speech recognition scores in various filtering/SNR conditions used in the actual experiment. The blank cells represent conditions that were not evaluated in the study.

	SNR (refer to individual RTSs)							
Cut-off frequency	−4	−2	−1	0	2	4	6	8
Low-pass filtered								
500						0.6	1.3	1.9
650						10.0	18.0	37.5
800					7.3	23.1	28.7	50.0
1100				0.6	9.3	18.8	46.9	64.7
1400		0.6	4.4	5.0	15.0	38.1	57.5	67.5
1700	1.3	6.7	7.3	13.3	43.1	56.9	81.3	82.0
3500	3.8	12.7	23.1	42.5	50.6	84.4	91.3	95.0
5000	11.1	28.9	32.2	60.0	74.4	88.9	93.3	98.3
6500	2.2	31.1	44.4	56.7	74.4	91.1	94.4	96.7
8000	11.1	30.0	48.9	58.9	74.4	77.8	97.8	100.0
12000	15.6	29.4	43.8	56.3	74.4	85.6	96.3	96.3
High-pass filtered								
200	7.5	26.9	41.9	54.4	88.7	84.4	92.5	96.9
500	8.7	21.9	34.4	39.4	68.1	75.6	89.4	90.0
800	0.6	1.3	7.5	12.7	24.4	41.9	56.9	65.0
1100	0.6	3.1	6.3	8.7	19.4	29.4	55.6	60.0
1400		2.5	3.1	4.0	10.0	18.8	35.6	40.6
1700			0.6	3.8	2.5	3.1	6.7	12.5

used in the actual study, presentations at +7 dB SNR; and LP and high-pass (HP) filters set to 500, 800, 1100, 1400, and 1700 Hz were arbitrarily chosen. To select the SNR conditions for the actual experiment, a total of eight SNR conditions, i.e., −5, −4, −3, −1, 0, +1, +3, +5 dB SNR (with reference to individual RTSs) were used with LP filter set to 12 000 Hz cutoff. A total of 13 conditions were evaluated. Each participant took about one hour to complete the testing in the pilot study. After the pilot study, SNR and filtering conditions that would contribute important information to the study were selected. In addition, a few other SNR and filtering conditions were selected in order to yield more detailed information. The criteria used to select these conditions will be discussed in the Results section.

2. Study

A practice list was administered to obtain RTS in noise so as to familiarize participants with test stimuli and procedures. RTS was measured in noise to determine the speech level at which percent correct intelligibility was to be measured. As only 23 sentence lists were available after individual RTSs were obtained, each participant was evaluated using 22 to 23 randomly assigned conditions. This process took about one hour to complete. Mean intelligibility in each test condition was based on 16 sets of data. Based on results from the pilot study, RTS was obtained in a total of 115 filtering/SNR conditions (see Table III).

E. Data analysis

1. Pilot study to select filtering/SNR conditions

Percent intelligibility across the filtering/SNR conditions was compared. Conditions were selected or added to ensure as wide a range of scores as possible (from 0% to 100%) could be obtained. Among the conditions that yielded very similar results, only one was selected.

2. Determination of performance-intensity function

The performance-intensity (PI) function is defined as the change in intelligibility per dB change in SNR. The PI function was used to confirm whether intelligibility grows as a function of SNR in a linear relationship, with a slope of about 10% per dB change in SNR (Wong and Soli, 2005). For the present study, the PI function was estimated by using the data from the 12 000 Hz LP filtering condition at various SNRs. Intelligibility scores from 20% to 80% were used to estimate the PI function; beyond this range, plateau of scores did not allow accurate measurement.

3. Determination of crossover frequencies

The crossover frequency is defined as the frequency which divides the frequency range into two regions, each accounting for 50% of the information. To obtain crossover frequencies at each SNR, intelligibility at each LP and HP cutoff frequency was plotted. These data were examined and only the scores that contributed to the linear portion of the growth function were used to obtain two regression equations, one for the LP and another for the HP conditions; beyond this range, ceiling and floor effects might have affected the results. The intersection between the LP and HP curves at each SNR represents the crossover frequency for the CHINT materials. The crossover frequency was obtained by solving these equations. The same procedures were applied at various SNRs.

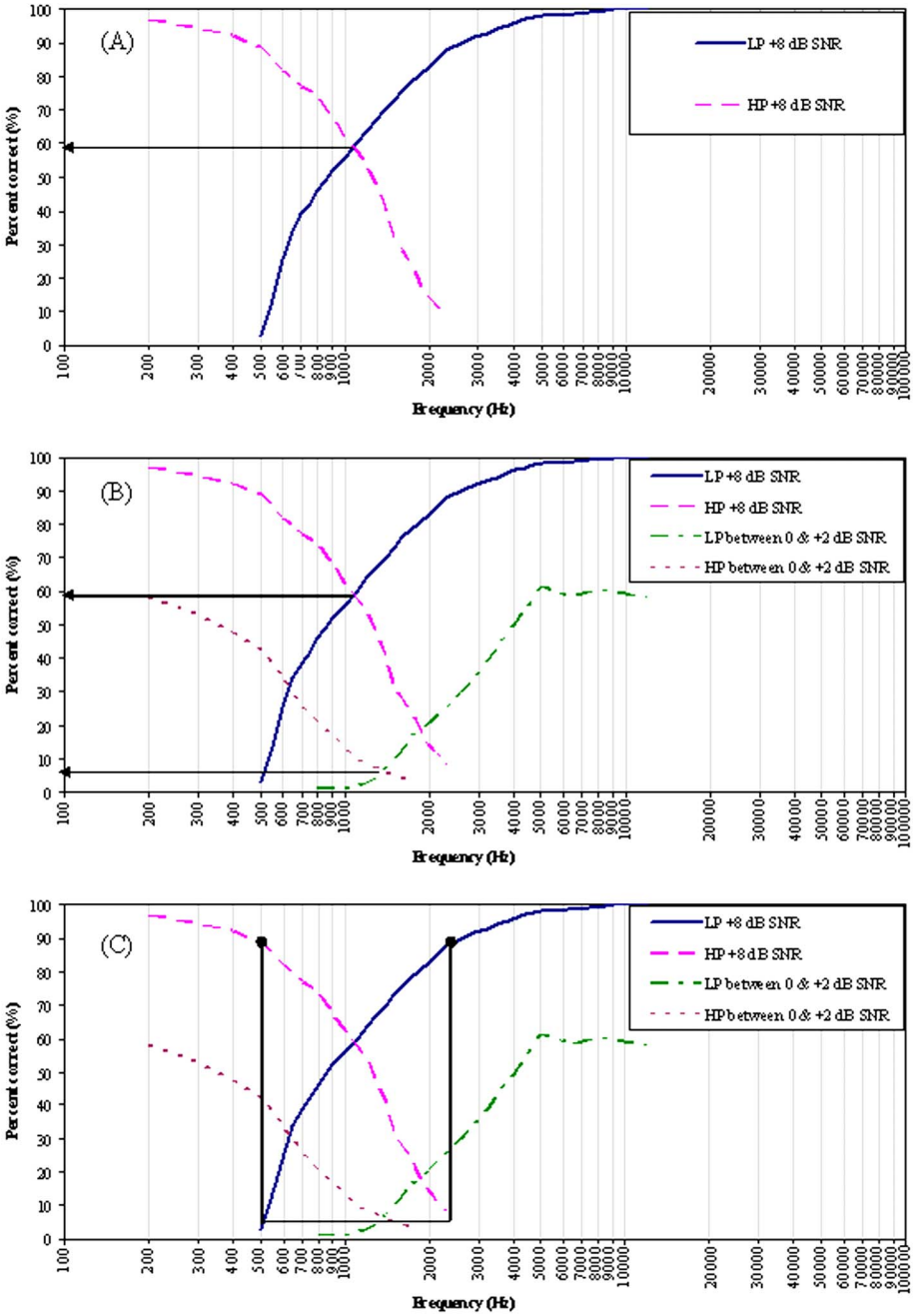


FIG. 1. The curve bisection procedure used to derive the RTF. Panel A denotes the value for 0.50 SII, panel B denotes the value for 0.25 SII, and panel C denotes the values used to derive 0.75 SII. Results from high-pass filtering conditions are represented by lines with upper ends that start from the left side of the graph; those from low-pass filtering conditions start from the right side.

4. Derivation of the relative transfer function

The relative transfer function (RTF) assumes that the maximum SII is equal to one. That is, the unfiltered condition with the highest score is assigned a SII value of 1.00 and the other conditions have SIIs relative to that value (Studebaker *et al.*, 1987; Studebaker and Sherbecoe, 1991). The curve bisection procedure described by Studebaker and Sherbecoe (1991, pp. 431 to 432) was used to derive the RTF. Briefly, percent correct scores for the LP and HP filtering conditions at the highest SNR (i.e., +8 dB SNR, with reference to individual RTSs) were first plotted as a function of filter cutoff frequency (see Fig. 1). The percent correct intelligibility corresponding to 0.5 SII was obtained using these two curves. That is, the intersection of these two curves represents 0.5 SII, because half of the total auditory area is available to the listener above this point and another half is below this point. The total area for this SNR is assumed to have an SII of 1.00 (Studebaker and Sherbecoe, 1991, p. 430). The procedures are shown in panel A of Fig. 1. The score at 0.50 SII was then used to determine the next point (i.e., the score corresponding to 0.25 SII) on the transfer function. Because there were no HP or LP curves that terminated at the score corresponding to 0.50 SII, data from the curves corresponding to 0 and 2 dB SNR were used to interpolate the data. The intersection of these two curves yielded the scores for 0.25 SII.

Scores for SII values above 0.50 were obtained by identifying points on the curves that complemented those below 0.5 SII (Studebaker and Sherbecoe, 1991, p. 431). The procedure is illustrated in panel B of Fig. 1. The 0.75 SII point was produced by extending a horizontal line for the score for 0.25 SII until it intersected the HP and LP curves for the

+8 dB SNR (with reference to individual RTSs) condition. These HP and LP curves as well as the horizontal line are shown in panel C of Fig. 1. Two vertical lines were then drawn, starting from these two intersection points, to connect to the upper ends of the LP and HP curves for the +8 dB SNR (with reference to individual RTSs) condition. The values corresponding to the top intersections of these lines and the HP and LP curves, as indicated by the two circles in panel C of Fig. 1, were then averaged to yield a final score for 0.75 SII.

The above bisection procedures were followed until a number of SII values with corresponding percent correct intelligibility were obtained. The SPSS 11.0 program was used to fit the percent correct intelligibility scores and the corresponding SII derived from the above procedures using several equations including Eq. (2). The best fit SII relative transfer function (RTF), together with its fitting constants were estimated.

5. Derivation of the frequency-importance function

The RTF was then used to derive the frequency-importance function (FIF), i.e., the relative importance of speech information contained in each frequency region defined by the area between filter cutoff frequencies (Henry *et al.*, 1998; Studebaker and Sherbecoe, 1991). The procedures described in Studebaker and Sherbecoe (1991, pp. 430 to 433) and Henry *et al.* (1998, p. 83) were followed. First, all HP and LP mean scores at each SNR condition were converted to SIIs using Eq. (3), which is a transformation of Eq. (2):

$$A = Q/P \log(1 - S^{1/N}). \tag{3}$$

The mean scores and their corresponding SIIs were substituted into Eq. (3) to obtain the fitting constants Q and N using SPSS 11.0 program. The P value was assumed to be 1.000. The HP and LP SII data were combined and averaged using the procedures set out in Studebaker and Sherbecoe (1991, pp. 430 to 432) to generate an average cumulative SII curve against the filter cutoff frequencies. Briefly, the mean SII across all SNRs for each filtering cutoff frequency was calculated. These SII values were then plotted against the filtering frequencies. The SPSS 11.0 program was used to identify the best fit curve for relating these parameters. As this graph represented the cumulative band-importance for the full range of frequencies (200 to 12 000 Hz), the contribution of each one-third octave band FIF was obtained by dividing the full range into appropriate bands, and subtracting the cumulative SII at the center frequency of the lower band from that of the higher band. Then, the relative FIF was expanded to an SII scale of 0 to 1. This was achieved by dividing every SII value by the sum of individual SIIs.

6. Derivation of the absolute transfer function

Once the FIF is determined, the slope of the RTF can be adjusted so that the best SII predicted by that function is now equal to its true absolute value (Studebaker and Sherbecoe, 1991). As the curve bisection procedure in RTF derivation assumes the best score obtained is equivalent to a perfect SII (or 1.0), adjustment to the slope of the RTF is required to obtain an absolute transfer function (ATF), which reflects the true relationship between SII and the test scores. The procedures described in Studebaker and Sherbecoe (1991, p. 433) were followed below to derive the ATF.

Equation (1) was used to identify the SIIs for the mean percent correct score in each test condition. As the SNRs were based on individual RTSs, a correction factor equivalent to the mean RTS (or 3.5 dB) was added to each condition before calculations. Using an iterative method of audibility index determination (Studebaker and Sherbecoe, 1991), the SII for each listening condition was calculated using Eq. (4):

$$\mathrm{SII} = \sum_{i=1}^{n} [(\mathrm{SNR}_{\mathrm{adjusted}} + K)/\mathrm{DR}] \times \mathrm{FIF}i, \tag{4}$$

where SNR adjusted is the SNR for each test condition adjusted by the mean RTS (or 3.5 dB), K is the assumed speech maxima above LTASS, DR is the assumed dynamic range for speech, FIFi is the FIF of frequency band i, and n is the total number of bands used in the calculation. First, mean scores between 5 and 95% were plotted against their SII values, using Eq. (2) as the fitting model which was also the best fit curve among others (e.g., linear regression analysis). As the value of K was unknown for the CHINT material, it was varied in 1 dB steps from 10 to 21 dB, and the DR was set at 30 or 40 dB to identify a combination of K and DR values that would yield the smallest mean square error. These K and DR values were based on ANSI-S3.5 (1969) standard where a 30 dB DR represents the range from +12 dB to −18 dB relative to the LTASS. This range was then modified to ±15 dB in ANSI-S3.5 (1997). In this study, the exploration of the K value was extended to 21 to include these values. While typical Cantonese speech DR was assumed to be 30 dB (Byrne *et al.*, 1994), a DR of 40 dB as suggested by Studebaker *et al.* (1999) was also evaluated for any improvement to the accuracy of intelligibility prediction.

III. RESULTS

A. Pilot study to select filtering/SNR test conditions

A speech intelligibility dropped dramatically when the cutoff frequency of the LP filter was reduced from 800 to 500 Hz, the 650 Hz LP filtering condition was added in the actual experiment. Because the 1700 Hz LP filtering condition failed to yield a high score (i.e., scores were lower than 90% correct), LP filtering conditions with cutoff frequencies at 3500, 5000, 6300, and 8000 Hz were added. A 200 Hz HP filtering condition was also added because the 500 Hz HP filtering condition failed to yield a high score. Thus, a total of 17 filtering conditions were used in the actual experiment. There were 11 LP filtering conditions with cutoff frequencies set at 500, 650, 800, 1100, 1400, 1700, 3500, 5000, 6300, 8000, and 12 000 Hz; and six HP conditions with cutoffs at 200, 500, 800, 1100, 1400 and 1700 Hz.

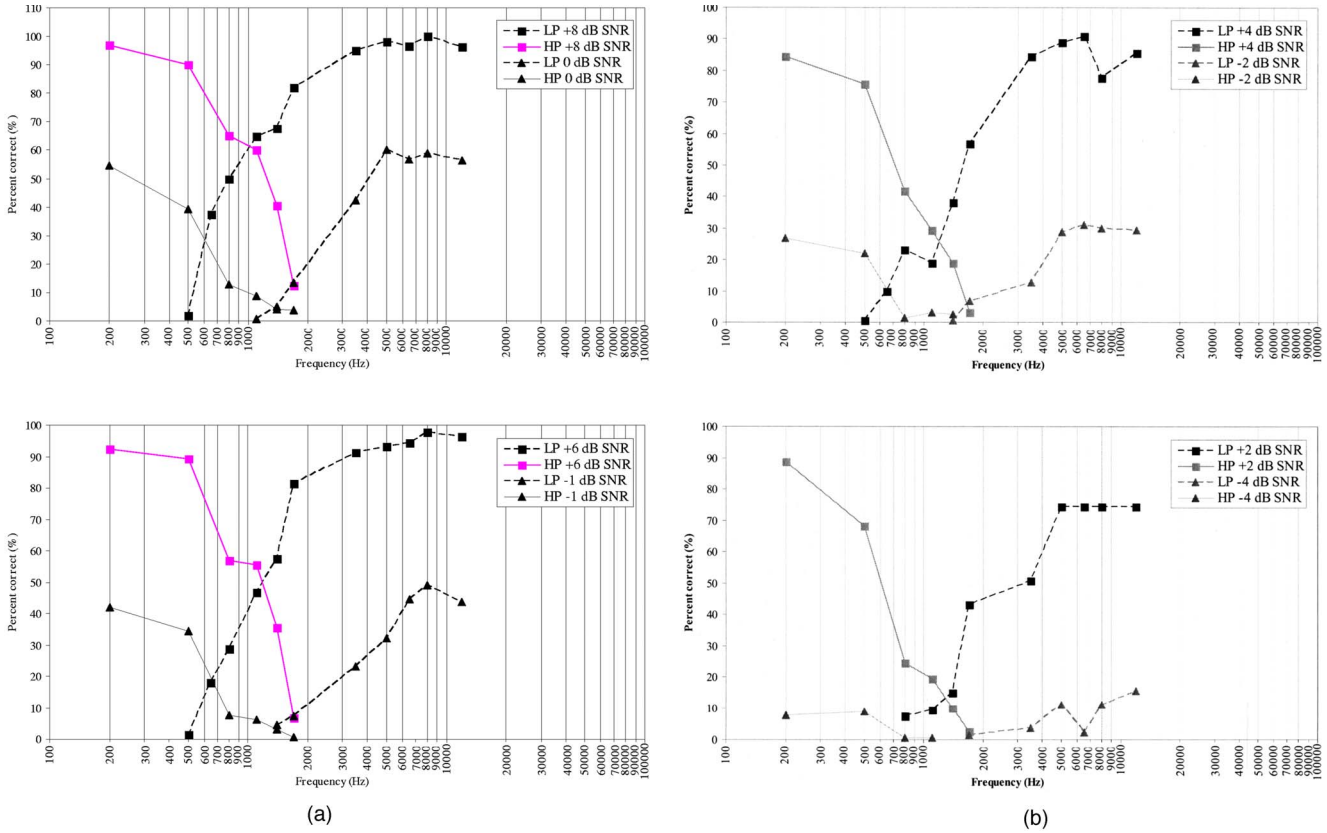


FIG. 2. Mean percent speech intelligibility, plotted as a function of cutoff frequency at various SNRs. Results from high-pass filtering conditions are represented by lines with upper ends that start from the left side of the graph; those from low-pass filtering conditions start from the right side.

As there was no substantial difference in scores between the −5 and −4 dB SNR conditions (i.e., 5% versus 8.3%), the −5 dB condition was not used for the actual experiment. Because the +7 dB SNR with LP filter cutoff at 1700 Hz condition yielded a score of only 80%, the +8 dB SNR condition was added in the actual experiment in an attempt to yield better scores. In addition, a preliminary SII was estimated using the curve bisection procedure described above. This suggested that testing using 2 dB SNR steps was adequate in generating results for SII calculations except that the −1 dB SNR condition should be retained because it yielded approximately 0.50 SII in the pilot study and would facilitate derivation of SII. Thus, in the actual study, eight SNR conditions at −4, −2, −1, 0, +2, +4, +6, and +8 dB were adopted.

As speech stimuli in some of the filtering/SNR conditions (e.g., LP filtering cutoff at 1400 Hz or below at −4 dB SNR) were consistently unintelligible, these conditions were excluded from further testing. Together, 115 filtering/SNR conditions (see Table III), instead of the 136 conditions (8 SNR × 17 filtering conditions) used in the pilot study were used in the actual study.

B. Results in various filtering/SNR conditions

The mean percent correct score in each filtering/SNR condition is reported in Table III and Fig. 2. These results suggest an improvement in intelligibility as the cutoff frequency of LP filtering was increased to about 3500 Hz and as the cutoff frequency of HP filtering was reduced to about 800 Hz. The scores also covered a wide range of performance.

C. Reception threshold of sentences and performance-intensity function

The mean RTS was −3.5 dB (s.d. 1.16). The PI function is shown in Fig. 3. Sentence intelligibility that ranged between 29.4% and 74.4% corresponded to −2 dB and +2 dB SNR (with reference to individual RTSs), respectively, in the full band condition and grew at a rate of 11.1% per dB SNR.

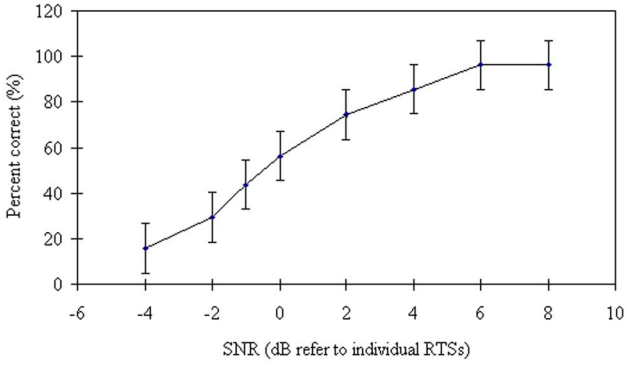


FIG. 3. PI function plotted as mean percent intelligibility at various SNRs (refer to individual RTSs). The bars represent ±1 standard error from the mean.

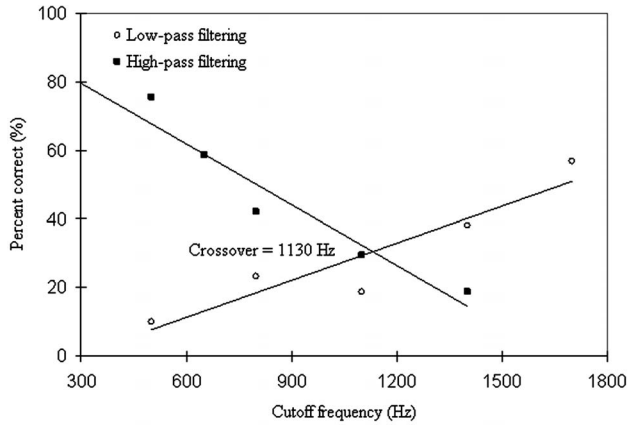


FIG. 4. Crossover frequency as the intersection between regression lines as a function of mean percent intelligibility at cutoff frequencies from 500 to 1700 Hz. Results from +4 dB SNR (refer to individual RTSs) conditions are used.

Using this PI function, the SNR for 50% correct performance is estimated at −0.3 dB, which is 3.2 dB above the mean RTS.

D. Crossover frequency

Linear regressions used to fit the data yielded crossover frequencies of 1069 Hz at −2 dB SNR, 1097 Hz at −1 dB SNR, 1110 Hz at 0 dB SNR, 1045 Hz at 2 dB SNR, 1130 Hz at 4 dB SNR, 1025 Hz at 6 dB SNR, and 1050 Hz at 8 dB SNR. Crossover frequency at −4 dB SNR (with reference to individual RTS) was not calculated because the intelligibility was very low across all filtering conditions and performance was probably affected by floor effects. The geometric average of these crossover frequencies is 1075 Hz. Mean percent performance at +4 dB SNR (refer to individual RTSs) for LP and HP filtering conditions is presented in Fig. 4.

E. Relative transfer function

In panel A of Fig. 1, the two LP and HP curves for the +8 dB SNR (refer to individual RTSs) condition are plotted. The intersection point (marked by a circle) between the two curves corresponded to 0.5 AI. The corresponding percent correct intelligibility (58%) served as the starting point at which the next two LP and HP curves were plotted. As none of the SNRs produced a 58% correct score, the next pair of LP and HP curves was estimated by interpolating data between the two curves (0 and +2 dB SNR refer to individual RTSs) that yielded scores closest to 58% in the unfiltered condition, as shown in panel B. The point where these two curves intersected was 0.25 SII. The value corresponding to 0.25 SII was about 7%. The 0.75 SII point was estimated by drawing a horizontal line through the 0.25 SII point until it intersected the LP and HP filtered curves in the best SNR condition. Two vertical lines were then drawn across the intersections until one met the upper end of the HP filtered curve, and the other met the upper end of the LP filtered curve. The circles in panel C indicated the values used to derive 0.75 SII and these values were averaged. Ten other points were estimated in a similar manner, yielding a total of 13 SII values with corresponding percent correct intelligibility as plotted in Fig. 5.

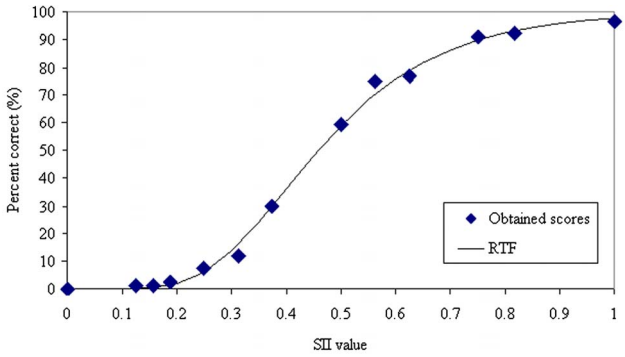


FIG. 5. Best-fit relative transfer function (RTF) and the 13 intelligibility scores (%) plotted as a function of SII values.

Equation (2) yielded the best fit SII relative transfer function (RTF), as compared to that of the other fit functions evaluated when the proficiency factor P was assumed to be 1.000. The fitting constants Q and N were found at 0.3638 and 12.2491, respectively. R^2 value of 0.9894 indicated that the model provided a good fit to the data. The RTF, plotted as a function of sentence recognition score against SII using the CHINT in the wideband condition, is also shown in Fig. 5.

F. Derivation of the frequency-importance function (FIF)

To derive the FIF, Eq. (2) was transformed to Eq. (3). The adjusted Q value was 0.3647, the value of N was 12.1488, and the R^2 value was 0.9996. Again, P was assumed to be 1.000. Values for the FIF, in one-third octave bands, are summarized in Table IV and Fig. 6. The FIF is characterized by a peak at 1600 Hz which is the frequency range of greatest importance for CHINT sentence recognition. Cumulative values of the CHINT FIF are plotted in Fig. 7, together with those of similar materials in English. As the FIFs for ANSI S3.5-1997 and Pavlovic (1984) were derived from the same data, the ANSI 3.5-1997 cumulative FIF is not plotted in Fig. 7. Frequency regions below 557 Hz and above 2331 Hz each accounted for 25% of importance weight. The midpoint of the FIF is at 1183 Hz.

TABLE IV. Frequency-importance function in one-third octave bands. The weights are expressed as percentages (%).

1/3-Octave band (Hz)	Center frequency (Hz)	Weight (%)	1/3-Octave band (Hz)	Center frequency (Hz)	Weight (%)
0–180	160	5.1	1120–1400	1250	8.1
180–224	200	2.2	1400–1800	1600	9.6
224–280	250	2.7	1800–2240	2000	8.4
280–355	315	3.6	2240–2800	2500	8.2
355–450	400	4.3	2800–3550	3150	7.8
450–560	500	4.8	3550–4500	4000	6.2
560–710	630	6.1	4500–5600	5000	4.2
710–900	800	7.0	5600–7100	6300	2.9
900–1120	1000	7.3	7100–9000	8000	1.5

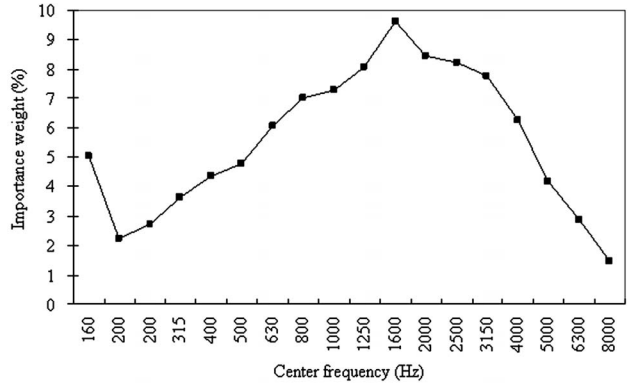


FIG. 6. Frequency-importance function (FIF) of the CHINT.

G. Derivation of the ATF

The slope of the RTF was adjusted to reflect its absolute value. The ATF is shown in Fig. 8. The iterative process of varying K and DR suggested that the smallest rms error was obtained with K set at 11.8 dB and DR set at 30 dB. These values provided the best fit of the data to the ATF. The corresponding Q and N values were 0.1894 and 12.1771 and the R^2 value was 0.8926 for predicting SII from intelligibility scores. The Q value was 0.1844 and the N value was 12.5769, with the R^2 at 0.9499 for predicting intelligibility scores using SII values. These R^2 values indicate that the model still provided a good fit to the data. Applying Eq. (3) to the mean scores, the SII for all filtering/SNR test conditions were obtained. These values are plotted in Fig. 8.

IV. DISCUSSION

A. Reception threshold of sentences and performance-intensity function

The mean RTS in noise found in this study is within the 95% confidence interval for normal hearing listeners found by Wong and Soli (2005). The slope of the PI function found in this study is also in agreement with the slope of 9.7% per dB found previously (Wong and Soli, 2005). These findings suggested that the CHINT is a consistent measure of speech intelligibility in noise. Although Studebaker *et al.* (1987) and

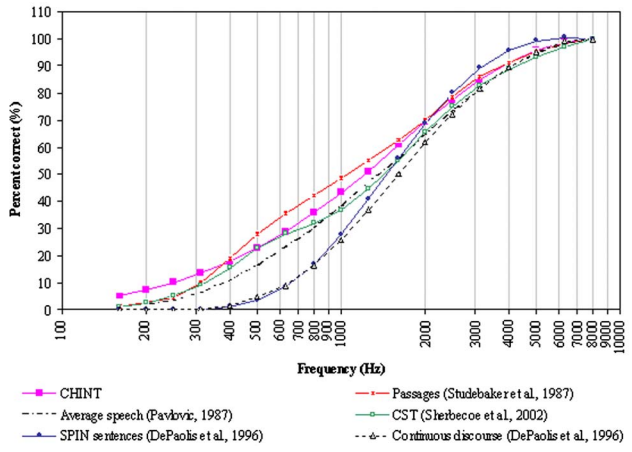


FIG. 7. Comparison of cumulative FIFs derived from the CHINT and other similar materials.

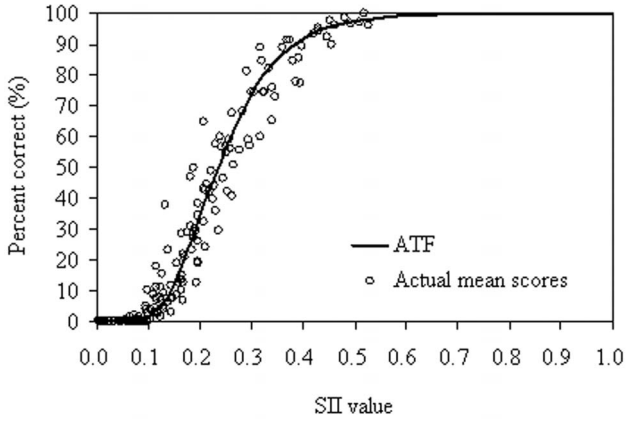


FIG. 8. Best-fit absolute transfer function (ATF) and actual mean scores (%) plotted as a function of SII.

Sherbecoe and Studebaker (2002) suggested that a steeper PI function is expected when speech and noise spectra are matched, the slope of the PI function obtained in this study is not as steep as might have been expected based on some earlier work that used talker spectrum matched maskers. In fact, the PI function is consistent with those reported for the CST by Sherbecoe and Studebaker and the English HINT by Eisenberg *et al.* (1998), and more gentle than those found by Plomp and Mimpen (1979), Hagerman (1982), and Studebaker *et al.* (1987). The CHINT materials were designed to yield a PI function slope of about 10% per dB so that they are more suitable for the HINT adaptive procedure (Wong and Soli, 2005; Nilsson *et al.*, 1994). Any influence due to clarity of speech or spectral matching between speech and noise would have been accounted for by this predetermined criterion of test development. Because test stimulus levels are specified in the same way, we are able to compare the PI function of the CHINT and the English HINT and conclude that they yielded a similar PI function (Sherbecoe and Studebaker, 2002).

B. The CHINT transfer function

As suggested by Sherbecoe and Studebaker (2002), comparing transfer functions (TFs) across studies is difficult because absolute TFs often are not reported. When absolute TFs have been reported, testing might not have been conducted using noise matched to the speech spectrum to control for filtering effects of hearing thresholds. Furthermore, *a priori* assumptions about the size of speech peaks have been made when relative TFs were converted to absolute TFs. Nonetheless, like the TFs of many English materials, The CHINT TF shows a monotonic relationship between SII values and speech recognition scores. The slope of the CHINT transfer function and the Q and N values are similar to those reported by Eisenberg *et al.* (1998) for the English HINT and Sherbecoe and Studebaker (2002) for the CST (see Table V). However, the N value for Cantonese HINT is much smaller than the mean number of phonemes (26.3) in CHINT sentences, in contrast with the English HINT sentences with a mean number of phonemes (16.8) per sentence matching the N value. It seems therefore, that Cantonese phonemes are not perceived as separate units but "chunks." This speculation

TABLE V. Comparison of transfer functions (TFs) and frequency-importance functions (FIFs) for various speech materials.

Authors	Material	Q	N	TF slope[a]	FIF (shape, peaks)
Current study	CHINT sentences	0.1844	12.58	11.0	bimodal, below 200 Hz and around 800–1600 Hz
DePaolis *et al.* (1996)	PB-50 words	0.641	2.436	≈4.0	unimodal, around 2000 Hz
	SPIN sentences	0.329	4.481	≈8.0	unimodal, around 2000 Hz
	Continuous discourse	0.353	8.943	≈7.0	unimodal, around 2000 Hz
Eisenberg *et al.* (1998)	HINT sentences	0.235	15.13	≈10.0	unimodal, 2000 Hz using average
	ANSI S3.5 standard	0.247	16.90	≈10.0	
Henry *et al.* (1998)	CNC monosyllables	0.474	2.518	—	unimodal, 2000 Hz
Sherbecoe and Studebaker (2002)	CST passages	0.227	10.26	10.6	bimodal, 500 and 1600 Hz
Studebaker and Sherbecoe (1991)	CID W-22 words	0.283	4.057	10.2	bimodal, 400 and 2000 Hz
Studebaker *et al.* (1993)	NU-6 words	0.404	3.334	6.4	bimodal, 500 and 2000 Hz
Studebaker *et al.* (1987)	Continuous discourse	—	—	18.7	bimodal, 500 and 2500 Hz

[a]TF slopes are in percent per 0.0333 SII and are based on observed or estimated scores between 20 and 80%. Numbers are either reported in the relevant studies or estimated by the authors according to reported TFs (≈ denotes approximation). TF slope data are not available in Henry *et al.* (1998).

requires further research to verify. However, an example may help illustrate this phenomenon.

The CHINT sentence, /tai$_6$ kɔ$_1$ siŋ$_4$ jɐt$_6$ hɐi$_2$ kʊŋ$_1$ si$_1$ kɔŋ$_2$ tin$_6$ wa$_2$/, means "my big brother is on the phone all day long at work." The Chinese word /kɔ$_1$/ means "brother" and would limit the word before it to those related to order of birth. The word /jɐt$_6$/ means "day" and would limit the word before it to mean the day before or after, or all day. The word /hɐi$_2$/ means "in" and refers specifically to a physical location. When followed by the character /kʊŋ$_1$/ (work), the next word must be /si$_1$/ which together with /hɐi$_2$/ and /kʊŋ$_1$/ mean at one's workplace. The words /kɔŋ$_2$/ and /wa$_2$/ both mean speaking and when spoken in a sequence, the only words that can fit between are /tin$_6$/ (electric), /tai$_6$/ (big), or /siu$_3$/ (laugh). The three monosyllables together mean talking on the phone, lying or joking. Therefore, it seems that individual Chinese speech sounds are not independent of each other and perhaps is related to the fact that Chinese polysyllabic words are made up of semantically meaningful monosyllabic parts. Chinese polysyllabic words seemed to have greater semantic and syntactic constraints than their English counterparts. This redundancy has made shorter Chinese sentences inappropriate for adaptive testing. Adverbial phrases were added to shorter sentences to derive the CHINT sentences to make them less redundant (Wong and Soli, 2005).

An SII value of 0.5 or higher would yield close to maximum intelligibility using CHINT sentences (97.6%). This would be consistent with findings of other materials (e.g., the CST) that are more redundant in content than single words (e.g., NU-6). At the same SII, 89.3% intelligibility is expected with the English HINT. As greater constraint on speech material (e.g., grammatical structure and context) and greater redundancy would yield higher percent intelligibility for a given AI (ANSI-S3.5 1969, p. 21; Studebaker *et al.*, 1987), we can conclude that the Cantonese materials are more redundant than the English HINT and materials that employ single-word stimuli such as the NU-6 (Studebaker *et al.*, 1993).

In summary, the CHINT sentences have fewer independent sounds than would be suggested by the number of phonemes in the sentences. The CHINT material is more redundant in context than similar materials such as the English HINT or single-word materials.

C. Crossover frequency and frequency-importance function

As the crossover frequency decreases, the relative importance of low-frequency information increases. Since the crossover frequency is lower for Cantonese than for all English speech materials (see Table I), we conclude that low frequencies in Cantonese contain more speech information than in English. Results from the Cantonese HINT FIF (Fig. 7) also show that when compared to similar English materials, the 1/3 octave band centered at 180 Hz carries more weight for speech understanding. As a result, the whole FIF is shifted down in frequency; 75% of CHINT information is located below 2331 Hz. Figure 7 also shows that the shape of CHINT cumulative FIF resembles those of average speech derived by Pavlovic (1987) and the ANSI S3.5-1997, with the exception that frequencies below 400 Hz are slightly more heavily weighted and frequencies above 4000 Hz exhibit reduced importance when compared to equivalent English materials.

Several reasons might contribute to differences in CF and FIF across materials. First, redundancy of materials may be a factor (Studebaker *et al.*, 1987; Studebaker and Sherbecoe, 1991). As discussed, CHINT appears to carry much redundant information. In fact, the crossover frequency of the CHINT material resembles that reported for continuous discourse by Studebaker *et al.* (1987) (at 1189 Hz). This contrasts with those reported for the W-22, with a crossover frequency at 1314 Hz (Studebaker and Sherbecoe, 1991), the English HINT, with crossover frequency at 1550 Hz (Eisenberg *et al.*, 1998), and nonsense syllables, with crossover frequency at 1980 Hz (French and Steinberg, 1947). Second, the shape of the FIF may differ depending on the bandwidth of the filter used to derive the function (DePaolis, 1996).

Third, the rate, clarity, and peak spectrum of the speech materials may have an effect on the FIF, so that for a given material, different talkers may yield different FIFs (Sherbe-

coe and Studebaker, 2002). This, however, is unlikely to have affected the shape of the CHINT FIF because spectrally matched noise was used (Studebaker *et al.*, 1994). The crossover frequency obtained in this study was slightly lower than the midpoint of the FIF (1183 Hz) and the crossover frequencies did not vary systematically with SNR (Studebaker *et al.*, 1993; Sherbecoe and Studebaker, 2002). Thus, the contribution of talker characteristics was small. The shift in importance weight toward lower frequency is probably due to a fourth factor—the tonal nature of Cantonese. Findings from research on tone recognition support this phenomenon (e.g., Fok Chan, 1974).

1. The role of fundamental frequency on Cantonese speech perception

Fundamental frequency (F0) contains information on pitch level and contour. F0 ranges from 80 to 210 Hz for males and 190 to 305 Hz for females (Baken, 1987; Evans *et al.*, 2006). F0 plays a crucial role in identifying the meaning of Cantonese words with identical phonemes (Fok Chan, 1974; Gandour, 1981, 1983; Lee *et al.*, 2002). While some studies found pitch contour and direction are more important than height (Fok Chan, 1974; Gandour, 1981; Cheung, 1992; Whalen and Xu, 1992), others found height a more important factor (Vance, 1976; Tse, 1977; Gandour, 1983; Lui, 2000). The CHINT FIF showed that, while the 1/3 octave band between 180 to 224 Hz contributed only minimally to intelligibility, frequencies below 180 Hz, where the fundamental frequency of male speakers lies (the CHINT was recorded using a male voice), seemed more importantly weighted. Ng (1981) also found that good Cantonese word discrimination can be achieved even when the signals have been LP filtered at 250 Hz. The contribution of tonal information is exemplified in the ability to acquire correct tone production by children with moderate to profound hearing loss and Dodd and So (1994) attributed this phenomenon to better hearing at low frequency.

2. Findings from other tonal language literature

Research on Mandarin, another Chinese dialect, also suggested that low frequencies play an important role in speech and tone recognition. Tone recognition could be preserved at a high level (94.6% correct), even with speech LP filtered at 300 Hz (Liang, 1963). Similarly, Fu *et al.* (1998) found that tone recognition of LP filtered Mandarin (at 500 Hz) was preserved. In another study, about 80% of Mandarin tones were correctly identified when speech was LP filtered at 750 Hz (Zhang *et al.*, 1981). However, the cues for tone recognition in Mandarin and Cantonese, however, are slightly different. The primary cues for Cantonese tones are pitch contour and level (Fok Chan, 1974). While fundamental frequency is the most important cue for tone recognition in both dialects, temporal (e.g., duration) and amplitude envelopes cue Mandarin sentence recognition when spectral information is absent, these cues are less crucial when more spectral information is available (Lin, 1988; Fu *et al.*, 1998; Whalen and Xu, 1992; Wei *et al.*, 2004). Similarly, Fu and Zeng (2000) found that tone duration and amplitude contours help in the identification of tone 3 in Mandarin, amplitude cues contribute to the discrimination of tone 4, and periodicity cues aid recognition of all five tones. When fundamental frequencies are absent, resolved and unresolved harmonics contribute to tone recognition (Stagray *et al.*, 1992). These results suggest that low-frequency information is important for tone recognition which, in turn, aids sentence recognition.

Overall, findings from this study suggested that low-frequency information is more important for speech understanding for Cantonese than for English. These results are consistent with findings in tone recognition experiments (e.g., Fok Chan, 1974).

V. SUMMARY AND CONCLUSION

To summarize, a SII for the CHINT material was established in this study. While the Q and N values were similar to those of English sentence materials, the N value was smaller than the average number of phonemes in each sentence. The slope of the ATF, the N value, the crossover frequency and the FIF of the CHINT suggest that low frequencies are more important for Cantonese speech recognition than English. Whether the redundancy of the CHINT material and/or the tonal nature of the language has affected this result remains uncertain. One way to separate these effects is to repeat the experiment using female recordings (with higher fundamental frequency). If similar results are obtained, the shift in importance weight at low frequency is probably related to the redundancy in the speech materials. These results also suggest that it is important to establish separate FIF and SII for various languages. The FIF obtained in this study may have important implications on how hearing aids and/or cochlear implants should be fitted to Cantonese speakers. The roles of low- or high-frequency information on speech intelligibility assessed using other Cantonese speech materials, and using materials in other tonal languages, need to be established.

ACKNOWLEDGMENT

The authors are grateful to Carol Cheung, Kammy Yeung, and Benny Zee for their assistance in data collection and analysis. Our gratitude also goes to all participants in the study, as well as to Phonak Hearing Center Hong Kong Ltd. and the University of Hong Kong Standard Chartered Community Foundation Hearing Center for their assistance in participant recruitment. This study was supported by a Research Grants Council CERG grant (HKU 7165/01H), Hong Kong, China.

Amlani, A. M., Punch, J. L., and Ching, T. Y. C. (**2002**). "Methods and applications of the audibility index in hearing aid selection and fitting," Trends Amplif. **6**, 81–129.

ANSI (**1969**). S3.5, *American National Standard Methods for the Calculation of the Articulation Index* (Acoustical Society of America, New York).

ANSI (**1997**). S3.5, *American National Standard Methods for Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).

Baken, R. J. (**1987**). *Clinical measurement of speech and voice* (Taylor and Frances, London).

Browning, L. K. (1974). "The Cantonese dialect with special reference to contrasts with Mandarin as an approach to determining dialect relatedness," Ph.D dissertation, Georgetown University.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wibraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J. Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Fronlenkov, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Cheung, P. P. (**1992**). "Tonal confusions in Cantonese at different signal-to-noise ratios," B.Sc. dissertation, University of Hong Kong.

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). "The perception of Cantonese lexical tones by early-deafened cochlear implantees," J. Acoust. Soc. Am. **111**, 2250–2256.

DePaolis, R. A., Janota, C. P., and Frank, T. (**1996**). "Frequency importance functions for words, sentences, and continuous discourse," J. Speech Hear. Res. **39**, 714–723.

Dodd, B. J., and So, L. K. H. (**1994**). "The phonological abilities of Cantonese-speaking children with hearing loss," J. Speech Hear. Res. **37**, 671–779.

Doyle, J., and Wong, L. L. (**1996**). "Mismatch between aspects of hearing impairment and hearing disability/handicap in adult/elderly Cantonese speakers: some hypotheses concerning cultural and linguistic influences," J. Am. Acad. Audiol **7**, 442–446.

Doyle, J., Schaefer, C., Dacakis, G., and Wong, L. L. N. (**2002**). "Hearing levels and hearing handicap in Cantonese speaking Australian," Asia-Pac. J. Speech Lang. Hear. **7**, 92–100.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., and Martinez, A. S. (**1998**). "Subjective judgments of clarity and intelligibility for filtered stimuli with equivalent speech intelligibility index predictions," J. Speech Lang. Hear. Res. **41**, 327–339.

Evans, S., Neave, N., and Wakelin, D. (**2006**). "Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice," Biol. Psychol. **72**(2), 160–163.

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.

Fok Chan, Y. Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (University of Hong Kong, Hong Kong).

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Fu, Q. J., and Zeng, F. G. (**2000**). "Identification of temporal envelope cues in Chinese tone recognition," Asia Pacific J. Speech Lang. Hear. **5**, 45–57.

Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.

Gandour, J. (**1981**). "Perceptual dimensions of tones: evidence in Cantonese," J. Chin. Linguist. **9**, 20–36.

Gandour, J. (**1983**). "Tone perception in Far Eastern languages," J. Phonetics **11**, 149–175.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79–87.

Henry, B. A., McDermott, H. J., McKay, C. M., James, C. J., and Clark, G. M. (**1998**). "A frequency importance function for a new monosyllabic word test," Aust. J. Audiol. **20**, 79–86.

Kamm, C. A., Dirks, D. D., and Bell, T. S. (**1985**). "Speech recognition and the articulation index for normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 281–288.

Killion, M. C., and Christensen, L. A. (**1998**). "The case of the missing dots: AI and SNR loss," Hear. J. **51**, 32–47.

Lau, C. C., and So, K. W. (**1988**). "Material for Cantonese speech audiometry constructed by appropriate phonetic principles," Br. J. Audiol. **22**, 297–304.

Lee, K. Y. S., Chiu, S. N., and van Hasselt, C. A. (**2002**). "Tone perception ability of Cantonese-speaking children," Lang Speech **45**, 387–406.

Li, R. (**1989**). "The classification of the Chinese dialects," FangYan. **4**, 241–259.

Liang, Z. A. (**1963**). "The auditory perception of Mandarin tones," Acta. Physiol. Sincia. **26**, 85–91.

Lin, M. C. (**1988**). "The acoustic characteristics and perceptual cues of tones in standard Chinese," Chin. Ling. **204**, 182–193.

Lui, J. (**2000**). "Cantonese tones perception in children," Unpublished B.Sc. dissertation, University of Hong Kong.

Macrae, J. H., and Brigden, D. N. (**1973**). "Auditory threshold impairment and everyday speech reception," Audiology **12**, 272–290.

Matthews, S., and Yip, V. (**1994**). *Cantonese: A Comprehensive Grammar* (Routledge, London).

Mueller, H. G., and Killion, M. C. (**1990**). "An easy method for calculating the articulation index," Hear. J. **43**, 14–17.

Ng, Y. H. (**1981**). "The effects of filtering on the intelligibility of Cantonese," M.Ed. dissertation, University of Manchester.

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.

Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.

Pavlovic, C. V. (**1987**). "Derivation of primary parameters and procedures for use in speech intelligibility predictions," J. Acoust. Soc. Am. **82**, 413–422.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Ramsey, S. R. (**1987**). *The Languages of China* (Princeton University Press, Princeton).

Rankovic, C. M. (**1991**). "An application of the articulation index to hearing aid fitting," J. Speech Hear. Res. **34**, 391–402.

Rankovic, C. M. (**1995**). "Prediction of articulation scores," J. Acoust. Soc. Am. **97**, 3358.

Sherbecoe, R. L., and Studebaker, G. A. (**2002**). "Audibility-index functions for the Connected Speech Test," Ear Hear. **23**, 385–398.

So, L. K. H., and Dodd, B. J. (**1995**). "The acquisition of phonology by Cantonese-speaking children," J. Child Lang **22**, 473–493.

Stagray, J. R., Downs, D., and Sommers, R. K. (**1992**). "Contributions of the fundamental, resolved harmonics, and unresolved harmonics in tone-phoneme identification," J. Speech Hear. Res. **35**, 1406–1409.

Stelmachowicz, P., Lewis, D., and Creutz, T. (**2002**). *Situational Hearing-Aid Response Profile (SHARP, version 6.0) User's Manual* (Boys Town National Research Hospital, Omaha).

Studebaker, G. A., and Sherbecoe, R. L. (**1991**). "Frequency-importance and transfer functions for recorded CID W-22 word lists," J. Speech Hear. Res. **34**, 427–438.

Studebaker, G. A., and Sherbecoe, R. L. (**1993**). "Frequency-importance functions for speech recognition," in *Acoustical factors affecting hearing aid performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 185–204.

Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130–1138.

Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993**). "Frequency-importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test," J. Speech Hear. Res. **36**, 799–807.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

Studebaker, G. A., Taylor, R., and Sherbecoe, R. L. (**1994**). "The effect of noise spectrum on speech recognition performance-intensity functions," J. Speech Hear. Res. **37**, 439–448.

Tillman, T. W., and Carhart, R. (**1966**). An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University auditory test no. 6. Technical report no. SAM-TR-66-55. San Antonio, TX: USAF School of Aerospace Medicine, Brooks Air Force Base.

Tse, J. K. P. (**1977**). "Tone acquisition in Cantonese: a longitudinal case study," J. Child Lang **5**, 191–204.

Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica **33**, 368–392.

Wei, C. G., Cao, K., and Zeng, F. G. (**2004**). "Mandarin tone recognition in cochlear-implant subjects," Hear. Res. **197**, 87–95.

Whalen, D. H., and Xu, Y. (**1992**). "Information for mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.

Wong, L., Hickson, L., and McPherson, B. (**2004**). "Hearing aid expectations among Chinese first-time users: Relationships to post-fitting satisfaction," Aust. New Zeal. J. Audiol. **26**, 53–69.

Wong, L. L. N., and Soli, S. D. (**2005**). "Development of the Cantonese Hearing in Noise Test (CHINT)," Ear Hear. **26**(3), 276–289.

Wong, A. O., and Wong, L. L. (**2004**). "Tone perception of Cantonese-speaking prelingually hearing-impaired children with cochlear implants," Otolaryngol.-Head Neck Surg. **130**, 751–758.

Zhang, J. L., Qi, S. Q., Song, M. Z., and Liu, Q. X. (**1981**). "On the important role of Chinese tones in speech intelligibility," Acta Acust. (Beijing) **4**, 237–24.

Auditory and nonauditory factors affecting speech reception in noise by older listeners

Erwin L. J. George,[a] Adriana A. Zekveld, Sophia E. Kramer, S. Theo Goverts, Joost M. Festen, and Tammo Houtgast
ENT/Audiology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 17 March 2006; revised 11 December 2006; accepted 17 January 2007)

Speech reception thresholds (SRTs) for sentences were determined in stationary and modulated background noise for two age-matched groups of normal-hearing ($N=13$) and hearing-impaired listeners ($N=21$). Correlations were studied between the SRT in noise and measures of auditory and nonauditory performance, after which stepwise regression analyses were performed within both groups separately. Auditory measures included the pure-tone audiogram and tests of spectral and temporal acuity. Nonauditory factors were assessed by measuring the text reception threshold (TRT), a visual analogue of the SRT, in which partially masked sentences were adaptively presented. Results indicate that, for the normal-hearing group, the variance in speech reception is mainly associated with nonauditory factors, both in stationary and in modulated noise. For the hearing-impaired group, speech reception in stationary noise is mainly related to the audiogram, even when audibility effects are accounted for. In modulated noise, both auditory (temporal acuity) and nonauditory factors (TRT) contribute to explaining interindividual differences in speech reception. Age was not a significant factor in the results. It is concluded that, under some conditions, nonauditory factors are relevant for the perception of speech in noise. Further evaluation of nonauditory factors might enable adapting the expectations from auditory rehabilitation in clinical settings. [DOI: 10.1121/1.2642072]

PACS number(s): 43.71.Ky, 43.71.An, 43.66.Ba, 43.66.Sr [JHG] Pages: 2362–2375

I. INTRODUCTION

In everyday life, many older listeners have difficulties understanding speech, especially in the presence of background noise or reverberation (Plomp, 1978; Duquesnoy and Plomp, 1980; Nabelek and Robinson, 1982). A review by Plomp (1978) showed that the percentage of the population with problems in perceiving speech approximately doubles with every decade in age, from 16% at an age of 60 years to about 64% at an age of 80, to nearly everyone above 86 years of age. Although a lack of audibility, due to audiometric hearing loss or masking noise, appears to be a major component, it is generally not enough to fully account for differences in speech reception among individual listeners (Eisenberg *et al.*, 1995; Bacon *et al.*, 1998; Summers and Molis, 2004; Dubno *et al.*, 2002, 2003).

Plomp (1978) formulated a model description for the speech reception threshold (SRT) based on two auditory parameters: (i) hearing loss due to attenuation, related to a raised absolute hearing threshold, and (ii) hearing loss due to distortion, which is considered to reflect suprathreshold deficits in hearing (Stephens, 1976; Glasberg and Moore, 1989). Examples of such suprathreshold deficits are reduced temporal and spectral auditory resolution and a loss of normal auditory compression. The inter-relationship between these deficits and their relation with the hearing threshold is still under discussion (Ludvigsen, 1985; Moore *et al.*, 1999; Oxenham and Bacon, 2003). Both reduced temporal resolution and reduced spectral resolution are thought to adversely affect speech perception in noise, specifically when masker levels fluctuate over time (Glasberg *et al.*, 1987; Festen and Plomp, 1990; Glasberg and Moore, 1992; Festen, 1993; Baer and Moore, 1993, 1994; Boothroyd *et al.*, 1996; Dubno *et al.*, 2003; George *et al.*, 2006). Some studies, however, indicate that even listeners with significantly broadened spectral filters still have sufficient spectral resolution to resolve the spectral cues important for speech intelligibility (Ter Keurs *et al.*, 1993a, b).

In addition, speech reception in background noise may be affected by nonauditory processes (see, e.g., Humes, 2005). Perception of speech is a process that does not only involve the peripheral auditory organ, but also depends on information processing in the central auditory pathway and on nonauditory functions, like working memory capacity and speed of information processing (Gatehouse *et al.*, 2003; Lunner, 2003; Hällgren, 2005). More generally said, speech reception is thought to be affected by an interaction between, on the one hand, bottom-up or "stimulus-driven" processes, and, on the other hand, top-down or "knowledge-driven" factors (Goldstein, 2002). The relative contribution of auditory and nonauditory processes to speech reception is, however, still under discussion.

In the late 1980s, Van Rooij *et al.* constructed a test battery comprising auditory, cognitive, and speech perception tests (Van Rooij *et al.*, 1989). In this study, they found a significant contribution of cognitive factors to speech perception in noise. However, based on the results of two subsequent studies in older subjects (Van Rooij and Plomp, 1990; 1992), they finally concluded that age-related differences in

[a]Electronic mail: elj.george@vumc.nl

 0001-4966/2007/121(4)/2362/14/$23.00

speech perception in noise are most likely due to differences in auditory factors, notably differences in audiometric hearing thresholds. Moreover, their data indicated that auditory and cognitive factors independently contribute to speech perception, and that the importance of cognitive factors does not change significantly with increasing age.

Results by Pichora-Fuller *et al.* (1995) suggest that audiometric thresholds cannot fully account for the difficulty that elderly listeners experience in understanding speech in noise. They did not find any general age-related changes in cognition either; so cognitive factors do not seem responsible for the deteriorated speech recognition. Based on their results, however, they introduced a processing model, in which auditory difficulties adversely affect speech understanding both directly, by altering the amount of correctly perceived words, and indirectly, because effortful listening consumes resources that could otherwise be allocated to cognitive processes necessary for speech understanding. Consistent with this model, it was suggested by Hällgren (2005) that the relative importance of top-down or cognitive functions increases when speech information is degraded, either by hearing impairment, or by the presence of background noise or reverberation.

Moreover, increasing evidence exists that, averaged over groups of listeners, age can affect the processing of sounds, independently of hearing loss (Snell, 1997; Grose *et al.*, 2001; Dubno *et al.*, 2002; Gordon-Salant and Fitzgibbons, 2004; Gifford and Bacon, 2005). Recently, Divenyi *et al.* (2005) showed that the deterioration of speech reception with age is accelerated significantly relative to the decline in audiometric measures, which may indicate an accelerating decline of central processing. This finding suggests that the effect of age on speech reception might be related to nonauditory factors, as proposed earlier by, for instance, Gordon-Salant and Fitzgibbons (1997) and Humes (2002).

The importance of nonauditory factors is also apparent when considering the use of language proficiency (Van Wijngaarden, 2002, 2004) or, as more commonly investigated, sentence context (Boothroyd and Nittrouer, 1988; Nittrouer and Boothroyd, 1990; Gordon-Salant and Fitzgibbons, 1997; Dubno *et al.*, 2000). The relevance of context was already noticed by Warren (1970), who demonstrated that listeners can perceive missing phonemes by using the redundancies in speech at the acoustic, phonetic, phonological, and/or lexical level. Grant *et al.* (1998) referred to this function as perceptual closure, i.e., the ability to form linguistic wholes from perceived fragments. In a subsequent study, Grant and Seitz (2000) showed that this ability is modality-aspecific and may vary substantially across hearing-impaired subjects. Moreover, their results indicate that the importance of the use of context increases under degraded listening conditions, consistent with the above-mentioned processing model by Pichora-Fuller *et al.* (1995).

A common test to investigate the ability to make use of sentence context is the Speech Intelligibility in Noise or SPIN test (Kalikow *et al.*, 1977; Bilger *et al.*, 1984; for a review, see Elliott, 1995). However, SPIN performance does not only depend on context use, but may also be related to interindividual differences in auditory factors. In other words, measuring modality-independent factors by means of auditory stimuli may give rise to a confounding effect between auditory and nonauditory factors.

An alternative approach to confirm the importance of nonauditory factors to speech reception is to assess the relationship between auditory speech-reception performance and visual speech-reading abilities, as performed by for instance Watson *et al.* (1996). They found significant correlations between overall auditory and visual performance, and suggested that this association likely reflects the shared relevance of one cognitive function in visual and auditory language comprehension.

The current experiment aims at determining differences in speech reception between normal-hearing and hearing-impaired listeners and investigating the contribution of auditory and nonauditory factors in accounting for speech intelligibility in noise. In this paper, auditory factors are defined as factors that are related to processing in the peripheral and central auditory pathways, while nonauditory factors as meant here are related to modality-aspecific central, cognitive, or linguistic skills.

Both normal-hearing and sensorineural hearing-impaired, older listeners were selected to participate in the present experiment. To be able to adequately compare differences between the groups, the groups were matched according to age, thus avoiding different performance due to heterogeneity of age between the groups. Speech reception was assessed by adaptive measurements of the SRT in stationary and modulated background noise, following the procedure as described by Plomp and Mimpen (1979). To optimize audibility at all frequencies, individual hearing thresholds were used to adapt the spectrum of the noises to reach levels equal to the estimated middle of the dynamic range for each listener, as performed earlier by George *et al.* (2006). Measurements in modulated noise were included because differences between normal-hearing and hearing-impaired listeners are expected to be more prominent in nonstationary maskers (see, for instance, Festen and Plomp, 1990). Moreover, nonstationary maskers are more representative for listening conditions in everyday life (Kramer *et al.*, 1996).

Correlations will be studied between SRT in noise and measures of auditory and nonauditory performance, after which stepwise regression analyses will be performed within both groups separately. Auditory performance will be assessed by measuring each listener's pure-tone audiogram and by testing both spectral and temporal acuity (F and T), which are regarded to reflect the individual auditory-filter and temporal-window width. To assess the contribution of nonauditory factors, a visually presented test was used, which was developed by Zekveld *et al.* (2007) to be as similar as possible to the adaptive SRT measurement. The outcome of this test will be referred to as the text reception threshold (TRT). In this test, everyday Dutch sentences were presented visually on a screen, partially masked by an adaptively changing masking pattern. Results by Zekveld *et al.* (2007) show that there is a significant correlation between the ability to read masked text (TRT) and auditory speech-reception (SRT) in a group of normal-hearing participants. The current

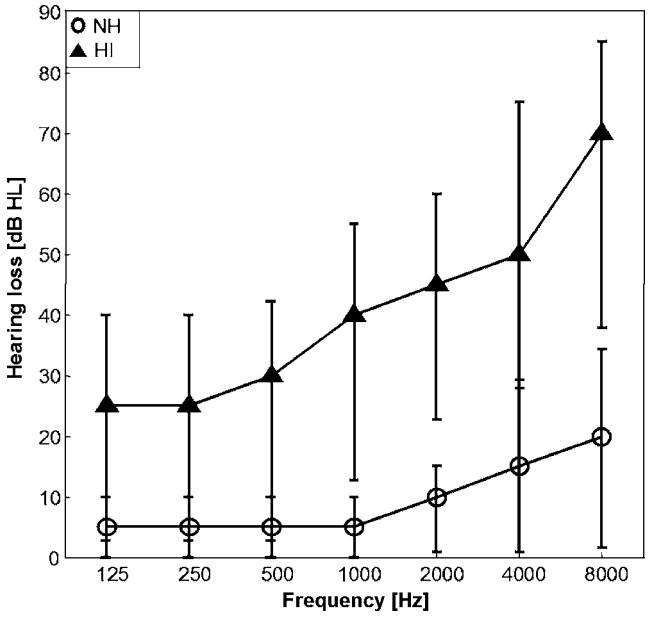


FIG. 1. Median pure-tone hearing thresholds (*re*: ISO-389-1991) and 5th and 95th percentiles for the normal-hearing (NH, $N=13$) and the hearing-impaired (HI, $N=21$) participants.

paper will extend these results with measurements in a group of hearing-impaired listeners and age-matched normal-hearing listeners.

II. EXPERIMENT AND METHOD

A. Participants

Thirteen normal-hearing (NH) and twenty-one sensorineural hearing-impaired (HI) listeners participated in this experiment. The hearing-impaired participants (12 females, 9 males) were patients of the audiology department of the VU University Medical Center, selected to have a symmetrical sensorineural hearing loss, with pure-tone thresholds up to 60 dB HL and interaural threshold differences smaller than 10 dB. The age of the hearing-impaired listeners ranged from 46 to 81 years, with an average of 65.5 years. The normal-hearing listeners (8 females, 5 males) were acquaintances of the hearing-impaired participants, selected to have pure-tone hearing thresholds better than 15 dB HL at 0.25, 0.5, 1.0, and 2.0 kHz and better than 30 dB at 4 kHz. The age of the normal-hearing listeners ranged from 53 to 78 years, with an average of 63.5 years. Figure 1 shows the median and spread of hearing loss for both groups of participants.

All participants were native speakers of Dutch and reported normal or corrected-to-normal vision. Their color vision was screened with Ishihara plates (Ishihara, 1989) and classified as normal.

B. Description of the tests

1. Pure-tone thresholds

Each experimental run started with the measurement of the listener's pure-tone hearing thresholds at octave frequencies between 125 and 8000 Hz, using the same apparatus as during all other measurements (see Sec. II C). The audiogram was used to shape the spectrum of the auditory stimuli presented in the subsequent tests. In the regression analyses, only the pure-tone average (PTA) is included as a predictor variable, defined as the average pure-tone hearing threshold of the subject's best ear over octave frequencies 0.5, 1.0, and 2.0 kHz.

2. Spectral and temporal acuities (F / T)

Spectral and temporal acuities of each listener's best ear were determined by employing an adaptive measurement procedure as introduced and validated by Hilkhuysen *et al.* (2005). Validation was performed by measuring 18 normal-hearing listeners. They showed auditory filter and time-window widths which were free of noteworthy learning effects, and which varied with presentation level and frequency and corresponded to values as commonly found in the literature. The measurement procedure was also used and explained in detail in a study investigating the effects of spectral and temporal acuity on masking release for speech (George *et al.*, 2006).

We determined spectral and temporal resolution by measuring the thresholds of short tone sweeps in spectrally or temporally modulated maskers (grids), and relating these results to the threshold in unmodulated noise. In the measurement procedure, listeners were asked to report the number of tone sweeps (zero to three) they were able to detect in (i) steady-state noise without grid; (ii) noise containing a spectral grid with a 50% duty cycle on a log-frequency scale; and (iii) noise containing a temporal grid with a 50% duty cycle. In all three noises, the tone sweeps to detect were sinusoids with a duration of 200 ms, sweeping upward over a range of 1.6 octaves centered around 1 kHz (0.57–1.74 kHz) at a speed of 8 octaves/s. Thus, the sweep reached its center frequency after 100 ms. The masker duration was 2.2 s, and the possible tone sweeps could start at 0.6, 1.0, or 1.4 s after masker onset. To determine detection thresholds, the level of the tone sweeps (for the steady-state noise) or the gap width of the noise-grid maskers was varied adaptively in a one-up-one-down 4-AFC-procedure (Levitt, 1971), starting above detection threshold for all listeners.

The observed masking release for the noise grids as compared to steady-state noise was used to estimate auditory filter and time-window widths. In our experiment, both temporal and spectral acuity were determined in the frequency region around 1 kHz, at a presentation level halfway up the listener's dynamic range. Thus, signals are spectrally optimized with respect to individual hearing thresholds. This means that audibility is optimized (threshold-related effects are minimized) and the outcome measures can be assumed to be related to suprathreshold processing. A drawback of this method is, however, that the overall presentation level is different for each listener. Effects of presentation level will thus have to be considered.

Spectral resolution is known to deteriorate with increasing presentation level for normal-hearing listeners (Dubno and Schaefer, 1992; Sommers and Humes, 1993a,b), while temporal resolution is enhanced at higher levels (Jesteadt *et al.*, 1982; Fitzgibbons, 1983; Fitzgibbons and Gordon-Salant, 1987). In our previous study (George *et al.*, 2006),

these effects of presentation level on spectral and temporal resolution were estimated by regression lines based on measurements in a group of normal-hearing listeners. It was shown that deteriorated spectral resolution for moderately hearing-impaired listeners could almost fully be accounted for by the "normal" effects of presentation level, indicating that deteriorated spectral resolution does not qualify as an actual suprathreshold deficit. In contrast, level-corrected temporal resolution was deteriorated for most hearing-impaired listeners, indicating that it is an actual suprathreshold deficit. The deviation of measured temporal and spectral resolution from the regression lines was suggested to be the most appropriate measure for the actual amount of suprathreshold deficits of a specific listener. Therefore, in the current study, spectral and temporal acuities are corrected for presentation level in the same way.

Both level-corrected spectral and temporal acuities are included as predictor variables in the regression analyses. They are denoted by "ΔF" and "ΔT," where the use of the capital Greek delta indicates that effects related to presentation level have been accounted for, as described by George *et al.* (2006), such that larger-than-normal values can be considered to reflect deteriorated suprathreshold processing.

3. Speech reception threshold

SRT measurements were performed using a simple adaptive one-up-one-down procedure as described by Plomp and Mimpen (1979), in stationary background noise, as well as in a masker modulated with a 16-Hz square wave with a duty cycle of 50%. The long-term average spectra of the two maskers were the same. The appropriate masker and a list of 13 Dutch everyday sentences were presented monaurally to the listener's best ear, sentence by sentence. Sentences were read by a female speaker (Plomp and Mimpen, 1979) and were unknown to the listener. The long-term rms level of the masker was kept fixed, while the speech rms level was varied adaptively to estimate the SRT, i.e., the speech-to-noise-ratio at which 50% of the sentences are reproduced without error. In each condition, the first sentence was presented at a level below threshold and repeated, at 4-dB higher levels with each repetition, until the listener was able to reproduce it correctly. The remaining 12 sentences were presented only once, following the adaptive procedure, with a 2-dB step size. The SRT was estimated as the average signal-to-noise level of sentences 5–14. The 14th sentence was not presented, but its level was determined by the response to the 13th sentence.

To optimize audibility at all frequencies, individual hearing thresholds were used to adapt the spectrum of both stationary and modulated noise to reach octave masker levels equal to the estimated middle of the dynamic range for each listener. The lower limit of the dynamic range was chosen to be the individual pure-tone threshold, while the upper limit is the uncomfortable loudness level (UCL), chosen at 115 dB SPL for all listeners. Because pure-tone thresholds were only measured at octave frequencies from 125 and 8000 Hz, intermediate threshold levels were interpolated. The overall rms level of each of the two maskers, averaged over participants, was 70.9 dB(A) for the normal-hearing group and 92.6 dB(A) for the hearing-impaired group, as can be derived from the audiograms given in Fig. 1. The shape of the speech spectrum was modified to be the same as that of the noise, while the level of the speech was varied, following the adaptive procedure as described earlier. The bandwidth of the noise and the speech signals was restricted to frequencies between 223 and 4490 Hz.

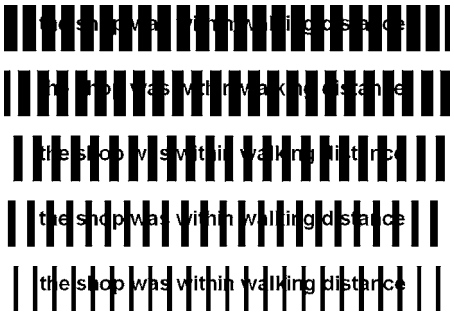


FIG. 2. Typical stimuli to measure the text reception threshold (TRT): a sentence masked by a vertical bar pattern. Between sentences, the degree of masking was adaptively varied. The field back color was white, text color was red, and the color of the mask was black. The shown percentages of unmasked text are 28%, 40%, 52%, 64%, and 76%, respectively. Here, a Dutch sentence from the lists by Versfeld *et al.* (2000) was translated into English. Adopted from Zekveld *et al.* (2007).

SRT in stationary noise and SRT in modulated noise are identified as criterion variables in the regression analyses, and are denoted by "SRT_{STAT}" and "SRT_{MOD}," respectively.

4. Text reception threshold

In order to determine whether nonauditory processes could play a role in the reception of speech, a visually presented test was used that is similar to the auditory SRT test. This test was developed earlier by Zekveld *et al.* (2007), who used it to investigate the relation between the ability to comprehend speech in noise and the ability to read masked written text in normal-hearing subjects. A list of 13 everyday Dutch sentences, adopted from Versfeld *et al.* (2000), was presented visually on a computer screen, sentence by sentence. Each presented sentence was partially masked by vertical bars with a specific degree of masking, adaptively changing over sentences. The field back color was white, text color was red, and the color of the mask was black. The text of each sentence appeared in a word-by-word fashion. The timing of the appearance of each word was based on the original recordings that were available for each sentence. The sentence disappeared 3.5 s after presentation of the last word of the sentence, while the mask remained visible until the next trial started with the presentation of a newly created mask.

The number of vertical bars in the masking pattern was determined such that for most masking percentages, each letter was partly masked and partly unmasked. Examples of sentences masked by this pattern are shown in Fig. 2. The exact position of the mask relative to the sentence was randomly determined.

The amount of masking was varied adaptively, following an one-up-one-down procedure similar to the SRT test, to determine the text reception threshold or TRT, defined as the

TABLE I. Group averages (M), standard deviations (S), reliabilities (r_{tt}), standard errors of measurement (SEM), and two-tailed t-statistics of the test outcomes for the normal-hearing (NH) and the hearing-impaired (HI) participants. Test reliabilities r_{tt} have been calculated from test-retest correlations r_{tr} using the Spearman-Brown formula: $r_{tt}=2r_{tr}/(1+r_{tr})$), cf. Nunnally, 1967. SEM is defined as $S^*\sqrt{(1-r_{tt})}$. Group differences are considered significant if $p<0.05$.

			NH ($N=13$)				HI ($N=21$)					
		Unit	M	S	r_{tt}	SEM	M	S	r_{tt}	SEM	t	p
Auditory	ΔF	Hz	212.9	22.8	0.60	14.4	220.5	79.3	0.84	32.0	0.408	0.687
	ΔT	ms	3.59	0.85	0.98	0.12	6.98	3.07	0.92	0.84	4.776	<0.001
	PTA	dB HL	6.7	3.0	⋯	⋯	34.4	12.0	⋯	⋯	10.041	<0.001
	SRT_{STAT}	dB SNR	−3.2	1.1	0.60	0.7	−1.4	2.3	0.83	0.9	2.966	0.006
	SRT_{MOD}	dB SNR	−14.0	2.6	0.88	0.9	−9.0	4.4	0.96	0.9	4.139	<0.001
	SRT^{**}_{MOD} [a]	dB SNR	−13.1	3.3	0.82	1.4	−8.2	5.2	0.88	1.8	3.408	0.002
Nonauditory	Age	Years	63.5	9.3	⋯	⋯	65.5	9.9	⋯	⋯	0.576	0.569
	TRT	% text	58.2	4.3	0.88	1.5	58.8	3.3	0.87	1.2	0.427	0.674
	TRT^{**} [a]	% text	58.9	4.7	0.83	1.9	59.7	4.2	0.79	1.9	0.510	0.615

SRT^{**}_{MOD} and TRT^{**} are based on only one measurement in test and retest, instead of two (SRT_{MOD}) or three (TRT).

amount of unmasked text needed by the subject to comprehend 50% of complete sentences correctly. Zekveld *et al.* (2007) showed that changing the amount of unmasked text with a step size of 6% would yield a change in the proportion of correct responses comparable to the 2-dB steps in the SRT test. In each condition, the first sentence was presented at a masked text percentage below threshold and was repeated, with an increased percentage of unmasked text, until the listener was able to reproduce it correctly. Also similar to the SRT, a double step size (12%) was used for the first sentence. The remaining 12 sentences were presented only once, following an adaptive procedure with a 6% step size. The TRT was estimated as the average percentage of unmasked text of sentences 5–14. It is regarded here as a general measure of modality-aspecific cognitive and linguistic skills contributing to the perception of partially masked sentences. The TRT is included as a predictor variable in the regression analyses.

C. Instrumentation and general procedure

The experiment was run on a Dell personal computer, equiped with a Creative Labs Audigy external sound device and Beyer Dynamic DT48 headphones. Sound calibrations were performed with a Brüel & Kjær Artificial Ear (type 4152) and a Brüel & Kjær 2260 Observer conform ISO 389 (1991). All measurements were performed while the listener and the investigator were seated in a sound-insulated room. Spectral shaping of auditory signals was performed by using individual thresholds as inputs via a 1024-point windowed finite impulse response filter. This filter also corrected the headphones frequency response and restricted the bandwidth of auditory signals to frequencies between 223 and 4490 Hz.

All measurements were performed following a test-retest design in a single session, interrupted by several small breaks. The test and retest blocks each included the measurement of three TRT, two SRT_{MOD} and single SRT_{STAT}, F and T. The TRT and the SRT_{MOD} measurements were performed more than once, both in test and retest, to improve reliability. Test and retest outcomes were averaged. The order in which the tests were presented was fixed. A session always started with the measurement of the audiogram and color vision screening. Auditory measurements were conducted monaurally, using the participant's best ear, which was chosen according to his or her audiogram (PTA), or, in case of equal PTAs, personal preference in telephone conversation.

D. Statistical analysis

Statistical analyses were performed using SPSS for Windows, release 11.0.1. Kolmogorov-Smirnov tests were performed to check the normality of variables, which showed that all variables were normally distributed within each group. Considering the expected nonequal variances in the two groups, two-tailed t-tests assuming nonequal variances were used to examine group differences between the test outcomes of the normal-hearing and the hearing-impaired participants. Correlation coefficients were calculated for each of the groups separately, to investigate which predictor variables are significantly associated with speech reception in stationary and modulated noise. Although it is common practice to scale down the significances when performing multiple comparisons or correlations (Miller, 1981), it was decided to adopt a tolerant criterion for significance, considering the exploratory nature of our study. All effects reaching p values below 0.05 are indicated by asterisks.

Finally, to account for predictor cross correlations, stepwise multiple regression analyses were performed for both groups to investigate which predictor variables could most effectively account for intersubject variance in speech reception. The variables PTA, ΔF, ΔT, TRT and age were included as predictor variables in these analyses, while SRT_{STAT} and SRT_{MOD} were included as criterion variables. The percentages of explained variance (R^2) reported in the following have already been corrected for the available degrees of freedom.

III. RESULTS

A. Differences between groups

Table I displays group averages, standard deviations, reliabilities, standard errors of measurement, and t-statistics of the test outcomes for normal-hearing and hearing-impaired

participants. The test outcomes are ordered by origin (auditory or nonauditory). To investigate to which extent the multiple measurements of both the TRT and the SRT_{MOD} tests increase reliability and significance of group differences, Table I also shows data when these variables would have been measured only once in test and retest, like all other test outcomes.

Results of the t-test show that group differences for the SRT in modulated noise (SRT_{MOD}) are larger than for the SRT in stationary noise (SRT_{STAT}). Even when SRT_{MOD} is based on only one test-retest measurement (SRT^{**}_{MOD}), the difference in speech reception between the groups appears to be more prominent in modulated noise compared to stationary noise, as indicated by the larger t-value. To statistically test this hypothesis, a two (group) by two (background noise) repeated-measures ANOVA was performed on SRT_{STAT} and SRT^{**}_{MOD}, which showed significant effects of group ($p=0.004$) and background noise ($p<0.001$), but also a significant interaction between group and background noise ($p=0.01$). This interaction indicates that the difference in speech reception between the groups is significantly larger in modulated noise compared to stationary noise. Thus, measuring the SRT in modulated noise instead of in stationary noise indeed increases the ability to discriminate between normal-hearing and hearing-impaired participants.

In addition, it can be concluded from the t-statistics that the performances of both groups differ significantly for auditory tests, like temporal acuity (ΔT) and audiogram (PTA). The latter is not surprising, since the groups were selected on the basis of their differences in audiometric thresholds. In contrast, spectral acuity (ΔF) does not appear to differentiate between normal-hearing and hearing-impaired listeners, as shown before by George *et al.* (2006). Moreover, Table I shows that the normal-hearing and hearing-impaired groups do not perform significantly different on the TRT test.

The fact that differences between the two groups are mainly auditory in nature indicates that differences in speech reception between the groups are also likely to be mainly governed by auditory factors. To investigate whether nonauditory factors nevertheless are associated with interindividual variance in speech reception in noise, results were investigated within both groups separately, thus partializing out the group effect.

B. Correlation analyses within groups

Table II gives product-moment correlations between the tested predictor variables and SRT in stationary noise (SRT_{STAT}) and in modulated noise (SRT_{MOD}) for both the normal-hearing and the hearing-impaired participants. Age is included as an extra predictor variable.

It can be seen in Table II that, for the normal-hearing listeners, nonauditory factors have the largest association with speech intelligibility, especially in modulated noise. In contrast, for the hearing-impaired group, Table II shows that nonauditory factors do not appear to be significantly associated with individual differences in speech reception. Instead, mainly auditory factors, specifically temporal acuity (ΔT), and audiogram (PTA) appear to be related to speech intelligibility in this group, both in stationary and nonstationary noise. It should be noted, however, that the differences in ranges of auditory and nonauditory measures between both groups contribute to the observed differences in correlations, as will be discussed in Sec. IV A.

TABLE II. Product-moment correlations between the predictor variables and SRT in stationary noise (SRT_{STAT}) and SRT in modulated noise (SRT_{MOD}). All correlations are calculated separately for the normal-hearing (NH) and the hearing-impaired (HI) participants. Significant correlations are displayed in bold, p values are denoted by asterisks: $^{(*)}p<0.05$; $^{(**)}p<0.01$; $^{(***)}p<0.001$.

		NH ($N=13$)		HI ($N=21$)	
		SRT_{STAT}	SRT_{MOD}	SRT_{STAT}	SRT_{MOD}
Auditory	ΔF	0.12	0.22	−0.05	0.13
	ΔT	−0.26	−0.09	**0.52**$^{(*)}$	**0.73**$^{(***)}$
	PTA	0.39	0.48	**0.71**$^{(***)}$	**0.73**$^{(***)}$
Nonauditory	Age	0.37	0.43	0.29	0.39
	TRT	**0.61**$^{(*)}$	**0.80**$^{(***)}$	0.34	0.42

Figures 3 and 4 show, for the SRT in stationary and in modulated noise, the relationship with the audiogram (PTA) and temporal acuity (ΔT), respectively. Figures 3 and 4 illustrate that audiogram and temporal resolution can differentiate between the normal-hearing and hearing-impaired groups, as indicated by the t-statistics in Table I. However, they also confirm that auditory factors cannot account for all of the variance in SRT, specifically within the normal-hearing group. The correlation analysis shows that part of this variance may be related to interindividual differences in nonauditory factors, as measured by the TRT.

To further investigate the relation between the TRT and SRT in stationary and modulated noise, SRT has been plotted as a function of the TRT for both groups in Fig. 5. It can be seen that the range (i.e., average and variance) of the TRT, as given in Table I, is the same for the normal-hearing and hearing-impaired listeners. This illustrates the results from the t-test performed earlier: the groups cannot be distinguished on the basis of the TRT. Figure 5 will be rediscussed later in the light of the results of the stepwise multiple regression analyses.

Finally, it can be seen in Figs. 3–5 that the ranges of both SRT_{STAT} and SRT_{MOD} are different for the hearing-impaired group and for the normal-hearing group. This illustrates that the hearing-impaired group performs significantly worse than the normal-hearing group on both SRTs, as shown by the t-statistics in Table I. The difference in SRT between both groups appears more prominent in modulated noise, though.

C. Stepwise multiple regression within groups

As mentioned earlier, the above presented correlation analysis on speech reception is not ideal. It gives a distorted picture, because the predictors are cross correlated, as can be seen in Table III. This possibly leads to induced correlations between a predictor and speech intelligibility. To control for

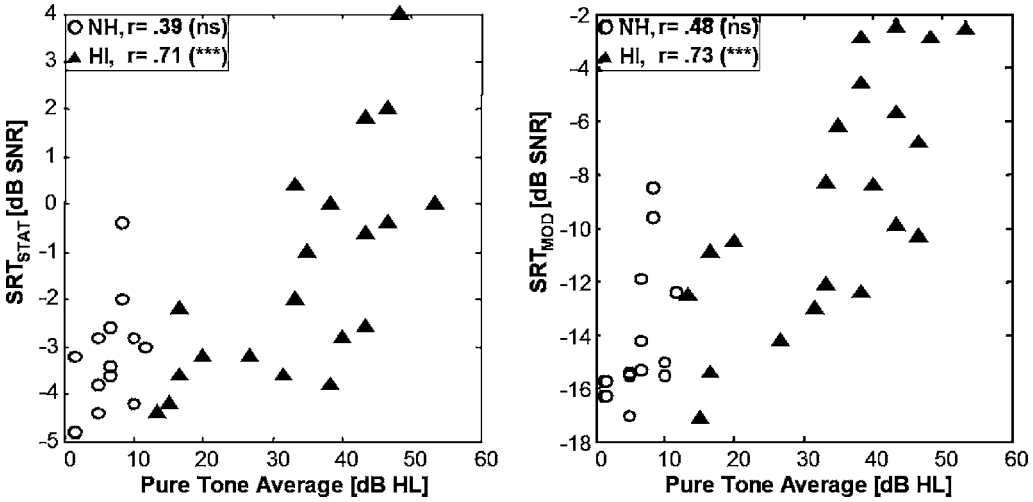


FIG. 3. SRT in stationary noise (SRT_{STAT}, left) and in 16-Hz block-modulated noise (SRT_{MOD}, right) as a function of pure-tone average (PTA), for the normal-hearing (NH, $N=13$) and hearing-impaired (HI, $N=21$) participants. Also shown are the correlation coefficients and their significance: (ns) nonsignificant; $^{(*)}p<0.05$; $^{(**)}p<0.01$; $^{(***)}p<0.001$.

these cross correlations, stepwise multiple regression analyses were performed, again for each group of participants separately.

For the normal-hearing group, results from the stepwise multiple regression analysis show that the TRT is the predictor variable that contributes most to explaining variance in speech reception, both in stationary and in modulated noise. On its own, the TRT accounts for 31% of the intersubject variance in speech reception in stationary noise and for 60% of the variance in speech reception in modulated noise, as already indicated by the correlations between TRT and SRT displayed in Table II. When the TRT is included in the model, no other predictor significantly explains variance over and above the variance explained by the TRT ($p>0.46$ in both cases).

For the hearing-impaired group, results from the stepwise multiple regression analysis are displayed in Table IV. In stationary noise, only PTA significantly contributes to explaining differences in speech reception, accounting for 47% of the intersubject variance. When PTA is included in the model, no other predictors are significantly correlated with the residual ($p>0.34$). In modulated noise, mainly temporal acuity (ΔT) contributes to explaining variance in speech intelligibility for hearing-impaired participants, accounting for 48% of the variance on its own. The TRT explains only 9% of the variance ($p=0.06$) when it is the only predictor included, but becomes a significant term ($p=0.0006$) when temporal acuity is included in the model first. Together, ΔT and the TRT explain 73% of the variance in speech intelligibility in modulated noise. When they are both included in the model, no other predictors significantly contribute anymore ($p>0.09$).

Thus, the regression analysis shows that a nonauditory component, as measured by the TRT, does appear to significantly contribute to explaining variance in SRT_{MOD} for the hearing-impaired subjects, in contrast with what was to be expected from the correlations shown in Table II. Even though the correlation between the TRT and SRT_{MOD} ($r=0.42$) is nonsignificant, the TRT is significantly associ-

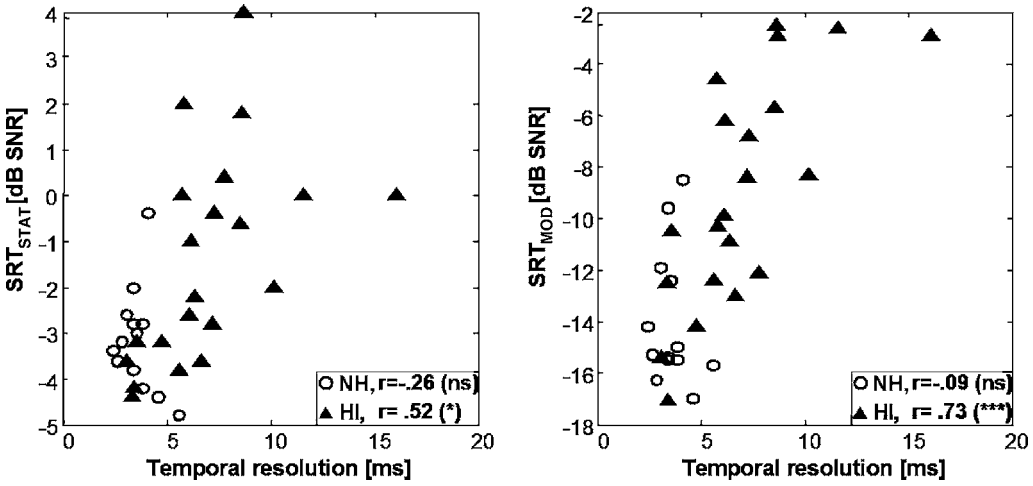


FIG. 4. SRT in stationary noise (SRT_{STAT}, left) and in 16-Hz block-modulated noise (SRT_{MOD}, right) as a function of temporal resolution (ΔT), for the normal-hearing (NH, $N=13$) and the hearing-impaired (HI, $N=21$) participants. Correlations and significances indicated as in Fig. 3.

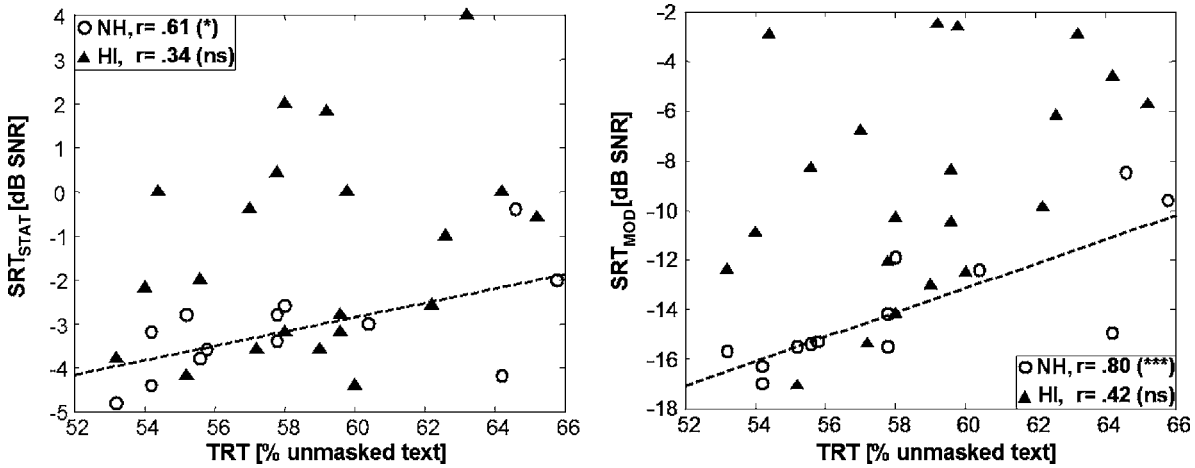


FIG. 5. SRT in stationary noise (SRT_{STAT}, left) and in 16-Hz block-modulated noise (SRT_{MOD}, right) as a function of text reception threshold for text masked with a vertical bar pattern (TRT), for the normal-hearing (NH, $N=13$) and the hearing-impaired (HI, $N=21$) participants. The dashed lines are linear regression lines fitted on the data of the normal-hearing participants. Correlations and significances indicated as in Fig. 3.

ated with SRT_{MOD} when the effect of temporal acuity is partialized out by the first step in the regression analysis.

In summary, there is a difference in the relative contribution of factors accounting for variance in speech reception between normal-hearing and hearing-impaired listeners. When there are no or only minor auditory problems, as in the normal-hearing participants, variance in speech reception in noise is mainly governed by nonauditory factors. However, when variance in auditory processing increases as a consequence of deterioration in the auditory system, interindividual differences in speech reception appear to be governed by both auditory and nonauditory factors.

IV. DISCUSSION

A. Differences between groups

Our experimental results indicate that nonauditory factors, as measured by the TRT, are the most important source of variance in speech reception for the normal-hearing listeners. For the hearing-impaired listeners, both auditory and nonauditory factors appear to influence intersubject differences in speech reception.

This observed difference in contribution of auditory and nonauditory factors can be explained by the differences in ranges of the auditory and nonauditory factors between the two groups. Participants in the normal-hearing group all perform reasonably well on auditory factors, i.e., the range of, for instance, PTA and ΔT is only fairly small compared to the range in the hearing-impaired group. This difference in ranges contributes to the observed relatively small correlation between auditory factors and SRT for normal-hearing participants. Thus, the relatively small range of auditory factors for the normal-hearing group, compared to the larger range of auditory factors in the hearing-impaired group, may be considered the main reason for the apparent difference in the relative contribution of auditory and nonauditory factors between both groups.

In contrast, the range of the nonauditory TRT is about the same for both groups, making the TRT more likely to contribute to variance in both groups. Indeed, the TRT contributes to explaining variance in SRT in both groups, especially in modulated noise. Nevertheless, the observed correlation between TRT and SRT_{MOD} is smaller for the hearing-impaired group, due to the influence of auditory factors, which explain the main part of the variance. Thus, the relative contribution of the TRT to speech reception is larger in the normal-hearing group (explaining 60% of the variance) than in the hearing-impaired group (25% after temporal acuity is partialized out in the first step of the regression analysis). However, these relative contributions can be expressed in absolute terms by multiplying them with the total amount of unexplained variance (i.e., the squared standard deviation) within a group. In absolute terms, then, the TRT explains a similar amount of variance in SRT_{MOD} in both groups

TABLE III. Product-moment cross correlations between the predictor variables. Only data from the hearing-impaired participants ($N=21$) were included in the calculations. Significant correlations are displayed in bold, p values are denoted by asterisks, indicated as in Table II.

		Auditory			Nonauditory	
		ΔF	ΔT	PTA	Age	TRT
Auditory	ΔF	⋯				
	ΔT	0.42	⋯			
	PTA	−0.28	**0.59**$^{(**)}$	⋯		
Nonauditory	Age	−0.22	0.09	0.42	⋯	
	TRT	−0.35	−0.08	0.35	0.28	⋯

TABLE IV. Results from the stepwise multiple regression analyses: significant contributors to explaining the variance in SRT in stationary noise (SRT_{STAT}) and SRT in modulated noise (SRT_{MOD}). The analysis was performed only on the data from the hearing-impaired participants ($N=21$). Shown are the successive contributing predictor variables and the percentage of the variance that they significantly account for when included in the model, either on their own (R^2) or cumulatively combined (cum. R^2). Both measures are corrected for the available degrees of freedom. All shown models have a significance $p<0.001$.

	SRT_{STAT}			SRT_{MOD}		
	Predictor	R^2	cum. R^2	Predictor	R^2	cum. R^2
Step 1	PTA	0.47	0.47	ΔT	0.48	0.48
Step 2	···	···	···	TRT	0.09	0.73

($4.2\ dB^2$ in the normal-hearing and $4.9\ dB^2$ in the hearing-impaired group). This finding indicates that nonauditory factors affect speech reception in modulated noise, independent of the amount of hearing loss.

B. Differences between stationary and modulated noise

Considering the results within the hearing-impaired group, there seems to be a fundamental difference between the factors that explain variance in speech intelligibility in either stationary or modulated noise. The present results show that speech reception in stationary noise is mainly governed by auditory factors, while both auditory and nonauditory factors account for intersubject variance in modulated noise.

In stationary noise, PTA is the main factor accounting for intersubject variance in speech intelligibility. This is in agreement with results by Van Rooij and Plomp (1992), who concluded that the audiogram is the most adequate predictor of speech reception in stationary noise. In modulated noise, temporal acuity (ΔT) and TRT appear to be the main factors explaining intersubject differences in speech reception. This does not mean that PTA does not correlate with SRT in modulated noise, as can be seen in Table II ($r=0.73$). It does mean, however, that PTA does not explain as much variance in SRT_{MOD} as temporal acuity and the TRT do together, and that, in the regression analysis, the effect of PTA is covered by these variables.

These findings suggest that PTA is an estimate for general auditory performance and, as such, is related to speech reception in both stationary and modulated noise. Measuring speech reception in modulated noise, showing larger interindividual differences, apparently enables specification of the variance in speech reception into temporal auditory processing (temporal acuity) and nonauditory factors (TRT). Moreover, measuring the SRT in modulated noise appears to increase discrimination between normal-hearing and hearing-impaired participants, as indicated by the t-statistics in Table I. Therefore, modulated noise may be preferred over stationary noise to measure speech reception for clinical purposes, specifically because nonstationary backgrounds are more common in everyday situations (Kramer *et al.*, 1996).

C. Relation among speech reception (SRT), PTA, and presentation level

In the speech reception tasks in the current experiment, individual hearing thresholds were used to adapt the spectrum of the background noise and the speech signal, i.e., the noise spectrum was fixed halfway up the dynamical range for each participant. This method optimizes audibility at all frequencies for each participant, but, consequently, gives rise to interindividual spectrum and level differences. As mentioned earlier, the average rms level of each of the two maskers was 70.9 dB(A) for the normal-hearing group and 92.6 dB(A) for the hearing-impaired group. Using the SRT to quantify the subject's ability to perceive speech in noise does not take these differences between listener groups into account.

Moreover, optimizing audibility for each listener has as a direct consequence that presentation level and hearing threshold are related, and that their effects on speech reception cannot be fully distinguished. This means that part of the correlation between the audiogram (or PTA) and the SRT, as reported earlier, might be attributed to the effects of presentation level. Indeed, overall presentation level, expressed in dB(A), is significantly related to SRT in the hearing-impaired group, both in stationary ($r=0.44$, $p=0.05$) and in modulated noise ($r=0.57$, $p=0.007$).

A measure of speech intelligibility performance which is able to handle intersubject audiogram and spectrum differences is the Speech Intelligibility Index or SII (ANSI S3.5-1997), which gives an estimate of the amount of speech information available in a certain condition, using the individual's audiogram and the signal and masker spectrum levels as inputs. In addition to accounting for threshold and spectrum differences, the SII model also includes a level distortion factor, which takes the deterioration of speech recognition at higher presentation levels into account. A SII of about 0.30–0.35 is commonly considered to be enough to reach 50% speech intelligibility for normal-hearing listeners. Hearing-impaired listeners generally need more speech information, which is regarded as the result of less efficient processing of the information due to suprathreshold deficits.

The variables SRT_{STAT} and SRT_{MOD} were tranferred to SII values by using the model as introduced and validated by Rhebergen and Versfeld (2005), which was also applied in our earlier study (George *et al.* 2006). Multiple stepwise regression analyses were performed as before, with SII_{STAT} and SII_{MOD} as dependent variables, to investigate which pre-

dictor variables could most effectively account for intersubject variance in SII. For easy comparison, the amounts of explained variance in SRT, as obtained before in Sec. III C, will be repeated in the following between square brackets.

Results of the stepwise regression analyses on SII values show that, in the normal-hearing group, the TRT is the predictor variable that contributes most to accounting for variance in SII values. On its own, the TRT accounts for 24% [31%] of the intersubject variance in SII_{STAT} ($r=0.55$, $p=0.05$) and for 58% [60%] of the variance in SII_{MOD} ($r=0.78$, $p=0.002$). When the TRT is included in the model, no other predictor significantly explains variance over and above the variance explained by the TRT ($p>0.43$ in both cases).

For the hearing-impaired group, only PTA significantly contributes to intersubject differences in SII, accounting for 21% [47%] of the variance in SII_{STAT} ($r=0.50$, $p=0.02$). When PTA is included in the model, no other predictors are significantly correlated with the residual ($p>0.26$). In modulated noise, temporal acuity (ΔT) contributes most to explaining variance in speech reception in the hearing-impaired group, accounting for 49% [48%] of the intersubject variance in SII_{MOD} ($r=0.72, p<0.001$). The TRT is not significantly related to SII_{MOD} ($p=0.16$), but becomes a significant term ($p=0.01$) when temporal acuity is included in the model first. Together, ΔT and the TRT explain 62% [73%] of the variance in SII in modulated noise. When they are both included in the model, no other predictors significantly contribute anymore ($p>0.30$).

The results of the regression analyses on SII values are comparable to the earlier obtained results on SRT values. The largest difference is the reduced amount of variance explained by PTA in the hearing-impaired group in stationary noise. This was to be expected, since it is a direct consequence of taking interindividual audibility, spectrum, and level differences into account. Even though the SII may underestimate the deteriorating effects of level on speech reception, as argued by Studebaker *et al.* (1999), the correlation between overall presentation level and speech reception, as observed before in the hearing-impaired group, is not significant anymore when the SII values are considered ($r=0.38$, $p=0.09$ in stationary noise; $r=0.32$, $p=0.15$ in modulated noise).

Within the current set of predictor variables, PTA is still the best (least poor) predictor of speech reception (SII) in stationary noise for hearing-impaired listeners, even though level and audibility difference have been taken into account. This confirms our earlier suggestion that the correlation between PTA and speech reception is not related to audibility differences. Instead, it can be understood by considering PTA as a good estimate for general auditory performance: the development of hearing loss (PTA) generally accompanies deterioration of suprathreshold processing, and vice-versa. As shown in Table III, PTA is indeed significantly related to temporal acuity, and it is not unlikely that a larger PTA also reflects deteriorated intensity coding, recruitment (Stephens, 1976), or the loss of normal auditory compression. In this light, it is understandable that PTA relates to SRT in stationary noise, even when audibility effects are accounted for.

In summary, these SII results corroborate with our earlier conclusions that, in the normal-hearing group, variance in speech reception in noise is mainly governed by nonauditory factors. In the hearing-impaired group, interindividual differences in speech reception appear to be governed by both auditory and nonauditory factors, especially in modulated noise.

D. Relation among speech reception (SRT) and ΔF, ΔT, and age

The observed noncontribution of deteriorated spectral resolution (ΔF) to variance in speech reception does not appear to be consistent with literature (Patterson *et al.*, 1982; Noordhoek *et al.*, 2001). This may be explained, however, by the fact that deteriorated spectral acuity for mild to moderate hearing-impaired listeners may be largely accounted for by increased presentation level, as demonstrated by George *et al.* (2006). When presentation level is accounted for, differences in spectral resolution between normal-hearing and hearing-impaired participants are only minor, making spectral acuity a poor candidate to explain variance in speech reception (see the nonsignificant correlations in Table II). In contrast, Table II shows that reduced temporal resolution (ΔT) is significantly related to speech reception in both stationary and modulated noise. In stationary noise, there is a significant correlation between ΔT and SRT, but the regression analysis shows that this effect disappears after taking PTA into account, while it is just the other way around in modulated noise. This can be explained by the fact that, in modulated noise, the effect of deteriorated temporal resolution is more prominent. That is, sufficient temporal acuity is necessary for a listener to take advantage of the relatively silent periods or gaps in modulated noise (Glasberg *et al.*, 1987; Festen and Plomp, 1990; Peters *et al.*, 1998; Snell *et al.*, 2002).

Age does not appear to be significantly related to speech reception in noise, as implied by the correlations in Table II and by results of the regression analyses. This finding does not appear to be in line with earlier findings, which did report a detrimental effect of age on speech reception (e.g., Gustafsson and Arlinger, 1993; Snell *et al.*, 2002). One possible explanation for this noncontribution of age might be that age-related deficits are most apparent in complex auditory tasks or backgrounds (Gordon-Salant, 1987; Souza and Turner, 1994; Pichora-Fuller and Souza, 2003). That is, perhaps our listening conditions were not challenging enough for the participants. However, this explanation seems unlikely, since the adaptive procedure used in the speech reception task varies the signal-to-noise ratio around the point at which the listener reaches 50% sentence intelligibility, thus avoiding "easy listening conditions." A more likely explanation is found by suggestions in literature (Pichora-Fuller *et al.*, 1995; Watson *et al.*, 1996; Gordon-Salant and Fitzgibbons, 1997; Gatehouse *et al.*, 2003) that the effect of age on speech reception in noise may be mediated by the influence

of cognitive effects on hearing ability. In fact, in the current experiment, age was significantly related to SRT_{MOD} ($p=0.012$) when temporal acuity was partialized out by the first step in the regression analysis for hearing-impaired listeners. The correlation between TRT and SRT_{MOD}, however, was larger ($p=0.0006$), giving rise to a larger cumulatively explained amount of variance. When TRT is included in the regression model, the contribution of age is not significant anymore. This implies that the inclusion of TRT in the model reduces age-related variance in SRT_{MOD}. Thus, even though the relationship between age and TRT was not statistically significant, they share a common component related to speech reception, which is better expressed in terms of nonauditory factors (TRT) than in terms of age.

Finally, it should be noted that the observed percentages of variance accounted for (R^2), as mentioned in Sec. III, may be underestimates, considering the nonunity reliabilities of the predictor variables, as shown in Table I. The R^2 were substantial, although only a relatively small number of participants was included in the current experiment, which may indicate that there may be little systematic variance left to be explained over and above the variance explained by the predictor variables included in the present study.

E. Relation between speech reception (SRT) and text reception (TRT)

The TRT does not only contribute significantly to explaining variance in SRT_{MOD} in the hearing-impaired group, but is also the main source of variance in speech reception in the normal-hearing group. The obtained correlation in this group between SRT_{MOD} and TRT ($r=0.80$) suggests that the perception of speech in nonstationary noise and of masked text depend partly on the same modality-aspecific processing mechanisms. This finding confirms the results of Zekveld *et al.* (2007), who observed this relationship between SRT and TRT in a normal-hearing group with a wider age range. Moreover, this result is in line with results of, for instance, Amitay *et al.* (2002), who demonstrated the relationship between spoken and written language comprehension in a study on reading disabilities. Consistent with the suggestions of Grant *et al.* (1998) and Zekveld *et al.* (2007), we suggest that this shared component reflects a modality-independent skill to perform perceptual closure, i.e., the process by which the masked portions of a stimulus are completed in order to identify the object (Snodgrass and Kinjo, 1998). Specifically, the closure as meant here is related to the extent to which listeners restore missing phonemes or words by using sentence context, i.e., the redundancies in speech at the acoustic, phonetic, phonological, and/or lexical level (Warren, 1970).

Consistent with findings by Grant and Seitz (2000), it was observed that this ability, as expressed by the TRT, may vary substantially across hearing-impaired subjects. Apparently, some hearing-impaired subjects are better at using sentence context for speech recognition than others. This may enable the application of the TRT for clinical purposes.

To clarify this point, Fig. 5 should be reconsidered. It illustrates the results from the stepwise regression analyses, showing that for both the normal-hearing and the hearing-impaired participants, the deteriorating SRTs in the two noise conditions appear to be associated, to some extent, with deteriorating TRT. For an individual normal-hearing participant, the SRT is mainly governed by nonauditory factors, that also govern the TRT. Hence, for normal-hearing participants, the relationship between SRT and TRT is relatively strong. The dotted linear regression lines represent this relationship in Fig. 5. For hearing-impaired participants, the relation between the TRT and SRT in noise appears less straight-forward. The normal-hearing regression lines might be considered as a "limit of performance" or "baseline" for hearing-impaired subjects. When a data point is close to the regression line, the SRT can be considered essentially normal, given the TRT, or, to put it differently, the SRT score can be considered to be solely related to nonauditory factors. An individual's vertical deviation from the normal-hearing regression line may then be thought of as reflecting the relative contribution of deteriorated auditory factors to the elevated SRT. The stepwise regression results indicate that, in modulated noise, this deviation is related to deteriorated temporal acuity (ΔT).

Thus, the combined measurements of SRT_{MOD} and TRT makes it possible to estimate the relative contribution of auditory and nonauditory factors to speech recognition in noise. This may enable the clinical examiner to determine in part the origin (auditory or nonauditory) of deteriorated speech reception, using the TRT test as an additional diagnostic clinical tool. A listener with a relatively poor TRT score will be less able to use context information to improve sentence readability or intelligibility. Thus, to reach equal speech intelligibility, this listener will need a relatively higher signal-to-noise ratio compared to a listener with a better TRT score. The expectations concerning the benefit from auditory rehabilitation by, for instance, hearing-aids, can then be adapted likewise (see Pichora-Fuller and Souza, 2003).

The success of clinical application of the combination of TRT and SRT_{MOD} does, however, depend on the assumption that the relative contributions of auditory and nonauditory factors to speech reception do not change with deteriorating auditory performance or age. Differently said, it is assumed that auditory and nonauditory factors are independently used to process speech, such that the deterioration of auditory processing does not change the relative contribution of nonauditory factors to speech reception.

Although it has been suggested by Pichora-Fuller *et al.* (1995) that auditory and nonauditory (cognitive) functions are linked in a processing model, there is as yet no reason to doubt this assumption. The fact that the TRT accounts for the major part of the variance in speech reception in the normal-hearing group suggests that the TRT is capable of measuring modality-aspecific functions associated with the reception of speech. Moreover, results in Table III show that both temporal acuity and the TRT are independent of age. In addition, they are mutually independent. Finally, the amount of variance explained by the TRT, expressed in absolute terms, is similar for the normal-hearing and the hearing-impaired group (see Sec. IV A). This indicates an independent utilization of auditory and nonauditory cues, in

accordance with results by Van Rooij and Plomp (1992). Nevertheless, further research should clarify whether the above-mentioned assumption is valid.

Finally, the text reception threshold should be more widely validated for normal-hearing listeners to determine the sources responsible for the variance in TRT. It is now assumed to be a general measure of nonauditory factors relevant for speech reception, presumably related to a modality-independent skill to perform perceptual closure. Recently, however, a study has started in our laboratory investigating the relation between the SRT and several variants of the TRT for listeners with Dutch as a second language, who are clearly less experienced in applying semantic, syntactic, or lexical information to improve the readability or recognition of Dutch sentences. Future studies like these may enable specification of the TRT measure into less general cognitive or linguistic skills which contribute to speech perception.

V. CONCLUSIONS

The main results arrived at in this study can be summarized as follows:

(I) The two groups of normal-hearing and hearing-impaired participants, matched for age, do not perform significantly differently on a nonauditory test (TRT). Differences in speech reception between both groups are thus likely to be mainly governed by auditory factors.

(II) Differences in speech reception (SRT) between the normal-hearing and hearing-impaired participants are more prominent in modulated noise than in stationary noise. Therefore, modulated noise may be preferred over stationary noise to measure speech reception for clinical purposes, i.e., to assess whether speech recognition in daily life is deteriorated compared to normal.

(III) For the normal-hearing listeners, nonauditory factors, as measured by the TRT, are the most important source of variance in the SRT, both in stationary and in modulated noise.

(IV) For the hearing-impaired participants, interindividual differences in speech reception in stationary noise are mainly governed by auditory factors, in particular by the audiogram. In contrast, both auditory (temporal resolution) and nonauditory (TRT) factors account for intersubject variance in speech reception in modulated noise.

(V) The combined measurement of the SRT in modulated noise and the TRT may be clinically relevant to determine part of the origin (auditory or nonauditory) of deteriorated speech reception, possibly adapting the expectations from auditory rehabilitation.

ACKNOWLEDGMENTS

This research was supported by the Heinsius-Houbolt Foundation, The Netherlands. Thanks are due to Hans van Beek for technical support and for his programming work on the TRT. Nienke Versteeg and Anushka Lala are thanked for their contributions to this experiment.

Amitay, S., Ahissar, M., and Nelken, I. (**2002**). "Auditory processing deficits in reading disabled adults," J. Assoc. Res. Otolaryngol. **3**, 302–320.

ANSI (**1997**). "American national standard methods for the calculation of the Speech Intelligibility Index," ANSI S3.5-1997, American National Standards Institute, New York.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807–1818.

Divenyi, P. L., Stark, P. B., and Haupt, K. M. (**2005**). "Decline of speech understanding and auditory thresholds in the elderly," J. Acoust. Soc. Am. **118**, 1089–1100.

Dubno, J. R., Ahlstrom, J. B., and Horwith, A. R. (**2000**). "Use of context by young and aged adults with normal hearing," J. Acoust. Soc. Am. **107**, 538–546.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal-hearing," J. Acoust. Soc. Am. **113**, 2084–2094.

Dubno, J. R., and Schaefer, A. B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110-2121.

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**, 537–544.

Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Hear. Res. **38**, 222–233.

Elliott, L. L. (**1995**). "Verbal auditory closure and the Speech Perception in Noise (SPIN) test," J. Speech Hear. Res. **38**, 1363–1376.

Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by an interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.

Fitzgibbons, P. J. (**1983**). "Temporal gap detection in noise as a function of frequency bandwidth, and level," J. Acoust. Soc. Am. **74**, 67-72.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1987**). "Temporal gap resolution in listeners with high-frequency sensorineural hearing loss," J. Acoust. Soc. Am. **81**, 133-137.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal-hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Gatehouse, S., Naylor, G., and Elberling, C. (**2003**). "Benefits from hearing aids in relation to the interaction between the user and the environment," Int. J. Audiol. **42**, Suppl. 1: S77–85.

George, E. L. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.

Gifford, R. H., and Bacon, S. P. (**2005**). "Psychophysical estimates of nonlinear cochlear processing in younger and older listeners," J. Acoust. Soc. Am. **118**, 3823–3833.

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Glasberg, B. R., and Moore, B. C. J. (**1992**). "Effects of envelope fluctuations on gap detection," Hear. Res. **64**, 81–92.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. (**1987**). "Gap detection and masking in hearing-impaired and normal-hearing subjects," J. Acoust.

Soc. Am. **81**, 1546–1556.
Goldstein, E. B. (**2002**). *Sensation and Perception* (Wadsworth-Thomson, Pacific Grove, CA).
Gordon-Salant, S. (**1987**). "Age-related differences in speech recognition performance as a function of test format and paradigm," Ear Hear. **8**, 277–282.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.
Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808–1817.
Grant, K. W., and Seitz, P. F. (**2000**). "The recognition of isolated words and words in sentences: Individual variability in the use of sentence context," J. Acoust. Soc. Am. **107**, 1000–1011.
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.
Grose, J. H., Hall, J. W., and Buss, E. (**2001**). "Gap duration discrimination in listeners with cochlear hearing loss: Effects of gap and masker duration, frequency separation, and mode of presentation," J. Assoc. Res. Otolaryngol. **2**, 388–398.
Gustafsson, H. A., and Arlinger, S. D. (**1993**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.
Hällgren, M. (**2005**). "Hearing and cognition in speech comprehension," Doctoral thesis, Linköping University, Sweden.
Hilkhuysen, G. L. M., Houtgast, T., and Lyzenga, J. (**2005**). "Estimating cochlear-filter shapes, temporal-window width and compression from tone-sweep detection in spectral and temporal noise gaps," J. Acoust. Soc. Am. **117**, 2598–2599.
Humes, L. E. (**2002**). "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am. **112**, 1112–1132.
Humes, L. E. (**2005**). "Do 'auditory processing' tests measure auditory processing in the elderly?," Ear Hear. **26**, 109–119.
International Organization for Standardization (**1991**). "Acoustics-Standard reference zero for the calibration of pure-tone air conduction audiometers," ISO 389:1991(E). Available from the American National Standards Institute, New York.
Ishihara, S. (**1989**). *The Series of Plates Designed as a Test for Colour Deficiency* (Kanehara, Tokyo).
Jestead, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950-962.
Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.
Kramer, S. E., Kapteyn, T. S., Festen, J. M., and Tobi, H. (**1996**). "The relationships between self-reported hearing disability and measures of auditory disability," Audiology **35**, 277–287.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Ludvigsen, C. (**1985**). "Relations among some psychoacoustic parameters in normal and cochlearly impaired listeners," J. Acoust. Soc. Am. **78**, 1271–1280.
Lunner, T. (**2003**). "Cognitive function in relation to hearing aid use," Int. J. Audiol. **42**, Suppl. 1: S49–58.
Miller, R. G. (**1981**). *Simultaneous Statistical Inference* (Springer, New York).
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psycho-acoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.
Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.
Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.
Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**2001**). "Relations between intelligibility of narrow-band speech and auditory functions, both in the 1-kHz frequency region," J. Acoust. Soc. Am. **109**, 1197–1212.
Nunnally, J. C. (**1967**). *Psychometric Theory* (McGraw-Hill, New York).
Oxenham, A. J., and Bacon, S. P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear. **24**, 352–366.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Peters, R. W., Moore, B. C. J., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.
Pichora-Fuller, M. K., and Souza, P. E. (**2003**). "Effects of aging on auditory processing of speech," Int. J. Audiol. **42**, 2S11–2S16.
Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.
Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.
Rhebergen, K. S., and Versfeld, N. J. (**2005**). "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. **117**, 2181–2192.
Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.
Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). "Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity," J. Acoust. Soc. Am. **112**, 720–727.
Snodgrass, J. G., and Kinjo, H. (**1998**). "On the generality of the perceptual closure effect," J. Exp. Psychol. Learn. Mem. Cogn. **24**, 645–658.
Sommers, M. S., and Humes, L. E. (**1993a**). "Auditory filter shapes in normal-hearing, noise-masked normal and elderly listeners," J. Acoust. Soc. Am. **93**, 2903-2914.
Sommers, M. S., and Humes, L. E. (**1993b**). "Erratum: Auditory filter shapes in normal-hearing, noise-masked normal and elderly listeners [J. Acoust. Soc. Am. **93**, 2903-2914 (1993)]," J. Acoust. Soc. Am. **94**, 2449-2450.
Souza, P. E., and Turner, C. W. (**1994**). "Masking of speech in young and elderly listeners with hearing loss," J. Speech Hear. Res. **37**, 655–661.
Stephens, S. D. G. (**1976**). "The input for a damaged cochlea—A brief review," Br. J. Audiol. **10**, 97–101.
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.
Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.
Ter Keurs, M., Festen, J. M., and Plomp, R. (**1993a**). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547–1552.
Ter Keurs, M., Festen, J. M., and Plomp, R. (**1993b**). "Limited resolution of spectral contrast and hearing loss for speech in noise," J. Acoust. Soc. Am. **94**, 1307–1314.
Van Rooij, J. C. G. M., and Plomp, R. (**1990**). "Auditory and cognitive factors in speech perception by elderly listeners. II. Multivariate analysis," J. Acoust. Soc. Am. **88**, 2611–2624.
Van Rooij, J. C. G. M., and Plomp, R. (**1992**). "Auditory and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion," J. Acoust. Soc. Am. **91**, 1028–1033.
Van Rooij, J. C. G. M., Plomp, R., and Orlebeke, J. F. (**1989**). "Auditory and cognitive factors in speech perception by elderly listeners. I. Development of test battery," J. Acoust. Soc. Am. **86**, 1294–1309.
Van Wijngaarden, S. J., Bronkhorst, A. W., Houtgast, T., and Steeneken, H. J. (**2004**). "Using the Speech Transmission Index for predicting non-native speech intelligibility," J. Acoust. Soc. Am. **115**, 1281–1291.
Van Wijngaarden, S. J., Steeneken, H. J., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.
Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.
Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393.

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.

Zekveld, A. A., George, E. L. J., Kramer, S. E., Goverts, S. T., and Houtgast, T. (**2007**). "The development of the Text Reception Threshold test: A visual analogue of the Speech Reception Threshold test," J. Speech Lang. Hear. Res. (to be published).

Discrimination of interval size in short tone sequences

Toby J. W. Hill and Ian R. Summers[a)]
Biomedical Physics Group, School of Physics, University of Exeter, Exeter EX4 4QL, United Kingdom

(Received 2 August 2006; revised 16 November 2006; accepted 21 January 2007)

This study investigates the discrimination of small changes of interval size in short sequences of musical tones. Major, minor and neutral thirds were varied in increments of 15 cents. The nine subjects had varying degrees of amateur musical experience—their level of musical training was lower than that of professional musicians. In some experiments the stimuli were presented purely melodically and in others they were presented together with a sustained tone at a higher pitch. Some subjects were able to make use of the additional cues from beats in the latter case. Category widths for identification were measured at around 70 cents and just-noticeable differences in frequency were measured at around 10 cents. Little significant variation of inter-stimulus sensitivity index d' was observed across the stimulus sets, i.e., there was little evidence for "anchors" or "landmarks" within the range of tunings employed. However, for major thirds, discrimination of the 15 cent increment between 400 and 415 cents was reduced compared to discrimination of other 15 cent increments within the stimulus sets. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2697059]

PACS number(s): 43.75.Cd, 43.75.Bc, 43.66.Hg, 43.66.Fe [DD] Pages: 2376–2383

I. INTRODUCTION

In Western music, the intervals used within the musical scale (minor second, major second, minor third, etc.) may be subject to small variations of tuning according to particular performance practices (Rasch, 1983). For example, when using equal temperament (based on a logarithmic division of the octave into 12 equal steps) the major third is intended to be 400 cents; when using a tuning scheme with a "just" major third (i.e., a fundamental-frequency ratio of 5:4) this interval is intended to be 386 cents. (Note: One cent is the unit obtained by logarithmic division of the octave into 1200 equal steps.) The present study is concerned with the extent to which such small changes in tuning are apparent to the listener.

The literature documents a large number of studies related to perception of the tuning of musical intervals. Experiments cover measurements of just-noticeable difference in frequency for pure and complex tones (e.g., Harris, 1952; Nelson *et al.*, 1983; Moore and Glasberg, 1990; Sek and Moore, 1995), measurements of consonance and dissonance (e.g., Kameoka and Kuriyagawa, 1969a, 1969b; Tufts *et al.* 2005), and a wide range of topics relating to the perception of tone sequences and information transfer via such sequences (Watson *et al.*, 1975, 1976; Deutsch, 1980; Spiegel and Watson, 1984; Kidd and Watson, 1992; Schellenberg and Trehub, 1994; Parncutt and Cohen, 1995; Thompson *et al.*, 2001; Creel *et al.*, 2004; Smith and Schmuckler, 2004).

Several previous studies have investigated perception of the tuning of melodic or harmonic intervals presented in isolation (e.g., Burns and Ward, 1978; Hall and Hess, 1984; Vos, 1986; Burns and Campbell, 1994). Equivalent experiments in the context of "real" music are problematical—the inherent complexity of the musical material and the listener's cognitive response means that subjects' performance is difficult to analyze. Hence, investigators (e.g., Rasch, 1985; Vos, 1988) have chosen to work on musical fragments or short tone sequences which provide a quasi-musical context for measurements, while avoiding undue complexity.

Rasch (1985) used musical fragments in the form of two simultaneous tone sequences, with experiments involving the mistuning of tones in one or both of the sequences. Melodic (interval-width) and harmonic (beat) cues were found to contribute to the detection of mistuning. Vos (1988) used similar musical fragments, with experiments involving rating and paired comparison of six common intonation systems. Overall acceptability was found to relate to the "purity" of the intervals (i.e., their closeness to just intonation), suggesting that harmonic (beat) cues were dominant in this case.

An aim of the present study is to further explore this "middle ground" between experiments on single intervals in isolation and experiments based on real music. Four experiments have been carried out using short tone sequences to provide a quasi-musical context. Identification tasks were used to determine subjects' ability to detect small changes in tuning. There is evidence (Burns and Ward, 1978; Burns and Campbell, 1994; Perlman and Krumhansl, 1996) that the perceptual "landscape" contains "anchors" or "landmarks" at particular positions in the pitch range. For example, sensitivity to small pitch changes might be less around unfamiliar intervals such as the "neutral" third (approximately 350 cents) and greater around familiar intervals such as the major and minor thirds (equal tempered at 400 and 300 cents and just intonation at 386 and 316 cents). However, other investigators (Parncutt and Cohen, 1995) have not observed this effect. In the present study, stimuli were designed to investigate the possibility of anchors or landmarks within the range of tunings employed.

[a)]Author to whom correspondence should be addressed. Electronic mail: i.r.summers@exeter.ac.uk

 0001-4966/2007/121(4)/2376/8/$23.00

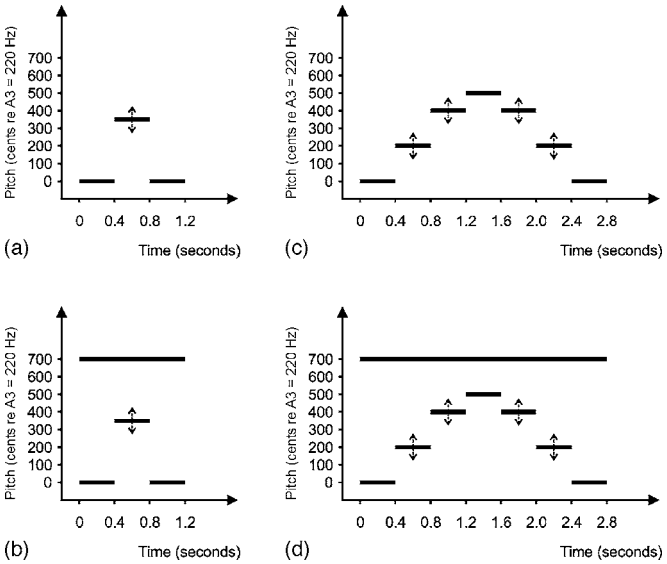


FIG. 1. Schematic pitch-time diagrams for the tone sequences employed in the four experiments: (a) Experiment 1, melodic third; (b) Experiment 2, harmonized third; (c) Experiment 3, melodic scale-like sequence; (d) Experiment 4, harmonized scale-like sequence. The double-headed arrows indicate tones whose pitch was varied.

II. METHOD

A. Overview

This study includes four related experiments. In each experiment subjects were presented with short sequences of musical tones, each sequence being identical except for small changes in the tuning of the component tones. The task was to identify the tuning system used for each sequence—a choice from five alternatives. As indicated above, the experimental design is intended to provide a quasi-musical context within which specific psychophysical measurements may be made. It is not the intention to measure pre-learned notions of "good" or "bad" intonation, but rather to see whether particular intonations are perceptually more distinct (perhaps as a result of being good or bad).

The tone sequences are illustrated in Fig. 1. For Experiment 1 each sequence was of three tones, rising and then falling by an interval of a third. For Experiment 2 these three-note sequences were accompanied by a sustained tone at a higher pitch (at an interval of a perfect fifth above the start note of the sequence). For Experiment 3 each sequence was of seven tones, rising and falling in a scale-like manner. For Experiment 4 these scale-like sequences were accompanied by a sustained tone at a higher pitch (again, at an interval of a perfect fifth above the start note of the sequence). The double-headed arrows in Fig. 1 indicate the tones whose pitch was varied to achieve the different tuning systems—the 2nd tone within each sequence of three in Experiments 1 and 2, and the 2nd, 3rd, 5th and 6th tones within each sequence of seven in Experiments 3 and 4. Experiments 3 and 4 were intended to provide a more complex task than Experiments 1 and 2, but with additional cues. An important issue in relation to experiments on intonation is the relative importance of melodic (or interval-width) and harmonic (or beat) cues in the perception of the various stimuli. The intention in the present study was to provide melodic cues to pitch changes in Experiments 1 and 3 and both harmonic and melodic cues in Experiments 2 and 4.

Details of the stimulus tunings are given in Tables I and II. The experimental pitches were designed in sets of five based around the equal-tempered major third (400 cents), the equal-tempered minor third (300 cents) and the neutral third (350 cents). Preliminary pilot studies indicated that increments of 15 cents would not only yield meaningful results but would also allow close approximations of the just-intonation major and minor thirds (386 and 316 cents, respectively) to be incorporated. In the tone sequences for Experiments 3 and 4 the tones at a major second above the start note (i.e., tones 2 and 6 in the sequence—see Fig. 1) were tuned so as to bisect the interval of the major third above the start note (see Table II), following the scheme adopted in many tuning systems by which a major third contains two equal-size major seconds (Rasch, 1983).

TABLE I. Stimulus tunings used in Experiments 1 and 2. Each variant of the test involves five stimuli. Stimuli are presented melodically in Experiment 1 and accompanied by a higher-pitch tone in Experiment 2.

Test variant	Stimulus label	Tone 1 in sequence (cents re A3)	Tone 2 in sequence (cents re A3)	Tone 3 in sequence (cents re A3)	Higher pitch in Experiment 2 (cents re A3)
Major 3rd	*a*	0	430	0	700
	b	0	415	0	700
	c	0	400	0	700
	d	0	385	0	700
	e	0	370	0	700
Neutral 3rd	*ℓ*	0	380	0	700
	m	0	365	0	700
	n	0	350	0	700
	o	0	335	0	700
	p	0	320	0	700
Minor 3rd	*v*	0	330	0	700
	w	0	315	0	700
	x	0	300	0	700
	y	0	285	0	700
	z	0	270	0	700

TABLE II. Stimulus tunings used in Experiments 3 and 4. Each test involves five stimuli. Stimuli are presented melodically in Experiment 3 and accompanied by a higher-pitch tone in Experiment 4.

Stimulus label	Tones 1, 7 in sequence (cents re A3)	Tones 2, 6 in sequence (cents re A3)	Tone 3, 5 in sequence (cents re A3)	Tone 4 in sequence (cents re A3)	Higher pitch in Experiment 4 (cents re A3)
a	0	215	430	500	700
b	0	207.5	415	500	700
c	0	200	400	500	700
d	0	192.5	385	500	700
e	0	185	370	500	700

The wave forms used for the tones in these experiments were designed to have an unambiguous pitch, to sound "musical" but not like any instrument in particular, and to be of a duration (400 ms) representative of "real" music. The pitch on which the intervals were constructed was A3, with a fundamental frequency of 220 Hz. This pitch was found to be relatively comfortable to listen to repetitively and is in the midrange of pitches used melodically in music. Moreover, it has been suggested that the perception of tones at this pitch is representative of perception over a wide range of other musical pitches (Vos and van Vianen, 1985). Each tone comprised ten frequency components, the fundamental and the next nine harmonics, with a spectral amplitude envelope falling at 6 dB oct^{-1} [equivalent to 1/(harmonic number)]. The tones had rise and release times of 40 ms, giving them a bland, organ-like timbre. Stimuli were specified in software, using a 40 kHz sample rate.

B. Apparatus

The experiments were conducted in a room designed for audiological purposes, with low reverberation. The subjects were tested singly and were seated in the middle of the room in front of a computer monitor and keyboard on which they received instructions and entered their responses. The experiments were self-paced by the subject. The stimuli were presented to the subject via two loudspeakers, which carried identical signals, positioned on either side of the monitor. The sound level at the subject's head was measured to be 62 dB(A) for the melodic stimuli and 64 dB(A) for the harmonized stimuli. This represents a comfortable listening level.

C. Subjects

Subjects were unpaid volunteers, predominantly graduate students in physics, with varying degrees of amateur musical experience. Their level of musical training was lower than that of professional musicians. The same nine subjects, eight male and one female, participated in each of the four experiments. Subjects' ages ranged from 20 to 45. None of the subjects reported any hearing impairment, and none possessed the ability of absolute pitch determination.

D. Paradigm and protocol

A single-interval, five-alternative, forced-choice identification paradigm was employed in each of the four experiments. Subjects were asked to identify the tuning system employed for each of a series of melodic fragments, using a labeling system (see Tables I and II) which was demonstrated at the start of the experiment. The experimental protocol comprised a sequence of three sections that was repeated within each test session. First, in a demonstration block, subjects were presented with the stimuli and their identifying labels (e.g., ℓ, *m*, *n*, *o*, *p*—each variant of the test involved five stimuli). Subjects were able to repeat this demonstration block on request. Second, subjects received a training block in which they identified stimuli and received trial-by-trial feedback and an overall score—the initial training block in a session comprised 25 items and all subsequent training blocks comprised ten items. At the end of a training block subjects were offered further demonstration or training or they could proceed to the third section of the protocol: the test blocks. Each test block comprised 25 stimuli without feedback, presented in a balanced, pseudo-random order in which each stimulus occurred five times. At the end of each test block subjects were informed of their performance, with a breakdown of score for each of the five stimuli. This protocol, which is similar to that reported by Mori and Ward (1995), was intended to assist subjects in self-motivation (by being aware of their performance) and in maintaining their decision criteria and concentration levels. The modular structure also facilitated the statistical assessment of learning effects. It was not the intention to intensively train subjects in order to measure their optimal performance—rather the intention was to overcome any asymptotic tendencies in the response function due to subjects gaining familiarity with the task and stimuli, in order to measure their latent ability to do the task.

Computer simulations (Hill, 2000) of absolute identification tasks suggested that a minimum of 40 trials per stimulus category (i.e., 200 trials in the case of five stimulus categories) are required per subject in order to satisfactorily extract values for information transfer and sensitivity index from individual confusion matrices (see following sections). Consequently, nine test blocks (each of 25 trials) were used for each variant (major third, neutral third and minor third) of Experiments 1 and 2. In each case this gave an experimental session lasting typically between 45 and 60 min. Pilot studies suggested that longer sessions might produce problems with subjects maintaining their attention. Therefore, for the longer, more complex stimuli of Experiments 3 and 4 (see Fig. 1) it was decided to use two test sessions of five blocks in order to keep the average session time below 60 min. Subjects were free to pause or take a short break in

TABLE III. Results from Experiment 2: pooled confusion matrix for the major-third variant of the test.

Stimulus	Response				
	R_1 (a)	R_2 (b)	R_3 (c)	R_4 (d)	R_5 (e)
S_1 (a)	306	86	10	3	
S_2 (b)	63	226	85	26	5
S_3 (c)	17	83	264	37	4
S_4 (d)	3	21	98	252	31
S_5 (e)		4	7	76	318

between test blocks and most subjects opted to take one break halfway through a session. In Experiments 1 and 2 the order in which subjects completed major-, neutral- and minor-third tests was permutated among the subjects to balance learning effects.

E. Data analyses

Individual-subject data for each identification task were analyzed to give values for information transfer *IT* over the stimulus set and sensitivity index d' between stimuli within the set. These quantities indicate the extent to which the stimulus categories are perceptually distinct. The full-range sensitivity index D' (i.e., cumulative d' across the full stimulus range) is closely related to *IT*, since both $D'$and *IT* indicate the number of discriminable categories within the stimulus range (Braida and Durlach, 1972).

1. Calculation of IT

For an identification task with s stimulus categories, s response categories and n test items, experimental data may be represented by an $s \times s$ confusion matrix with n entries. In this case, $s=5$ and $n=225$ or 250 (for a single subject). Information transfer to the subject was calculated from the confusion matrix according to the formula given by Miller and Nicely (1955), incorporating the correction suggested by Miller (1955) for the case when n is not large.

2. Calculation of d′ and D′

(Implicit in the discussion which follows is the assumption that the identification task is one dimensional, i.e., that the various stimuli within the stimulus set are ranged along a single perceptual dimension. This assumption is necessary in order for d'analysis of confusion matrices to be tractable. The experimental results provide evidence that the tasks in the present study are, to some extent, multidimensional—see below for discussion of this—but do not suggest that the one-dimensional assumption represents a major distortion of subjects' strategies.)

Braida and Durlach (1972) proposed a method by which d' values can be calculated from a confusion matrix. This method works well for well-populated matrices, i.e., when n is very large. However, there are numerical problems for smaller values of n. For this reason a similar but more robust method (Hill, 2000), which is better able to cope with smaller n, has been used in the present study. In summary, this method pools all "high" errors and all "low" errors for a given stimulus category in order to produce more reliable estimates of inter-stimulus d's.

In the present study, d' values are calculated from individual-subject confusion matrices. In a very few instances a negative d' is obtained between neighboring stimulus categories because of a subject's anomalous response pattern, in which cases the negative value has been set to zero. For some of the higher scoring subjects, a given stimulus category may produce a zero error count, in which cases an infinite inter-stimulus d' is suggested. However, this is a consequence of the quantized nature of the subjects' response patterns—a more realistic d' estimate is obtained by attributing a zero error count to a "true" count in the range 0–0.5 and hence calculating d' from an error count of 0.25.

3. Calculation of just-noticeable differences and category widths

A just-noticeable difference *JND* can be defined as the stimulus change which corresponds to $d'=1$. For a particular set of experimental data, a *JND* value (in cents) can thus be calculated as the inverse slope of a cumulative plot of d' vs stimulus separation (in cents) or, equivalently, as the quotient R/D' of the stimulus range R and the full-range sensitivity index D'. ($R=60$ cents for the experiments in this study.)

The *IT* calculated for a particular stimulus set is conventionally interpreted as indicating the number of stimulus categories 2^{IT} within the stimulus range R. The category width r, i.e., the separation (in cents) required for two stimuli to be reliably categorized as different, may thus be calculated using the formula $r=R/(2^{IT}-1)$.

III. RESULTS

Data are available for eight identification tasks: for the various major 3rds in Experiments 1, 2, 3 and 4, for the various neutral thirds in Experiments 1 and 2, and for the various minor thirds in Experiments 1 and 2. All eight tasks produce broadly similar confusion matrices. An example is given in Table III, which shows data pooled over subjects for identification of major thirds in Experiment 2. It can be seen that the degree of difficulty of the task produces an acceptable distribution of errors. Mean values of d' for discrimination within the various stimulus sets (averaged over single-subject values calculated from individual confusion matrices) are given in Figs. 2 and 3. These figures show mean d' for nearest neighbor stimuli ($a:b$, $b:c$, $c:d$, etc.), i.e., the perceptual distance corresponding to a 15 cent incre-

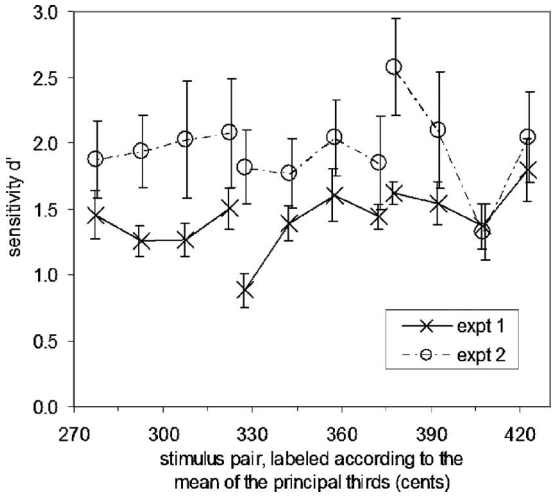


FIG. 2. Sensitivity indices d' for nearest-neighbor stimulus pairs: data from Experiments 1 and 2, averaged over nine subjects; error bars show the standard error. The stimulus pairs are labeled according to the mean of the principal thirds, e.g., a:b as 422.5 cents, y:z as 277.5 cents.

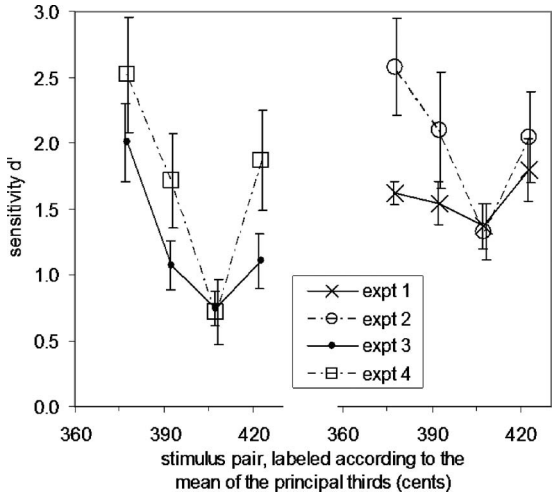


FIG. 3. Sensitivity indices d' for nearest-neighbor stimulus pairs: data from Experiments 3 and 4, averaged over nine subjects; error bars show the standard error. Corresponding data from Experiments 1 and 2 are shown for comparison. The stimulus pairs are labeled according to the mean of the principal thirds, e.g., a:b as 422.5 cents, d:e as 377.5 cents.

ment, as it varies with the position of the increment within the overall pitch range of the stimulus set. Cumulative d' values (corresponding to larger pitch increments) may be calculated from these data, if required. Note that in Experiments 3 and 4, in addition to cues available from pitch increments in the principal third, cues are also available from pitch increments in the tones at a major second above the start note. Hence, in these experiments the experimental results relate to discrimination of changes in the overall tuning system, although for convenience these changes are labeled according to pitch changes in the principal third.

Table IV shows mean values of *IT* (averaged over single-subject values calculated from individual confusion matrices) and mean values for full-range sensitivity index D' (averaged over single-subject values calculated from slopes of individual cumulative d' plots). Table IV also gives values for category width r (calculated from *IT*), for *JND* (calculated from D') and for the quotient r/JND. The quoted uncertainties are standard errors.

A. Results from Experiment 1: Three-tone melodic sequence

Analysis of variance (ANOVA) on the overall scores for each test block, between and within each test session, showed no significant learning effects. The variation of inter-stimulus d' across the stimulus set is shown in Fig. 2. Differences between the three variants of the experiment—major, neutral and minor thirds—were found not to be significant at the $p=0.05$ level [one-way ANOVA on data pooled over each class of third; three levels of "interval" variable; $F(2,16)=3.26$, $p=0.065$]. Differences within the three variants of the experiment were investigated using three separate one-way ANOVAs (four levels of "subinterval" variable)—a significant effect was found at the $p=0.05$ level in the neutral-third case [$F(3,24)=6.17$, $p=0.003$] but not in the major-third case [$F(3,24)=1.65$, $p=0.20$] or the minor-third case [$F(3,24)=0.99$, $p=0.41$]. (Note: No attempt was made to adjust the threshold on the test statistic to take account of multiple implementations of the ANOVA.)

Inter-stimulus d' values have a mean of around 1.5 for the pitch increments of 15 cents. Hence the values of full-range sensitivity index D' (for 60 cent range) are around 6.0, and the *JND* values (for $d'=1$) are around 10 cents (Table IV). The mean values for *IT* are around 1 bit, equivalent to perfect transmission of around two categories. Hence the cat-

TABLE IV. Summary results from Experiments 1, 2, 3 and 4: mean values for information transfer *IT*, category width r, full-range sensitivity index D'; just-noticeable difference JND and the quotient r/JND; the quoted uncertainties are standard errors.

Experiment	*IT* (bits)	r (cents)	D'	JND (Cents)	r/JND
1 major	1.020±0.065	58±5	6.3±0.5	9.5±0.8	6.1
1 neutral	0.897±0.055	70±6	5.3±0.4	11.3±0.8	6.2
1 minor	0.900±0.045	69±5	5.5±0.4	10.9±0.8	6.3
2 major	1.153±0.132	49±8	8.0±1.1	7.5±1.1	6.5
2 neutral	1.132±0.112	50±7	7.5±0.9	8.0±1.0	6.3
2 minor	1.153±0.146	49±9	7.9±1.2	7.6±1.2	6.4
3 major	0.698±0.116	96±21	4.9±0.7	12.2±1.8	7.9
4 major	0.896±0.180	70±20	6.8±1.3	8.8±1.7	8.0

egory width r is approximately equal to the stimulus range R, i.e., 60 cents.

B. Results from Experiment 2: Three-tone harmonized sequence

As with Experiment 1, an ANOVA found no significant learning effects over the course of the experiment. The variation of inter-stimulus d' across the stimulus set is shown in Fig. 2. Differences between the three variants of the experiment—major, neutral and minor thirds—were found not to be significant at the $p=0.05$ level [one-way ANOVA on data pooled over each class of third; three levels of interval variable; $F(2,16)=0.35$, $p=0.71$]. Differences within the three variants of the experiment were investigated using three separate one-way ANOVAs (four levels of "subinterval" variable)—a significant effect was found at the $p=0.05$ level in the major-third case [$F(3,24)=4.34$, $p=0.014$] but not in the neutral-third case [$F(3,24)=0.31$, $p=0.82$] or the minor-third case [$F(3,24)=0.15$, $p=0.93$]. (Note: No attempt was made to adjust the threshold on the test statistic to take account of multiple implementations of the ANOVA.)

Inter-stimulus d' values have a mean of around 2.0 for the pitch increments of 15 cents—higher than in Experiment 1, indicating the effect of the extra cues provided by the additional sustained tone. Hence the values of full-range sensitivity index D' (for 60 cent range) are around 8.0, and the *JND* values (for $d'=1$) are around 7.5 cents—the latter correspondingly lower than in Experiment 1 (Table IV). A two-way ANOVA on the complete data sets for d' from Experiments 1 and 2 (12 levels of "interval/subinterval" variable; two levels of "experiment" variable; data summarized in Fig. 2) indicates that the additional tone in Experiment 2 provides a significant overall benefit, compared to Experiment 1 [$F(1,8)=6.55$, $p=0.034$]. The error bars in Fig. 2 indicate that, compared to Experiment 1, there is greater variation in performance over the subject group. Some subjects find the additional tone of little benefit, whereas other subjects make good use of the additional information. The mean values for *IT* are just over 1 bit, equivalent to perfect transmission of just over two categories. The category width r is consequently slightly less than the stimulus range R and is calculated to be around 50 cents.

C. Results from Experiment 3: Melodic scale-like sequence

A two-tailed t test and nonparametric runs test indicate no significant learning effects in Experiment 3, either in each experimental session or over the course of the whole experiment. The variation of inter-stimulus d' across the stimulus set is shown in Fig. 3. Differences across the stimulus range were investigated using a one-way ANOVA (four levels of subinterval variable) and a significant effect was found [$F(3,24)=15.44$, $p<0.001$]. Inter-stimulus d' values have a mean of around 1.2 for the pitch increments of 15 cents—lower than in Experiment 1, reflecting the additional complexity of the task. The value of full-range sensitivity index D' (for 60 cent range) is 4.9 ± 0.7, and the *JND* is 12.2 ± 1.8 cents—the latter correspondingly higher than in Experiment 1 (Table IV). The mean value for *IT* is approximately 0.7 bits. The category width r is consequently greater than the stimulus range R and is calculated to be around 100 cents.

D. Results from Experiment 4: Harmonized scale-like sequence

The results from Experiment 4 similarly exhibit no learning effects when examined by a two-tailed t test and a nonparametric runs test. The variation of inter-stimulus d' across the stimulus set is shown in Fig. 3. Differences across the stimulus range were investigated using a one-way ANOVA (four levels of subinterval variable) and, as for Experiment 3, a significant effect was found [$F(3,24)=12.11$, $p<0.001$]. Inter-stimulus d' values have a mean of around 1.6 for the pitch increments of 15 cents—lower than in Experiment 2 (reflecting the additional complexity of the task), but higher than in Experiment 3 (indicating the effect of the extra cues provided by the additional sustained tone). The value of full-range sensitivity index D' (for 60 cent range) is 6.8 ± 1.3, and the *JND* is 8.8 ± 1.7 cents—the latter correspondingly higher than in Experiment 2 and lower than in Experiment 3 (Table IV). A two-way ANOVA on the data sets for d' from Experiments 3 and 4 (four levels of subinterval variable; two levels of experiment variable; data summarized in Fig. 3) indicates that the benefit provided by the additional tone in Experiment 4, compared to Experiment 3, is (just) not significant at the $p=0.05$ level [$F(1,8)=4.53$, $p=0.066$]. The error bars in Fig. 3 again indicate a significant variation in performance over the subject group, as in Experiment 2. The mean value for *IT* is approximately 0.9 bits. The category width r is consequently slightly greater than the stimulus range R and is calculated to be around 70 cents.

E. Dimensionality of the task

As mentioned above, the method of d' analysis used here, similar to that reported by Braida and Durlach (1972), is technically only valid for stimuli that vary along a single dimension. In practice, the addition of the upper tone in Experiments 2 and 4 (see Fig. 1) may introduce a second dimension relating to beat frequencies.

For the small pitch changes used in these experiments, the beat frequency (in Hz) varies linearly with the pitch increment (in cents), to a good approximation. However, because the listener cannot distinguish between positive and negative beat frequencies, the perceived beat frequency has a less simple relation to the pitch increment. For example, for the major thirds in Experiments 2 and 4, the dominant beat frequencies are as follows: *a*, 44.0 Hz; *b*, 29.4 Hz; *c*, 14.9 Hz; *d*, 0.6 Hz; *e*, (–)13.6 Hz. Stimulus *d* is effectively beat free and stimuli *c* and *e* are effectively identical with respect to beat frequency. Similarly, for the minor thirds, the dominant beat frequencies are as follows: *v*, 12.5 Hz; *w*, 1.0 Hz; *x*, (–)10.4 Hz; *y*, (–)21.7 Hz; *z*, (–)32.9 Hz. Stimlus *w* is effectively beat free and stimuli *v* and *x* are effectively identical with respect to beat frequency. [For the neutral thirds the beat frequencies relating to mistuning from the

major third are ℓ, (–)4.2 Hz; *m*, (–)18.3 Hz; *n*, (–)32.4 Hz; *o*, (–)46.3 Hz; *p*, (–)60.1 Hz. The beat frequencies relating to mistuning from the minor third are ℓ, 51.5 Hz; *m*, 39.7 Hz; *n*, 27.9 Hz; *o*, 16.3 Hz; *p*, 4.8 Hz.]

According to Braida and Durlach's analysis, in the one-dimensional case the quotient r/JND of the category width r and the *JND* ($d'=1$) should be approximately 5. The data in the final column of Table IV show this quotient to be typically 6 or more in the present study, perhaps suggesting that the one-dimensional assumption represents a distortion of the true situation but not a major distortion. Further evidence is provided by error patterns within the pooled subject data (see, for example, Table III, which relates to Experiment 2), which are generally as expected for the one-dimensional case.

In Experiments 2 and 4 there is some direct evidence that the higher scoring subjects are making use of beat cues, i.e., for those subjects the stimulus set is not one dimensional. For example, for stimulus *e* in Experiment 2, some subjects produce more erroneous identifications as stimulus *c* than as the (nearest neighbor) stimulus *d*. This may be explained on the basis (see above) that beats give little or no information to distinguish stimulus *e* (370 cents) and stimulus *c* (400 cents) from each other, but provide a significant cue to distinguish these two stimuli from stimulus *d* (385 cents).

The effect of beat cues in Experiment 4 is complex. One of the higher scoring subjects commented that stimulus *d* (see Table II) could be reliably identified by the lack of beats on the 3rd and 5th tones in the sequence (whose relation to the sustained tone at 700 cents is very close to a beat-free 316 cents), whereas stimulus *c* could be reliably identified by the lack of beats on the 2nd and 6th tones in the sequence (whose relation to the sustained tone at 700 cents is very close to a beat-free 498 cents). The task for that subject was then to distinguish the other three stimuli which provided no obvious beat cues.

F. General discussion of Experiments 1, 2, 3 and 4

The addition of the continuous upper tone in Experiments 2 and 4 enhances subjects' mean performance compared to Experiments 1 and 3, respectively, giving higher values for the full-range sensitivity index D', and lowering *JND*s by 2 or 3 cents (Table IV). Comparison of mean data for major thirds in Experiment 1 and Experiment 2 with those of Experiment 3 and Experiment 4 shows that the addition of extra melodic content leads to lower values for the full-range sensitivity index D' and an increase of 2 or 3 cents in *JND*, reflecting the greater complexity of the task. However, this reduction of performance is not seen for the better subjects—they maintain a similar performance, presumably because the increased complexity of the task is balanced by the additional cues available within the more complex stimuli.

As mentioned above, a significant variation of inter-stimulus d' might be expected across the stimulus range—for example, sensitivity might be less around unfamiliar intervals, such as the neutral third (approximately 350 cents) and greater around familiar intervals, such as the major and minor thirds (equal tempered at 400 and 300 cents and just intonation at 386 and 316 cents). However, in the results from Experiments 1 and 2, the perceptual landscape appears generally featureless—Fig. 2 shows that within each experiment there is little variation of inter-stimulus (nearest neighbor) d' across the stimulus range. This is true for small movements within the stimulus range (as shown by the individual lines in Fig. 2) and also for larger movements within the stimulus range (i.e., moving between major, neutral and minor thirds in the data shown in Fig. 2). The grossly mistuned intervals (365 and 335 cents) and the "halfway" neutral third (350 cents) are not distinguished much better or worse than the stimuli representing "in tune" musical intervals (equal-tempered thirds at 400 and 300 cents and just-intonation thirds at 386 and 316 cents). The statistically significant variation for neutral thirds in Experiment 1 is not observed in Experiment 2, and may thus be viewed with some uncertainty. The statistically significant variation for major thirds in Experiment 2 is not observed in Experiment 1—however, a similar significant variation is also observed in Experiments 3 and 4, and hence, on the basis of these results for major thirds, it may be concluded that the perceptual landscape is not entirely featureless. In fact, results from all four experiments suggest that the $b:c$ discrimination between major thirds of 415 and 400 cents is reduced compared to the other pairs of nearest-neighbor major thirds—indicated by the dip in all four lines in Fig. 3. (As an alternative statistical treatment, a simple binomial analysis suggests a probability of less than 3% for chance observation that one discrimination from four is consistently easier or harder than the rest over the four experiments—hence we may conclude that, for this particular subject group, the $b:c$ discrimination is significantly different from the rest.) It is not obvious why perceptual cues should be weaker for this particular discrimination. The variation across the range of major-third stimuli (Fig. 3) is more marked in Experiments 3 and 4 (where, in principle, there are cues from the tuning of the major seconds as well as from the major thirds) than in Experiments 1 and 2 (where cues are only available from the major thirds).

IV. CONCLUSIONS

In the quasi-musical context of these experiments, category widths for identification were measured at around 70 cents and *JND*s in frequency for the complex tones were measured at around 10 cents. (As expected, the addition of the continuous upper tone in Experiments 2 and 4 produced an enhancement of subjects' mean performance compared to Experiments 1 and 3, apparently due to the availability of beat cues.) The measured values for category width r and *JND* compare well with the semitone (around 100 cents), which forms the basis of conventional musical practice, and the comma (around 20 cents) which distinguishes intervals such as just intonation and Pythagorean major thirds. The measured category width in less than a semitone, indicating that intervals of a semitone or more will be reliably identified; the measured *JND* is less than a comma, suggesting that

differences between intervals such as just intonation and Pythagorean major thirds are likely to be apparent to the listener. The measured values for category width and *JND* also compare quite well with the findings of Parncutt and Cohen (1995)—falloff in performance for a melodic identification task was observed when the melody was constructed with step sizes less than 40 cents, and close-to chance performance was observed with a step size of 10 cents.

The sensitivity index d' for a 15 cent increment was typically calculated in the range 1.0–2.0. No significant difference in response was found between discrimination of major, minor or neutral thirds, although sensitivity was slightly higher for major thirds. Similarly, little significant difference in response was found between intervals within each of the major, minor or neutral stimulus sets (although, for major thirds, discrimination of the 15 cent increment between 400 and 415 cents was reduced compared to discrimination of other 15 cent increments within the stimulus sets). This is in line with the findings of Parncutt and Cohen (1995), whose results for a melodic identification task with various step sizes in the range 25–133 cents showed no peak in performance at the "in tune" step size of 100 cents. However, results for the present study do not correspond with the findings of some previous studies (Burns and Ward, 1978; Burns and Campbell, 1994) which indicated nonuniformity across the stimulus range. In the latter study, the average sensitivity index d' for a 25 cent increment was found to be around 1.0, but with periodic minima in d' across the range of stimulus tunings (for intervals at multiples of 100 cents). It should be noted that, in that study, the response categories (with 50 cent increments) did not match the stimulus categories. This hinders comparison with the present study, but it may be conjectured that the relatively constant d' values observed in the present study may be a consequence of the experimental paradigm, which may have encouraged subjects to disregard their previous musical experience. However, the lack of evidence for anchors or landmarks does not necessarily imply that these have no effect on perception in the context of these experiments—it may simply be that they stretch and compress perceptual space so as to produce little effect overall.

ACKNOWLEDGMENTS

The authors thank the participants for their generosity. This work was supported by the UK Engineering and Physical Sciences Research Council. Comments by an anonymous reviewer were particularly helpful.

Braida, L. D., and Durlach, N. I. (**1972**). "Intensity perception: II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483–502.

Burns, E. M., and Campbell, S. L. (**1994**). "Frequency and frequency-ratio resolution by possessors of absolute and relative pitch—examples of categorical perception," J. Acoust. Soc. Am. **96**, 2704–2719.

Burns, E. M., and Ward, W. D. (**1978**). "Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals," J. Acoust. Soc. Am. **63**, 456–468.

Creel, S. C., Newport, E. L., and Aslin, R. N. (**2004**). "Distant. melodies: Statistical learning of nonadjacent dependencies in tone sequences," J. Exp. Psychol. Learn. Mem. Cogn., **30**, 1119–1130.

Deutsch, D. (**1980**). "The processing of structured and unstructured tonal sequences," Percept. Psychophys. **28**, 381–389.

Hall, D. E., and Hess, J. T. (**1984**). "Perception of musical interval tuning," Music Percept. **2**, 166–195.

Harris, J. D. (**1952**). "Pitch discrimination," J. Acoust. Soc. Am. **24**, 750–755.

Hill, T. J. W. (**2000**). "Experiments on the perception of pitch increments in simple tone sequences," Unpublished Ph.D. thesis, University of Exeter, Exeter, UK.

Kameoka, A., and Kuriyagawa, M. (**1969a**). "Consonance theory. I. Consonance of dyads," J. Acoust. Soc. Am. **45**, 1451–1459.

Kameoka, A., and Kuriyagawa, M. (**1969b**). "Consonance theory. II. Consonance of complex tones and its calculation method," J. Acoust. Soc. Am. **45**, 1460–1469.

Kidd, G. R., and Watson, C. S. (**1992**). "The 'proportion-of-the-total duration (PTD) rule' for the discrimination of auditory patterns," J. Acoust. Soc. Am. **92**, 3109–3118.

Miller, G. A. (**1955**). "Note on the bias of information estimates," in *Information Theory in Psychology: Problems and Methods*, edited by H. Quastler (Free Press, Glencoe, IL), pp. 95–100.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Moore, B. C. J., and Glasberg, B. R. (**1990**). "Frequency discrimination of complex tones with overlapping and non-overlapping harmonics," J. Acoust. Soc. Am. **87**, 2163–2177.

Mori, S., and Ward, L. M. (**1995**). "Pure feedback effects in absolute identification," Percept. Psychophys. **57**, 1065–1079.

Nelson, D. A., Stanton, M. E., and Freyman, R. L. (**1983**). "A general equation describing frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **73**, 2117–2123.

Parncutt, R., and Cohen, A. J. (**1995**). "Identification of microtonal melodies—effects of scale-step size, serial order, and training," Percept. Psychophys. **57**, 835–846.

Perlman, M., and Krumhansl, C. L. (**1996**). "An experimental study of internal interval standards in Japanese and Western musicians," Music Percept. **14**, 95–116.

Rasch, R. A. (**1983**). "Description of regular twelve-tone musical tunings," J. Acoust. Soc. Am. **73**, 1023–1035.

Rasch, R. A. (**1985**). "Perception of melodic and harmonic intonation of two-part musical fragments," Music Percept. **2**, 441–458.

Schellenberg, E. G., and Trehub, S. E. (**1994**). "Frequency ratios and the discrimination of pure-tone sequences," Percept. Psychophys. **56**, 472–478.

Sek, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479–2486.

Smith, N. A., and Schmuckler, M. A. (**2004**). "The perception of tonal structure through the differentiation and organization of pitches," J. Exp. Psychol. Hum. Percept. Perform. **30**, 268–286.

Spiegel, M. F., and Watson, C. S. (**1984**). "Performance on frequency-discrimination tasks by musicians and nonmusicians," J. Acoust. Soc. Am. **76**, 1690–1695.

Thompson, W. F., Hall, M. D., and Pressing, J. (**2001**). "Illusory conjunctions of pitch and duration in unfamiliar tone sequences," J. Exp. Psychol. Hum. Percept. Perform. **27**, 128–140.

Tufts, J. B., Molis, M. R., and Leek, M. R. (**2005**). "Perception of dissonance by people with normal hearing and sensorineural hearing loss," J. Acoust. Soc. Am. **118**, 955–967.

Vos, J. (**1986**). "Purity ratings of tempered fifths and major thirds," Music Percept. **3**, 221–258.

Vos, J. (**1988**). "Subjective acceptability of various regular twelve-tone tuning systems in two-part musical fragments," J. Acoust. Soc. Am. **83**, 2383–2392.

Vos, J., and van Vianen, B. G. (**1985**). "Thresholds for discrimination between pure and tempered intervals: The relevance of nearly coinciding harmonics," J. Acoust. Soc. Am. **77**, 176–187.

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.

Resonance wood [*Picea abies* (L.) Karst.] – evaluation and prediction of violin makers' quality-grading

Christoph Buksnowitz[a] and Alfred Teischinger
University of Natural Resources and Applied Life Sciences, Vienna, Department of Material Sciences and Process Engineering, Institute of Wood Science and Technology, Peter Jordanstr. 82, A-1190 Vienna, Austria

Ulrich Müller
Compentence Center for Wood Composites and Wood Chemistry, St.-Peter-Str. 25, A-4021 Linz, Austria

Andreas Pahler
Holzforschung der TU Munich, Winzererstrasse 45, D-80797 Munich, Germany

Robert Evans
Ensis/CSIRO Forestry and Forest Products, Private bag 10, Clayton South, Victoria 3169, Australia

(Received 28 September 2006; revised 13 December 2006; accepted 20 December 2006)

The definition of quality in the field of resonance wood for musical instrument making has attracted considerable interest over decades but has remained incomplete. The current work compares the traditional knowledge and practical experience of violin makers with a material-science approach to objectively characterize the properties of resonance wood. Norway spruce [*Picea abies* (L.) Karst.] has earned a very high reputation for the construction of resonance tops of stringed instruments and resonance boards of keyboard instruments, and was therefore chosen as the focus of the investigation. The samples were obtained from numerous renowned resonance wood regions in the European Alps and cover the whole range of available qualities. A set of acoustical, anatomical, mechanical and optical material properties was measured on each sample. These measurements were compared with subjective quality grading by violin makers, who estimated the acoustical, optical and overall suitability for violin making. Multiple linear regression models were applied to evaluate the predictability of the subjective grading using the measured material characteristics as predictors. The results show that luthiers are able to estimate wood quality related to visible features, but predictions of mechanical and acoustical properties proved to be very poor. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2434756]

PACS number(s): 43.75.De [NF] Pages: 2384–2395

I. INTRODUCTION

Music has always played an important role in mankind's evolution. Owing to its high emotional and cultural importance, generations of musicologists, acousticians, chemists and material scientists have been inspired to investigate the interrelations of materials, sound and music (e.g., Bucur, 1983/2006; Holz, 1966/1984; Hori *et al.*, 2002; Hutchins 1992; Schwalbe and Becker, 1920).

The development of music has been influenced by the listeners' tastes, by outstanding musicians and their style of playing, by new trends in composition, but also by changes in the construction and the performance of musical instruments.

The superior sound of some masterpieces of instruments has remained a mystery up to the present. Numerous studies with different approaches have been performed to reveal those secrets, e.g., Yano *et al.* (1994) and Holz (1996), with their work on chemical treatment of resonance wood, Schumacher (1988) with a paper on the compliances of wood for violin plates, and Anderson and Strong (2005) with the idea to investigate the effect of an inharmonic partial on the pitch of pianos.

The delightful sound of, for example, the famous violins of Stradivari can be explained in part by the high quality of the construction and workmanship of the instruments. On the other hand, there is no doubt that the sound of musical instruments also depends on the quality of the raw material as well as the experience and intuition of the instrument maker to "fine-tune" their construction to cope with a wide range of resonance wood quality. Therefore it is worth taking another close look at the resonance wood quality of Norway spruce [*Picea abies* (L.) Karst.], which has earned a very high reputation for the construction of resonance tops of stringed instruments and resonance boards of keyboard instruments.

Most violin makers have only their senses to grade their raw material. The present study aims at linking the experience and knowledge of craftsmen and instrument makers with a basic material assessment to objectively define the material properties of resonance wood. A collective of resonance wood samples from numerous tone wood regions in Europe was used for the material assessment, which covered acoustical, anatomical, mechanical and optical properties on

[a]Author to whom correspondence should be addressed. Electronic mail: christoph.buksnowitz@boku.ac.at

 0001-4966/2007/121(4)/2384/12/$23.00

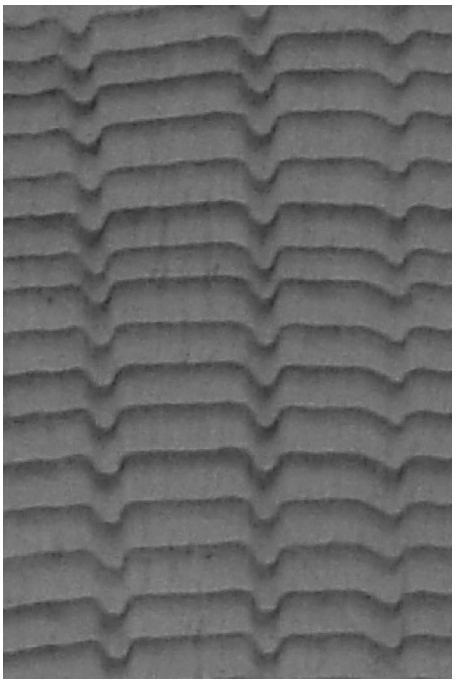


FIG. 1. Cross section of a "hazel-growth" Norway spruce showing three radial lines of indents.

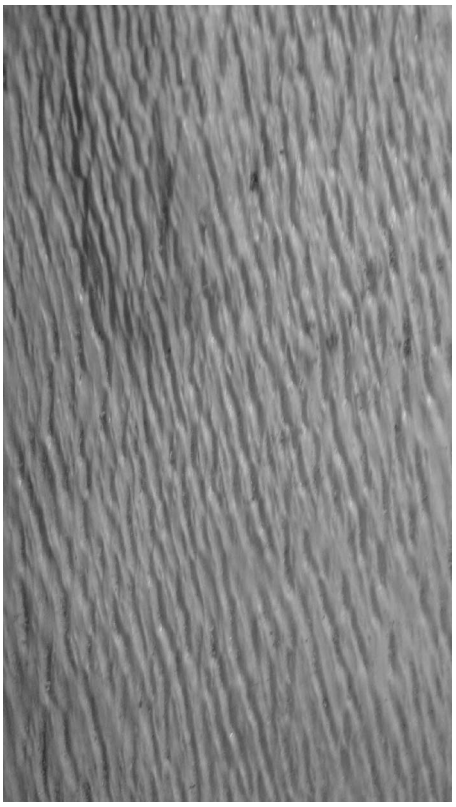


FIG. 2. Surface of a hazel-growth spruce log with removed bark.

several hierarchical levels. In addition the samples were subjectively graded in their acoustical, optical and overall suitability for violin making by instrument makers before they have been dissected for scientific analysis. The following questions should be answered by the comparison between measurable material properties and subjective quality grading of instrument makers.

- Can the instrument makers' choice for resonance wood be related to material characteristics objectively measured?
- Which material properties are taken into account by instrument makers?
- Which are the decisive criteria and governing factors for the grading of the resonance wood?

II. MATERIAL

Eighty four samples of Norway spruce (*Picea abies*. (L.) Karst.) from numerous resonance wood regions in Europe were collected. Seventy eight of them were raw violin tops, with approximate dimensions of 40 cm in longitudinal direction, 15 cm radially and 3–5 cm tangentially on the bark side. The samples were cut radially or split out of large diameter Norway spruce logs. Approximately 44% of the samples showed a growth anomaly called "hazel growth," "bear-claw" or "indented rings" (Figs. 1 and 2). Hazel growth is an abnormality in radial growth. The indentations of the rings towards the pith occur irregularly and mostly in adult wood (Ziegler and Merz, 1961). The samples were chosen to cover a representative variation of all available qualities of spruce resonance wood. Six nonresonance wood samples were added, which were graded as "joinery quality" by the retailer. These samples extended the set towards the lower qualities. The total sample set is shown in Table I giving information on the growth patterns and the origin. All samples were air dried and seasoned for at least one year. This period is considered the minimum to reduce internal growth stresses and drying stresses (Holz, 1972). The raw violin resonance tops (Fig. 3) were converted into small boards of identical thickness (16 mm) and surface quality (Fig. 4) prior to being graded by violin makers. Before dissecting the samples into the different specimens for subsequent tests (tension, compression, bending, etc.) they were conditioned at 65% relative humidity and 20 °C until equilibrium moisture content was achieved to guarantee precise, stable dimensions. The cutting plan for the specimens is shown in Fig. 5.

III. SUBJECTIVE GRADING

As a reference to the practical experience and the traditional knowledge of musical instrument makers, all resonance wood samples were subjectively graded by 14 renowned Austrian violin makers. To ensure an unbiased result, grading of the boards was performed in a "blind," meaning that graders (violin makers) had no information on the samples, which were randomly arranged on the tables of a lecture hall for the grading. The three separate grading criteria were the suitability (quality) for violins from an acoustical point of view, the suitability in respect to aesthetical (optical) requirements and finally an overall appraisal of the resonance wood quality. The participants were instructed to evaluate the three criteria separately from each other not using any tools. The violin makers were requested to grade the samples just like they are used to doing in practice. The parameters, which are used by violin makers to evaluate the

TABLE I. Regions of origin and growth patterns of resonance wood samples.

REGIONS of origin of the resonance wood samples	GROWTH PATTERN		TOTAL count
	NORMAL	INDENTS[a]	
Austria Tyrol	0	14	14
Austria Salzkammergut	3	5	8
Austria Styria	1	0	1
Germany Allgaeu	4	0	4
Germany Upper Bavaria	13	0	13
Italy Val di Fiemme	12	15	27
Italy South Tyrol	3	3	6
Switzerland Lucerne	5	0	5
Reference samples - origin unknown	6	0	6
TOTAL	47	37	84

[a]Hazelgrowth spruce.

samples, vary a lot. Some search for a specific feature others integrate over many properties in a complex intuitional decision. The following five grades were used: 1= perfectly suitable for violin making, 2= very suitable for violin making, 3= suitable for violin making, 4= suitable for violin making with restrictions, 5= not suitable for violin making.

The violin makers only used their senses, checked the planed surfaces, the cross sections, the knocking tone pitch, estimated weight and so forth. The participants filled in the results into a uniform questionnaire. The grading results were averaged over the 14 participants before being used in further statistical analysis. The three subjective grading criteria (acoustical quality, optical quality and overall quality) served as the three dependent variables, which were to be predicted by the objectively measured material properties. In the following, the different testing methods are briefly described. A detailed description of the testing methods can be found elsewhere (Buksnowitz, 2006).

IV. MATERIAL PROPERTIES

Most of the material properties were determined by testing several specimens taken from different radial positions in the same resonance wood sample. In the case of radially oriented specimens some measurements were performed quasi-continuously with a certain resolution (step width). For the purpose of statistical analysis the measurements were averaged for the whole sample, to match the subjective grading variables of the resonance wood boards. Table II gives an overview of the number of specimens per sample for each parameter measured and lists their dimensions and the radial positions from which they were taken.

A. Sound velocity

The sound velocity was measured at 54 kHz in the longitudinal direction as well as the radial direction. Before testing, the samples were conditioned until equilibrium moisture content (20 °C, 65% RH) was reached. The transit time was determined by means of an ultrasonic tester (PUNDIT-C.N.S. Electronics LTD. London, England) with a resolution of 0.1 μs. The transmitter was attached with constant pressure to all samples. No contact medium was used to transmit sound waves from the transmitter to the board. The sound velocity was calculated according to Eq. (1).

FIG. 3. Raw material for violin resonance tops.

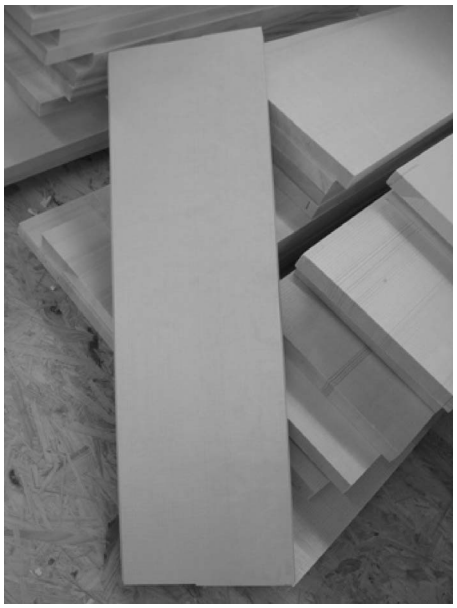


FIG. 4. Machine planed boards for subjective grading by instrument makers.

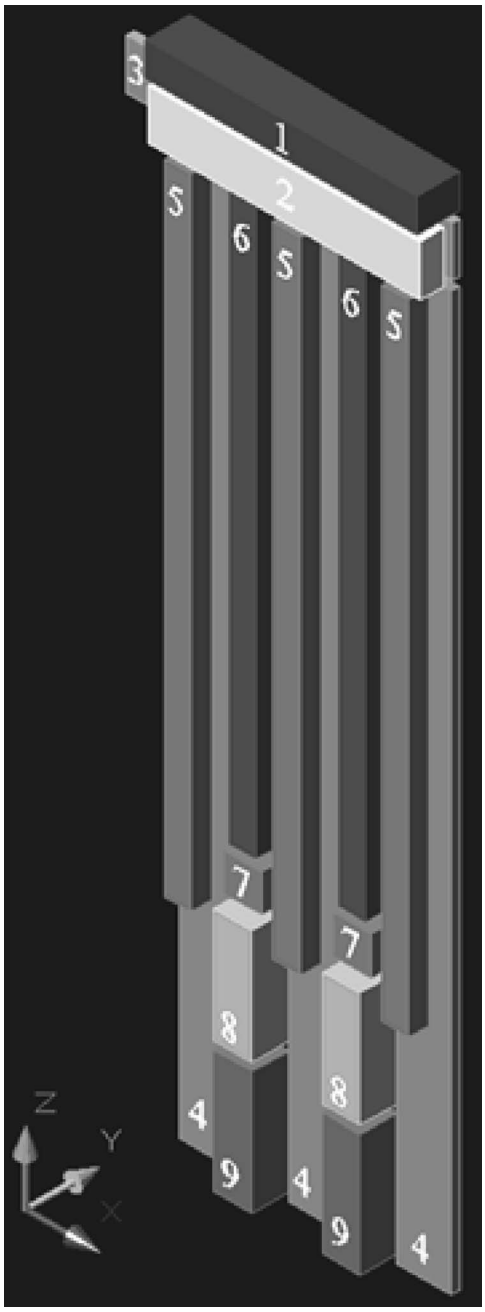


FIG. 5. Specimen cutting plan for experimental determination of material characteristics; Z=longitudinal direction, Y=tangential direction, X=radial direction. The specimen numbers refer to Table II, which gives the link to the measured parameters.

$$v = \frac{l}{t} \tag{1}$$

where v is the sound velocity [m/s], l is the specimen length, path length [m], and t is the transit time of ultrasound [s].

B. Logarithmic decrement (damping)

The logarithmic decrement (related closely to the loss tangent) was determined according to DIN EN ISO 6721-1 (2003) and DIN 6721-3 (1996). The specimens (longitudinal and radial strips of wood) were clamped on one end. The remaining free end was deflected to a defined position. After the release of the flexed end a laser device (M7L/2 sensor, MEL Mikroelektronik GmbH) picked up the amplitude of the damped vibration over time. The logged amplitude-time signal was used to calculate the logarithmic decrement applying Eq. (2):

$$\Lambda = \ln \frac{x_q}{x_{q+1}}, \tag{2}$$

where Λ is the logarithmic decrement [], x_q is the amplitude q, and x_{q+1} is the amplitude $q+1$ (directly following x_q).

The logarithmic decrement was averaged over 15 oscillations starting at a certain amplitude to avoid the noise in the initial phase of vibration. This parameter was investigated for longitudinal specimens (vibration in tangential plane) as well as for radial specimens (vibration in the plane of the cross section). Note that the relationship between the logarithmic decrement and the loss tangent is $\Lambda = 2\pi \tan \delta/2$ (Macdonald, 1966).

C. Resonance frequency

The resonance frequency (first bending mode) was derived from Fourier transform of the amplitude-time signal. The resonance frequency was computed by a standard computer program (Diadem) using the signal from the test setup described for derivation of logarithmic decrement.

D. Modulus of elasticity and modulus of rupture (MOR)

All mechanical experiments were performed on a universal testing machine (Zwick/Roell Z100). Before testing, the samples were conditioned until equilibrium moisture content (20 °C, 65% RH) was reached.

The limited material available per sample made it necessary to scale down the sample dimensions specified in the DIN standards. The following tests were performed to describe the stiffness of the material:

- three-point bending test according to DIN 52 186 (1978)—samples scaled down to half size.
- transversal bending test—three-point bending test—load applied on the radial surface (in the direction of bridge pressure on the resonance top of a violin or guitar).
- tension test according to DIN 52 188 (1979)—samples scaled down to half size.
- compression test according to DIN 52 185—samples scaled down.

Although strength (MOR) is only of secondary importance in the event of a constructional deficiency of the instrument or an impact, MOR values were recorded for each sample.

E. Hardness

The hardness test was performed according to the standard ÖNORM EN 1534 (2000) with the constraint that the sample sizes were scaled down. The assumption that the ratio of the sample size to the indent diameter is still acceptable was based on a publication by Gindl *et al.* (2004).

TABLE II. Number of specimens per sample, their dimensions and the radial positions for each parameter measured. For longitudinally oriented specimens five radial positions were possible. The applicable positions are marked x. Marks at all five positions indicate a radially oriented specimen. Pith=position closer to pith, bark=outermost position. Use No. to link Table II with the cutting plan in Fig. 5.

		RADIAL POSITION						DIMENSION OF SPECIMEN [mm]		
PARAMETER	No.	Pith		Mid		Bark	NUMBER/ RESOLUTION of measurement	LON	RAD	TAN
Sound velocity longitudinal	4	x		x		x	3x	320.0	20.0	3.5
Sound velocity radial	3	x	x	x	x	x	1x	18.5	120.0	3.0
Log. decr. longitudinal	4	x		x		x	3x	320.0	20.0	3.5
Log. decr. radial	3	x	x	x	x	x	1x	18.5	120.0	3.0
Res. freq. longitudinal	4	x		x		x	3x	320.0	20.0	3.5
Res. freq. radial	3	x	x	x	x	x	1x	18.5	120.0	3.0
MOE/MOR three-point bending	6		x		x		2x	210.0	10.0	10.0
MOE/MOR bend. transversal	3	x	x	x	x	x	1x	18.5	120.0	3.0
MOE/MOR tension	5	x		x		x	3x	235.0	10.0	7.5
MOE/MOR compression	8		x		x		2x	42.0	14.0	14.0
Hardness Brinell	7		x		x		2x	14.0	14.0	10.0
Density conditioned	1	x	x	x	x	x	0.05 mm	~5.0	100.0	~1.0
Ring width	1	x	x	x	x	x	at ring boundaries	~5.0	100.0	~1.0
Ring width variation coeff.	2	x	x	x	x	x	···	12.0	100.0	14.0
Late wood percentage	2	x	x	x	x	x	···	12.0	100.0	14.0
Tracheid diameter radial	1	x	x	x	x	x	0.05 mm	~5.0	100.0	~1.0
Tracheid diameter tangential	1	x	x	x	x	x	0.05 mm	~5.0	100.0	~1.0
Tracheid wall thickness	1	x	x	x	x	x	0.05 mm	~5.0	100.0	~1.0
Fiber length	2	x				x	length in μm	~18.5	100.0	~5.0
Microfibril angle	1	x	x	x	x	x	0.1 mm	~5.0	100.0	~1.0
Brightness	1	x	x	x	x	x	6.0 mm/1.0 mm	12.0	100.0	···
Red component	1	x	x	x	x	x	6.0 mm/1.0 mm	12.0	100.0	···
Yellow component	1	x	x	x	x	x	6.0 mm/1.0 mm	12.0	100.0	···
Max. swelling radial	9		x		x		2x	45.0	14.0	14.0

Abbreviations: log. decr. = logarithmic decrement (damping); res. freq. = resonance frequency 1st mode; MOE = modulus of elasticity; MOR = modulus of rupture (strength); coeff. = coefficient.

F. Anatomical features

Density, tracheid diameter, tracheid wall thickness, microfibril angle and ring width were measured by means of SilviScan-3, a system of instruments for nondestructive wood microanalysis. It combines x-ray densitometry (Evans *et al.* 1999; Gureyev and Evans, 1999; Washusen *et al.*, 2001), x-ray diffractometry (Evans *et al.*, 1996, Evans *et al.*, 1999) and image analysis (Evans, 1994) to determine parameters such as density, fiber diameter and microfibril angle (Evans *et al.*, 1996; Evans, 1999; Stuart and Evans, 1995; Evans and llic, 2001) at high resolution. In the current project the parameters mentioned above were measured with a radial resolution shown in Table II. For SilviScan measurements radially oriented samples (axial $\times$ tangential=7 $\times$2 mm) were dissected from the specimen with a double blade circular saw. One cross section was polished with a series of sand papers to guarantee proper surface quality for image analysis.

1. Density

The densities of the SilviScan specimens were determined gravimetrically on conditioned samples at 40% relative humidity and 20 °C.

2. Tracheid diameter and wall thickness

The high resolution video microscope of SilviScan-3 scanned the cross section of the samples. Subsequently the diameter of the tracheids in radial and tangential direction was derived from the scans by a special image analysis (Evans, 1994, Evans *et al.*, 1999). Tracheid wall thickness was calculated from density and tracheid diameter according to Evans (1994).

3. Microfibril angle

The microfibril orientation in the S2 layer of the tracheid cell wall was estimated using the SilviScan-3 x-ray diffractometry system (Evans, 1999).

4. Ring width

The ring widths were determined from radial density profiles generated by SilviScan-3 x-ray densitometry system.

5. Ring width variation coefficient

As an indicator for the growth ring regularity the ring width variation coefficient (εj) proposed by Holz (1972) was used—see Eq. (3).

$$\varepsilon_j = \left[\frac{1}{N-1} \sum_{i=1}^{N-1} \left(\frac{200 \Delta b_{ji}}{b_{ji} + b_{j(i+1)}} \right)^2 \right]^{1/2}, \quad (3)$$

where b_{ji} is the ring width [mm] $i=1,\ldots,N$, N is the total number of rings, and $\Delta b_{ji}=b_{ji}+b_{j(i+1)}, i=1,\ldots,N-1$.

The ring width as well as the latewood proportions were measured using a dendrochronological setup [stereo microscope Zeiss Stemi 2000C, charge coupled device (CCD) camera Sony CCD Iris, a screen, a table to move the specimen by micrometer] (Grabner and Wimmer 2006).

6. Late wood percentage

The measurements of the latewood and earlywood widths were analyzed by using the software program TSAP-win Professional 0.53. The latewood percentage was calculated according to Eq. (4).

$$\ell wp[\%] = \frac{\sum \ell ww}{\sum rw} \cdot 100, \quad (4)$$

where ℓwp is the latewood percentage [%], ℓww is the latewood width [mm], and rw is the ring width [mm].

7. Fiber length

The fiber length was separately determined for latewood and earlywood. At two radial positions (radially 10 mm away from each end of the sample) small chips of earlywood and latewood were isolated. If the ring at this position was too narrow to gain enough material, the consecutive rings were included. The wood was macerated in a solution of 2.65 g $K_2Cr_2O_7$+5 ml 65% HNO_3+25 ml H_2O. A detailed description of the maceration procedure is given by Jeffrey (1917). The tracheid lengths were measured on images, which were captured with a digital camera (Olympus DP 10) mounted on an incident light microscope (Olympus SZH 10 - research stereo) at a 20-fold magnification. The image analysis program was Olympus DP Soft 3.0. For further statistical analysis the fiber lengths of the earlywood and the latewood were averaged for the two radial positions and weighted with the sample's average earlywood and latewood portion.

8. Dimensional stability - swelling

In order to describe the dimensional stability, the swelling coefficients were determined according to the DIN 52 184 (1979) standard, with the constraint that the sample size was reduced to 45 mm in the longitudinal direction and 14 mm in radial and the tangential directions.

The maximal swelling coefficient was calculated according to Eq. (5)

$$\alpha_{\max} = \left[\frac{\ell_w - \ell_0}{\ell_0} \right] \cdot 100, \quad (5)$$

where $\alpha_{\max}$ is the maximal swelling coefficient [%], ℓ_w is the specimen dimension in completely soaked state [mm], and ℓ_0 is the specimen dimension in completely oven-dry state [mm].

The swelling coefficient in the longitudinal direction is an order of magnitude smaller than in the radial direction, and the resonance tops and boards of musical instruments are only a few millimeters thick in the tangential direction; only the swelling coefficient in the radial direction was considered in the statistical analysis.

9. Color

Color measurements of the samples were performed according to DIN 5033 using a CODEC 400 device (Phyma GmbH A-2531 Gaaden) with a 6 mm iris. One hundred overlapping and aligned measurement points across the radial surface of the specimen were averaged for further analysis and modeling. The color measurements were performed using the color model $L^\star a^\star b^\star$ of the Commission Internationale d'Eclairage.

V. ANALYSIS METHODS

For the analysis of the generated data matrix the software package SPSS 11 was used. A multiple linear regression model was applied to predict the value of a dependent scale variable (subjective grading by the violin makers) based on its assumed linear relationship to one or several predictors (measured material properties). In this study either the subjective grading of the acoustical quality (Model A), the subjective grading of the optical quality (Model B) or the subjective grading of the overall quality (Model C) by violin makers served as the dependent variable (marked with "x" in Table III). The according group of predictors (independent variables) used in the three different models can be derived from Table III (indicated with figures 0, 1, 2, 3 in column A, B and C). Although they are often applied for the assessment of resonance wood quality, no "composite variables" such as the radiation ratio: $R=c/\rho=(E/\rho^3)^{0.5}$ were used as predictors to prevent collinearity. A factor analysis routine was applied to gain better insight into variable dependencies in order to avoid collinearity.

The linear regression model assumes that there is a linear relationship between the dependent variable (y) and each predictor (x_n), which is described in Eq. (6).

$$y = b_1 \cdot x_1 + b_2 \cdot x_2 + \cdots + b_n \cdot x_n + a, \quad (6)$$

where x_i indicator independent variables, n is the number of independent variables, b_i is the coefficients of the independent variables estimated by the model, and a is the intercept.

Numerous statements in literature (Table III) point out that the assumption of a linear relationship between the subjective quality grading and the material characteristics is admissible. Although there might be a deviance from the perfect linear relation in some cases, at least the trend in the correlation can be revealed.

The multiple linear regression model estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. The coefficients are estimated using the least-squares method. Before applying the routine, the suitability of data has to be considered. The dependent and independent variables should be quantitative. Categorical variables need

TABLE III. Illustration of the experimental results, grading results and modeling results supplemented with corresponding values and suggestions from the literature.

		LITERATURE VALUE							MODELING DEPENDENT VARIABLES AND PREDICTORS		
		QUALITY-CORRELATION ▼									
VARIABLES	UNIT	No. in CITATION-LIST ▼			MEAN	SDEV	MIN	MAX	A	B	C
GRADING RESULTS											
Grading: acoustical quality	[ ]	…	…		2.84	0.59	1.93	4.15	x		
Grading: optical quality	[ ]	…	…		3.08	0.88	1.64	4.86		x	
Grading: total quality	[ ]	…	…		3.07	0.76	1.93	4.71			x
GROWTH PATTERN											
Hazel growth or normal	0/1[a]	…	[1]	±	…	…	…	…		0	0
ACOUSTICS											
Sound velocity longitudinal	[m/s]	5500, max. 5600–6000	[2]	+	6183	373	4906	6897	0		0
Sound velocity radial	[m/s]	449, >1000, 1681	[3]	+	1889	210	1042	2194	0		0
Log. dec. longitudinal	[ ]	as low as possible	[4]	–	0.0404	0.0018	0.0366	0.0461	0		0
Log. dec. radial	[ ]	as low as possible	[5]	–	0.1030	0.0109	0.0839	0.1441	0		0
Res. freq. longitudinal	[Hz]	high No. of res. freq. +	[6]	*	29.38	1.79	23.51	32.68	0		0
Res. freq. radial	[Hz]	high No. of res. freq. +	[7]	*	63.25	6.80	39.60	79.74	0		0
MECHANICS											
MOE 3-point bending	[N/mm^2]	*12 992, 10 000–12 000*	[8]	+	10718	1685	6422	14725			0
Strength 3-point bending	[N/mm^2]	*101.2, 49…78…136*	[9]	+	78.8	9.3	57.8	100.3			0
MOE transverse 3-p. b.	[N/mm^2]	*500–1000, 490*	[10]	±	852	181	304	1283			0
Strength transverse 3-p. b.	[N/mm^2]	–	[11]	+	10.2	1.9	6.0	15.1			0
MOE tension	[N/mm^2]	*13 500*	[12]	(+)	12340	2050	7224	16989			0
Strength tension	[N/mm^2]	*90, 88*	[13]	+	101.3	14.9	70.6	136.2			0
MOE compression	[N/mm^2]	*13 760*	[14]	(+)	12834	2608	6401	19786			0
Strength compression	[N/mm^2]	*50, 49*	[15]	+	36.9	4.1	28.2	45.4			0
Hardness Brinell	[N/mm^2]	*12*	[16]	(+)	9.88	1.70	6.18	13.28			0
DENSITY											
Density conditioned	[g/cm^3]	0.47, 0.4, >0.43, 0.427	[17]	–	0.432	0.033	0.354	0.501	1	0	0
ANATOMY											
Ring width	[mm]	1–2, <2	[18]	><	1.25	0.32	0.66	2.06		3	2
Ring width variation	[ ]	26–28, <30	[19]	–	18.4	5.5	11.4	35.7		1	1
Latewood percentage	[%]	20–36, 20–25	[20]	–	20.2	3.9	13.9	28.6		0	0
Tracheid diameter radial	[μm]	*35.5* (ew), *28.1* (lw)	[21]	(+)	32.57	1.09	29.27	35.02		5	4
Tracheid diameter tangential	[μm]	*20–40*	[22]	()	33.32	1.75	27.69	36.28		0	0
Tracheid wall thickness	[μm]	*3.5* (ew), *10.7* (lw)	[23]	(–)	2.49	0.18	2.07	2.84		0	0
Fiber length	[μm]	*2800, max. 4800*	[24]	+	5040	382	4013	5929		0	0
Microfibril angle	[μm]	*5–30, 2–44*	[25]	–	11.8	2.8	7.4	21.9		0	0
COLOR											
Brightness	[ ]	*84.23*	[26]	+	89.887	1.45	85.96	92.24		2	3
Red	[ ]	*3.09*	[27]	–	2.093	0.788	0.895	4.783		0	0
Yellow	[ ]	*19.45*	[28]	+	21.217	1.220	18.534	24.103		4	0
SWELLING											
Max. swelling radial	[%]	*3.7*	[29]	(–)	3.63	0.82	1.74	5.58			0

MULTIPLE LINEAR REGRESSION MODELS - RESULTS

R	Multiple correlation coefficient	0.300	0.800	0.802
R Square	Coefficient of determination	0.090	0.640	0.643
Durbin Watson	Autocorrelation test of residuals	1.027	1.616	1.535
Sig.	Significance of the total model (F-Test)	0.006	0.000	0.000
VIF	Highest out of the entered predictors; collinearity test	1.000	2.897	1.176

Abbreviations: MOE=modulus of elasticity,res. freq.=resonance frequency 1st mode,log. dec.=logarithmic decrement (damping),VIF=variance inflation factor,ew=earlywood,lw=latewood,3-p. b.=three point bending test.
0=normal growth pattern;1="hazel growth" with indented rings.
Explanation of the columns (column names in capitals): VARIABLES: All dependent variables and predictors that were used in this study, grouped in categories. UNIT: Unit of the variable—valid for the according values in Table III. LITERATURE VALUE: Information from literature, giving typical ranges or values for each parameter, to allow the evaluation of the measurements performed in the current study. Values and citations in italic style describe properties of normal Norway spruce wood or wood in general. Values in standard style are resonance wood specific. CITATION LIST: Numbers in squared brackets link to the references below the table. QUALITY CORRELATION: Gives basic information (derived from literature) on the correlation between the measured parameters and the resonance wood quality. (+= the higher the better the quality, −= the lower the better the quality, ⋆= more complex relation, ±= contradictory statements were found, >< = within the given range, ()= no statement available or an assumption). MEAN: mean value of the 84 valid cases (resonance wood samples). SDEV, MIN, MAX: the associated standard deviation, the minimum value and the maximum value. MODELING: Illustrating the results of the three different models A (prediction of the acoustical grading), B (prediction of the optical grading) and C (prediction of the over all grading). (x=dependent variable. 0,1,2,3...=independent variables (predictors) admitted in the respective model. 0 means that the admitted variable did not enter the model. 1,2,3... indicated the order of the variables that entered the model.)
Citation list to Table III appears before list of references.

to be recoded to binary (dummy) variables or other types of contrast variables. The only nonquantitative independent variable was the binary variable of growth pattern indicating if indented rings were observed or not.

Concerning the dependent variable itself, the measure was actually ordinal covering five different categories (grades). It was assumed that the difference in quality between grades 1 and 2 was as big as the difference in quality between, e.g., grades 4 and 5. Furthermore, the averaging of the 14 single grading results of the violin makers gave the variable a "quasi-continuous" character. It has been concluded that it is therefore admissible to apply a multiple linear regression routine instead of using discriminate analysis.

Before running a regression, a scatter plot between dependent and independent variables overlaid with a best-fit line was examined to determine whether a linear model is reasonable for these variables. Additionally, the following tests and assumptions were made: error term has a normal distribution with a mean value of 0, normal distribution of the dependent variable, independence of all observations, *p-p* plots of residuals, test for heteroscedasticity, check for outliers, test for autocorrelation of residuals (Durbin-Watson test), collinearity test (variance inflation factor)—see Table III. The predictor-variable selection method "stepwise forward" included the variable with the highest partial correlation in each step using the default significance level of 0.005 for entry and 0.01 for exclusion.

VI. RESULTS

The results from the test series, the subjective grading by violin makers as well as the multiple linear regression models, are shown in Table III. Generally the measurement results obtained in this project are in accordance with the resonance wood specific values that can be found in the literature (Table III. Literature Value) as far as references are available. The mean value with the corresponding standard deviation and the minimum/maximum values shall provide an insight into the characteristics of Norway spruce resonance wood (*Picea abies* (L.) Karst.). Table III also shows the concept of the three multiple linear regression models that have been set up to answer the main questions of this project.

The models A (acoustical grading), B (optical grading) and C (total grading) have different dependent variables (indicated by "x" in Table III). The material properties, which have been allowed as predictors in the respective model, are marked with figures in column A, B or C. The predictors that entered the model are made apparent by numbers higher than zero indicating the order of entry. A brief summary of the model results is given at the bottom of Table III. The multiple correlation coefficient R describes the strength of the linear correlation between the observed and model-predicted values of the dependent variable. R can reach values between 0 and 1. A high value indicates a strong relationship. R Square, the coefficient of determination, is the squared value of the multiple correlation coefficient. It shows the portion of the variation, which is explained by the model. The coefficient of determination yielded by Model A ($r^2=0.090$, Table III) shows that the model cannot explain a reasonable part of the variation within the subjective grading for acoustical suitability of the resonance wood samples for violin making. There is no significant correlation between acoustical material characteristics and the violin makers' grading. Model B ($r^2=0.640$) led to our conclusion that structural and optical properties are accessible to the senses of the violin makers and are used for the selection of the resonance wood in practice. The fact that neither mechanical wood properties nor acoustical material characteristics were able to contribute to the explanation of the variation within the subjective grading of the overall resonance wood quality (Model C; $r^2=0.643$) indicates that violin makers cannot assess these features without tools.

The Durbin-Watson is meant to detect serially correlated (or autocorrelated) residuals. As a rough rule one can say that values between 1.5 and 2.5 are acceptable. The variance inflation factor serves as a tool to detect collinearity in a linear regression model. Values above 10 indicate the existence of collinearity. Significance values of the F statistic below 0.05 mean that the variation explained by the model is not due to chance.

VII. DISCUSSION OF THE SUBJECTIVE GRADING MODELS

A. Model A: Prediction of the grading for acoustical quality

The multiple linear regression model A shows that it is not possible to predict the subjective grading of the acoustical quality of Norway spruce (*Picea abies* (L.) Karst.) resonance boards by using only acoustical characteristics.

Density entered the model as the only predictor, but density is rather a physical-anatomical parameter than an acoustical one. Nevertheless it determines acoustical and mechanical properties to a high extent (e.g., Kollmann and Côté, 1969; Niemz, 1993; Holz, 1967; Joppig, 2003; Bucur, 2006). This indicates that the violin makers use nonacoustical parameters to estimate the acoustical quality of their raw material.

From the inability of the model to describe a reasonable portion of the variance of the dependent variable, it can be assumed that the perception of acoustical properties without tools is too difficult to be used in practice. This leads to the conclusion that craftsmen seem to take a detour through wood properties that are easier to estimate to appraise the acoustical characteristics of resonance wood.

B. Model B: Prediction of the grading for optical quality

The approach to predict the subjective quality grading by optically perceptible material characteristics as predictors in a multiple linear regression model leads to better results than the acoustical model A.

The regression model B shows that violin makers strongly take into account structural-anatomical parameters. The annual ring structure is within the ambit of their approach. Special emphasis is not only on ring widths but also on their regularity.

Another property easy to consider is the color of the wood. According to the model results, color components assist the violin makers in their choice of resonance wood. Both color components that entered the model can be interpreted. The brightness describes the latewood proportion, the occurrence of compression wood as well as stain and decay, that make the wood appear darker. Brighter wood is generally preferred, favoring homogeneity and lower density. The color component yellow is positively correlated with the quality estimations, because it is a clear sign that the resonance wood sample does not contain compression wood, which is intensely red. Furthermore, violin makers tend to apply a yellow base coat onto the wood prior to varnishing. High yellow color components assist this procedure.

Finally, the radial tracheid diameter entered the model with a positive correlation to quality. This can be interpreted as a substitute for the earlywood proportion and the density. Tracheids in the earlywood zone have larger radial diameter than in the latewood (Burckel and Grissino-Mayer, 2003).

In general, a good model fit could be obtained with simple parameters. It can be concluded that the violin makers' decisions for the optical quality of the resonance wood samples are well based on macroscopically perceptible structural characteristics and therefore predictable by a multiple linear regression model to a reasonable extent.

C. Model C: Prediction of the grading for total quality

Prediction of the subjective total grading from the complete set of measured material characteristics gives a result similar to that from optical model B, indicating that violin makers are mainly guided in their decisions by structural-anatomical parameters. Tree ring related parameters are the most important ones for luthiers when estimating the overall quality for violin making. Special emphasis is not only on ring widths but also on their regularity, as evidenced by the fact that the ring width coefficient of variation entered the model as the first variable.

Again the brightness entered the model, which can be interpreted as in model B. The radial tracheid diameter entered this model with a positive correlation to quality just as in the optical model B, underlining its importance. The interpretation is identical to that one in the optical model B.

Once again a good model fit could be obtained with simple parameters. But it is surprising that none of the mechanical or acoustical predictors could enter the model. Only a combination of numerous wood properties leads to a considerably good correlation with the subjective grading results—a fact that is well known from the visual strength grading of timber, which yields correlation coefficients within the same range.

D. General discussion and conclusions

The three models A, B and C demonstrated that the acoustical and mechanical parameters do not influence the selection of resonance wood in the practice of violin making.

It is, on the other hand, clearly noticeable that most craftsmen make their decisions on the basis of optical perception. Knocking on the small board to estimate sound velocity or the pitch, or other methods to estimate acoustical material constants like damping seems to be too difficult to obtain a high correlation with the measurements of the corresponding parameter.

Would the transverse bending modulus of elasticity transversal to fiber have entered the model, if the boards had been thin enough to be bent as guitar makers do? We know that all parameters that entered the models are detectable by eye. As a result the speculation arises that the violin makers in fact choose a certain piece of wood, because the piece looks like one that was successfully turned into a good musical instrument in the past. The finding of a match is also strongly influenced by traditional knowledge and practical education of the craftsmen. Numerous descriptions and statements of luthiers and guitar makers (e.g., Borchardt, 2004; Romanillos, 1998) point out that the construction of the instrument is of major importance if the raw material fulfills certain minimum quality requirements. Instrument makers react to the piece of resonance wood by shaping it according to its unique properties.

The results at hand show that a prior estimation of acoustical properties is not vital to build good musical instruments, as the acoustical behavior can be modified during the construction process. But only regularly structured (Zieger, 1960; Holz, 1972; Holz, 1984), bright (Buksnowitz, 2006), stiff (Ziegenhals, 1999; Bariska, 1978) and light (Holz, 1966; Bucur, 1983; Feuerstein, 1935; Rajcan, 1991, Ziegenhals, 2001) resonance wood holds the potential to be turned into a good resonance top.

Master violins and other musical instruments can very rarely be investigated by applying destructive measurement techniques (e.g., tensile test for the determination of the modulus of elasticity), which limits new insights to cases where instruments are, e.g., accidentally destroyed (e.g., Schwalbe and Becker, 1920; Schwalbe and Schepp, 1925). For industrial manufacturers of musical instruments, reproducible quality plays an important role. Because the manufacturing process cannot react optimally to every single piece of wood individually, the different material properties cannot be compensated. In this case, knowledge of the correlation between measurable material characteristics and acoustical quality could serve as the basis for a new grading tool. The assessment of optical parameters such as color, and annual ring structure and material characteristics like density could be used to mimic the subjective methods of the violin makers. Measurements of sound velocity, damping or stiffness could be used as additional input data. Integrating acoustical and mechanical material properties in the decision for constructional details and shape could partially substitute for the individual treatment of every piece of resonance wood by luthiers. The fact that instruments occasionally become famous without meeting the optical requirements (e.g., they contain irregularities or small defects) can be seen as a clear sign that wood sounds good in an instrument for multiple reasons. Hazel growth can be interpreted as a growth irregularity occurring in very diverse specificity. Nevertheless, it is seen as something special by most of the violin makers and wood scientists (Ziegler and Merz, 1961; Romagnoli *et al.*, 2003; Feuerstein, 1935; Bariska, 1978; Zimmermann, 1996;

Zieger, 1960). On the other hand, hazel growth spruce is sometimes rejected from an acoustical point of view (Holz, 1966) and because it is more difficult with which to work. The question arises if it is the violin maker with the better developed skills, who dares to take the riskier piece of wood and makes it something outstanding, or if it is the piece of wood with some unique structure that was destined to develop an outstanding sound.

Many interesting questions may be answered with the data at hand. Taking a closer look at the parameters behind the violin makers' choice is just the first step to a better understanding of resonance wood. The presentation and discussion of the study's results at a meeting with the participating violin makers revealed that they are aware of a very diverse interpretation of resonance wood quality. They generally agreed on the fact that a more objective approach to evaluate the acoustical and mechanical potential of their raw material (resonance wood) is needed. Without the ability of the craftsmen to react on a variation of wood properties, no superior instruments could be built. The group agreed that assisting tools or precise guidelines to help in choosing the raw material, including mechanical and acoustical parameters, are desirable. The objective definition of quality classes or grades giving ranges of suitability for each parameter will therefore be the next step.

ACKNOWLEDGMENTS

We thank all violin makers who took part in the subjective grading of the resonance wood as well as numerous resonance wood retailers and individual persons for the donation of samples (Ciresa Italy, IVALSA CNR, Provincia Autonoma di Trento – Paneveggio state forest, Rivolta). Many thanks for their contribution to the project also go to Holzcluster Tirol, Holzcluster Steiermark, Schaffer Sägewerk-Holzexport GmbH and COST (European Cooperation in the field of Scientific and Technical Research).

CITATION LIST TO TABLE III

Note: For full citations refer to bibliography (references).
Explanation: The Link number is followed by the citations to the values cited in third column of Table III (appearing in the same order as in the table). A "/" separates these references from the ones that refer to the information on the quality correlations in the fifth column. These are followed by links to additional literature on that specific topic in squared brackets. Dashes at any of these three positions indicate that no literature was found on this topic.

1. —/Feuerstein (1935), Bariska (1978), Zieger (1960), Holz (1966) [Ramagnoli *et al.* (2003), Ziegler and Merz (1961), Zimmermann (1996)]
2. Bucur *et al.* (1999), Holz (2000)/Holz (2000) [Bucur (1983), Feuerstein (1935), Holz (1984), Kollmann (1951, 1983)]
3. Bucur *et al.* (1999), Holz (1984), Burmester (1965, 1968)/Holz (1984) [see 1]
4. Zieger (1960)/Bariska (1978), Beldan and Pescaru (1996), Gough (2000) [Biernacki and Beall (1993), Bucur (1983), Bucur and Böhnke (1994), Bucur and Feeney (1992), Holz (1967, 1972, 1973, 1984), Ille (1975), Kollmann (1983), Kollmann and Krech (1960), Sakai *et al.* (1990), Ziegenhals (2001), Zieger (1960)]
5. See 3.
6. Holz (1966)/— [Den Hartog (1985), Niemz (1993), Niemz *et al.* (1997)]
7. See 5.
8. Niemz *et al.* (1997), Sell (1989)/— [Bariska (1978), DIN 52 186 (1978), Holz (1966, 1973, 1984), Niemz *et al.* (1997), Koponen *et al.* (2004)]
9. Niemz *et al.* (1997), Wagenführ (1996)/—[—]
10. Treu and Hapla (2000), Hearmon (1948)/Ziegenhals (1999), Zimmermann (1996) [Holz (1984), Kollmann and Côté (1968)]
11. —/—[—]
12. Kollmann (1951/1982)/— [Ashby (1999), Bodig and Jayne (1982)]
13. Wagenführ (1996), Bosshard (1982)/—[—]
14. Kollmann and Côté 1969/—[—]
15. Wagenführ (1996), Bosshard (1982)/—[—]
16. ÖNORM B 3012 (2003), Burmester (1968)/—[—]
17. Burmester (1968), Bucur (1983), Bucur *et al.* (1989), Rajcan (1991)/ Feuerstein (1935) [Beuting and Klein (2003), Holz 1966, Joppig (2003)), Treu and Hapla (2000), Zieger (1960)]
18. Feuerstein (1935), Holz (1984)/Yano *et al.* (1994) [Beuting and Klein (2003), Blossfeld *et al.* (1962), Holz (1966), Ille (1976)]
19. Holz (1984)/Holz (1972), Zieger (1960) [Bariska (1978), TGL 15799/ 12 (1982), Ziegenhals (2001)]
20. Holz (1984)/Zieger (1960)/[Beuting and Klein (2003), Ziegenhals (2001)]
21. [Burckle and Grissino-Mayer 2003]
22. Bosshard (1982)/—[—]
23. Wagenführ (1996)/[—]
24. Wagenführ (1996)/Schnur (1985)[Burmester (1965), Ille (1976), Schultze-Dewitz (1959)]
25. Wagenführ (1984), Saranpää *et al.* (1997)/Hori *et al.* (2002) []
26. Kucera *et al.* (1998)/personal communication to Cremonese and Viennese Violin makers [—]
27. See 25.
28. See 25.
29. Mombächer (1988)/— [DIN 52 186 (1978), Ille (1975), Wagenführ *et al.* (2005)]

Anderson, B. E., and Strong, W. J. (**2005**). "The effect of an inharmonic partial on pitch of pianos," J. Acoust. Soc. Am. **117**, 3268–3272.
Ashby, M. F. (**1999**). *Materials Selection in Mechanical Design* (Butterworth/Heinemann, Oxford), ISBN 07506 4357 9, p. 202.
Bariska, M. (**1978**). "Klangholz, Holzinstrumente, Musik," ("Resonance wood, wooden instruments, music"), Naturwiss. Rundsch. **31**, 45–52.
Beldan, E. C., and Pescaru, P. (**1996**). "Research on the acoustic quality classes of resonance spruce wood in Romania", *Tenth International Symposium on Non-Destructive Testing of Wood*, Lausanne Switzerland, August 26–28, pp. 43–52.
Beuting, M., and Klein, P. (**2003**). "Holzkundliche und dendrochronologische Untersuchungen an Resonanzholz," ("Wood-scientific and dendrochronological investigations on resonance wood"), annual report of the Bundesforschungsanstalt für Forst- und Holzwirtschaft, 71–72.
Biernacki, J. M., and Beall, F. C. (**1993**). "Development of an acousto-ultrasonic scanning system for NDE of wood and wood laminates," Wood Fiber Sci. **25**, 289–297.
Blossfeld, O., Haasemann, W., and Haller, K. (**1962**). "Klangholz und Klangholzsortierung," ("Sound wood and sound wood grading"), Sozial. Forstwirtsch. **12**, 140–145.
Bodig, J., and Jayne, B. A. (**1982**). *Mechanics of Wood and Wood Composites* (Van Nostrand Reinhold, New York). ISBN 0-442-00822-8, p. 53.
Borchardt, G. (**2004**). Verbal communication in Cremona Italy, P.zza S.A.M. Zaccaria, 11.
Bosshard, H. H. (**1982**). *Holzkunde. Band 1. Mikroskopie und Makroskopie des Holzes (Woodscience. Vol. I. Microscopy and Macroscopy of Wood)*, 2nd ed. (Birkhäuser, Stuttgart), ISBN 3-7643-1328-5, p. 83.
Bucur, V. (**1983**). "Vers une appréciation objective des propriétés des bois du violon." ("An attempt at an objective appreciation of the properties of wood for violins"), Revue-Forestiere-Francaise (France) **35**, 130–137.

Bucur, V. (**2006**). *Acoustics of Wood*, 2nd ed., (Springer, New York) ISBN 1431-8563, Chap. 7.
Bucur, V., and Böhnke, I. (**1994**). "Factors affecting ultrasonic measurements in solid wood," Ultrasonics **32**, 385–390.
Bucur, V., and Feeney, F. (**1992**). "Attenuation of ultrasound in solid wood," Ultrasonics **30**, 76–81.
Bucur, V., Clément, A., Bitch, M., and Houssement, C. (**1999**). "Acoustic properties of resonance wood and distribution of inorganic components of the cell wall," Holz Roh-Werkst. **57**, 103–104.
Buksnowitz, C. (**2006**). Resonance wood of *Picea abies*, Doctoral thesis - Institute of Wood Science and Technology, Vienna University of Natural Resources and Applied Life Sciences - BOKU, Vienna.
Burckle, L., and Grissino-Mayer, H. D. (**2003**). "Stradivari, violins, tree rings, and the Maunder Minimum: A hypothesis," Dendrochronologia. **21**, 41–45.
Burmester, A. (**1965**). "Zusammenhang zwischen Schallgeschwindigkeit und morphologischen und mechanischen Eigenschaften von Holz," ("Correlation between sound velocity and morphological and mechanical wood properties"), Holz Roh-Werkst. **23**, 227–236.
Burmester, A. (**1968**). "Untersuchungen über den Zusammenhang zwischen Schallgeschwindigkeit und Rohdichte, Querzug-sowie Biegefestigkeiten von Spanplatten," ("Investigations on the correlation between sound velocity and density, bending strength and tensile strength transversial to the fiber direction of particle boards"), Holz Roh-Werkst. **26**, 113–117.
Den Hartog, J. P. (**1985**). *Mechanical Vibration* (Dover, New York) ISBN 0486647854.
DIN 52 033 (**1979**). "Farbmessung Grundbegriffe der Farbmetrik" ("Color measurement; basic terminology of color metrics"), German Institute for Standardization.
DIN 52 184 (**1979**). "Prüfung von Holz. Bestimmung der Quellung und Schwindung." ("Testing of wood; determination of swelling and shrinkage"), German Institute for Standardization.
DIN 52 185 (**1976**). "Prüfung von Holz. Bestimmung der Druckfestigkeit parallel zur Faser," ("Testing of wood; compression test parallel to grain"), German Institute for Standardization.
DIN 52 186 (**1978**). "Prüfung von Holz. Biegeversuch," ("Testing of wood; bending test"), German Institute for Standardization.
DIN 52 188 (**1979**). "Prüfung von Holz. Bestimmung der Zugfestigkeit parallel zur Faser," ("Testing of wood; determination of ultimate tensile stress parallel to grain"), German Institute for Standardization.
DIN 6721-1 (**2003**). "Kunststoffe. Bestimmung dynamisch-mechanischer Eigenschaften. Teil 1: Allgemeine Grundlagen," ("Plastics – Determination of dynamic mechanical properties – Part 1: General principles (ISO 6721-1:2001)"), German Institute for Standardization.
DIN 6721-3 (**1996**). "Kunststoffe. Bestimmung dynamisch-mechanischer Eigenschaften. Teil 3: Biegeschwingung Resonanzkurven-Verfahren." ("Plastics – Determination of the dynamic mechanical properties – Part 3: Flexural vibration, Resonance curve (ISO 621-3: 1994)"), German Institute for Standardization.
Evans, R. (**1994**). "Rapid measurement of the transverse dimensions of tracheids in radial wood sections from Pinus radiate," Holzforschung **48**, 168–172.
Evans, R. (**1999**). "A variance approach to the x-ray diffractometric estimation of microfibril angle in wood," Appita J. **52**, 283–289.
Evans, R., and llic, J. (**2001**). "Rapid prediction of wood stiffness from microfibril angle and density," For. Prod. J. **51**, 53–57.
Evans, R., Hughes, M., and Menz, D. (**1999**). "Microfibril angle variation by scanning x-ray diffractometry," Appita J. **52**, 363–367.
Evans, R., Stuart, S. A., and Van der Touw, J. (**1996**). "Microfibril angle scanning of increment cores by x-ray diffractometry," Appita J. **49**, 411–414.
Feuerstein, A. (**1935**). "Das Klangholz," ("Resonance wood"), Forstwiss. Centralbl. **57**, 617–624.
Gindl, W., Hansmann, C., Notgurga, G., Schwanninger, M., Hinterstoisser, B., and Jeronimidis, G. (**2004**): "Using water-soluble melamin-formaldehyd resin to improve the hardness of Norway spruce wood." J. Appl. Polym. Sci., **93**(4), 1900-1907.
Gough, C. (**2000**). "Science and the Stradivarius," Physics World, April 2000, http://physicsweb.org/article/world/13/4/8/1.
Grabner, M., and Wimmer, R. (**2006**). "Variation of different tree-ring parameters in samples from each terminal shoot of a Norway spruce tree," Dendrochronologia. **23**, 111–120.
Gureyev, T. E., and Evans, R. (**1999**). "A method for measuring vessel-free density distribution in hardwoods," Wood Sci. Technol. **33**, 31–42.
Hearmon, R. F. S. (**1948**). *The Elasticity of Wood and Plywood* (For. Prod. m. Res. Spec. Rep. No. 7, London).
Holz, D. (**1966**). "Untersuchungen an Resonanzhölzern. 1. Mitteilung: Beurteilung von Fichtenresonanzhölzern auf der Grundlage der Rohdichteverteilung und der Jahrringbreite," ("Research on resonance wood: Part 1: Quality assessment on the basis of density distribution and tree ring width"), Arch. Forstwes. **15**, 1287–1300.
Holz, D. (**1967**). "Untersuchungen an Resonanzholz. 3. Mitteilung: Über die gleichzeitige Bestimmung des dynamischen Elastizitätsmoduls und der Dämpfung an Holzstäben im hörbaren Frequenzbereich," ("Research on resonance wood: Part 3: Simultanious determination of the dynamic modulus of elasticity and damping on wooden rods in the audible frequency range"), Holztechnologie **8**, 221–224.
Holz, D. (**1972**). "Zur Beurteilung von Resonanzhölzern und einigen vergleichbaren Austauschstoffen durch Messungen akustisch wichtiger Eigenschaften," ("Evaluation of resonance wood and comparable substitutes by measurements of acoustical properties"), Dissertation TU-Dresden. Holtechnol. 14, (1973), 113–114.
Holz, D. (**1973**). "Untersuchungen an Resonanzholz. 5. Mitteilung: Über bedeutsame Eigenschaften nativer Nadel- und Laubhölzer im Hinblick auf mechanische und akustische Parameter von Piano-Resonanzböden," ("Research on resonance wood: Part 5: Quality assesment on the basis of density distribution and tree ring width"), Holztechnologie **14**, 195–202.
Holz, D. (**1984**). "Über einige Zusammenhänge zwischen forstlich-biologischen und akustischen Eigenschaften von Klangholz (Resonanzholz)," ("On relations between biological and acoustical properties of resonance wood"), Holztechnologie **25**, 31–36.
Holz, D. (**1996**). "Comments on: Chemical treatment of wood for musical instrument," J. Acoust. Soc. Am. **99**, 1795–1796.
Holz, D. (**2000**). "Die Birke als Klangholz," ("Birch as resonance wood"), Bayrische Landesanst. f. Wald Forstwirts. LWF-Bericht Nr. **28**, 92–98.
Hori, R., Wantanabe, U., Müller, M., Lichtenegger, H. C., Frazl, P., and Sugiyama, J. (**2002**). "The importance of seasonal differences in the cellulose microfibril angle in softwoods in determining acoustic properties," J. Mater. Sci. **37**, 4279–4284.
Hutchins, C. M. (**1992**). "A 30-year experiment in the acoustical and musical development of violin-family instruments," J. Acoust. Soc. Am. **92**, 639–650.
Ille, R. (**1975**). "Eigenschaften und Verarbeitung von Fichtenresonanzholz für Meistergeigen," ("Properties and processing of Norway spruce resonance wood for masters' instruments"), Holztechnologie **16**, 95–101.
Ille, R. (**1976**). "Eigenschaften und Verarbeitung von Fichtenresonanzholz für Meistergeigen (II)," ("Properties and processing of Norway spruce resonance wood for masters' instruments – Part II") Holztechnologie **17**, 32–35.
Jeffrey, E. C. (**1917**). "*The Anatomy of Woody Plants*" (University of Chicago Press, Chicago).
Joppig, G. (**2003**). "In Schwingung versetzt Hölzer im Instrumentenbau," ("Wood in musical instrument making"), Zeitschr. Zuschnitt. **12**, 22ff.
Kollmann, F. (**1951/82**). *Technologie des Holzes und der Holzwerkstoffe (Technology of solid wood and wooden materials)* (Springer-Verlag, Berlin). ISBN 3-540-11778-4, pp. 549, 552, 607.
Kollmann, F. (**1983**). "Holz und Schall – Theorie und Nutzanwendung," ("Wood and sound – theorie and applications"), Holz-Zentralbl. **109**, 201–202.
Kollmann, F., and Côté, W. A. (**1969**). *Principles of Wood Science and Technology. Solid Wood*. (Springer, Berlin), Chap. 7, pp. 294, 276.
Kollmann, F., and Krech, H. (**1960**). "Dynamische Messungen der elastischen Holzeigenschaften und der Dämpfung," ("Dynamic measurements of elastic wood properties and damping"), Holz Roh-Werkst. **18**, 41–54.
Koponen, T., Peura, M., Karppinen, T., Hæggström, E., Müller, M., Saranpää, P., and Serimaa, R. (**2004**). "Elastic properties and cell wall structure of Norway spruce as a function of year ring," *Proceedings of the Second International Symposium on Wood Machining*, Vienna, Austria, pp. 53–59.
Kucera, L. J., Niemz, P., and Fliesch, A. (**1998**). "Vergleichende Messungen zur Ermittlung der Eigenschaften von Fichtenholz mittels Eigenfrequenz und Schallgeschwindigkeit," ("Comparative measurements to determine Norway spruce wood properties using resonance frequency and sound velocity"), Holzforsch. Holzverwert. **5**, 96–99.
Macdonald, J. R. (**1966**). "Energy dissipation and attenuation under high-loss conditions," Br. J. Appl. Phys. **17**, 1347–1354.
Mombächer, R. (**1988**). *Holzlexikon. Nachschlagewerk für die Holz- und Forstwirtschaft (Wood encyclopaedia)*, (DRW-Verlag). ISBN 3-87181-318-4, p. 133.

Niemz, P. (**1993**). *Physics des Holzes und der Holzwerkstoffe.* ("Physics of solid wood and wooden materials"), (DRW, Leinfeld-Echterdingen), ISBN 3-87181-324-9, Chap. 8, pp. 111, 119.

Niemz, P., Kucera, L. J., and Pöhler, E. (**1997**). "Vergleichende Untersuchungen zur Bestimmung des dynamischen E-Moduls mittels Schall-Laufzeit- und Resonanzfrequenzmessung," ("Comparative measurements to determine the dynamic modulus of elasticity using sound velocity and resonance"), Holzforsch. Holzverwert. **5**, 91–93.

ÖNORM B 3012 (**2003**). "Wood species – Characteristic values to terms and symbols of ÖNORM EN 13556." (Austrian Standards Institute, Zwota).

ÖNORM EN 1534 (**2000**). "Parkett und andere Hölzfussböden. Bestimmung des Eindruckwiderstandes (Brinell). Prüfmethode." ("Wood parquet flooring – Determination of resistence to indentation (Brinell) – Test Method") (Austrian Standards Institute, Zwota).

Rajcan, E. (**1991**). "Die physikalisch-akustischen Charakteristiken von Holz als Material für die Produktion von Streichinstrumenten," ("Physical and acoustical characteristics of wood as raw material for the manufacturing of bowed instruments,") Instrumentenbau-Zeit **53**, 44–45.

Ramagnoli, M., Bernabei, M., and Codipietro, G. (**2003**). "Density variations in spruce wood with indented rings (*Picea abies*)," Holz Roh-Werkst. **61**, 311–312.

Romanillos, J. L. (**1998**). *Antonio De Torres: Guitar Maker – His Life and Work* (Kahn and Averill, London) ISBN 0933224931.

Sakai, H., Minimisawa, A., and Takagi, K. (**1990**). "Effect of moisture content on ultrasonic velocity and attenuation in woods," Ultrasonics **28**, 382–385.

Saranpää, P., Serimaa, R., Anderson, S., Pesonen, E., Suni, T., and Paakkari, T. (**1997**). "Variation of microfibril angle of Norway spruce [*Picea abies* (L.) Karst.] and Scots pine (*Pinus silvestris* L.) – comparing x-ray diffraction and optical methods," *Proceedings of the IAWA/IUFRO Int. Workshop* "Significance of Microfibril Angle to Wood Quality," Westport, New Zealand.

Schnur, K. (**1985**). "Klangholzanalysen," ("Resonace wood analysis,") Das Musikinstrumentenheft. **8**, 61–63.

Schultze-Dewitz, G. (**1959**). "Variation und Häufigkeit der Faserlänge der Fichte," ("Variation and frequency of tracheid lengths in norway spruce"), Holz Roh-Werkst. **17**, 316–326.

Schumacher, R. T. (**1988**). "Compliances of wood for violin plates," J. Acoust. Soc. Am. **84**, 1223–1235.

Schwalbe, C., and Becker, E. (**1920**). "Chemische Untersuchung des Holzes einer alten Amatigeige," ("Chemical investigations on wood of an original Amati violin"), Z. Angew. Chem. **33**, 272.

Schwalbe, C., and Schepp, R. (**1925**). "Zur Kenntnis des alten Italienischen Geigenholzes," ("Ancient Italian violin resonance wood") Z. Angew. Chem. **38**, 965–966.

Sell, J. (**1989**). *Eigenschaften und Kenngrössen von Holzarten. (Properties and Material Characteristics of Wood Species)*, (Baufachverlag, Lignum). ISBN 3-85565-223-6, p. 34.

Stuart, S. A., and Evans, R. (**1995**). "X-ray diffraction estimation of the microfibril angle variation in eucalypt wood," Appita J. **48**, 197–200.

TGL 15799/12, (**1982**). "Fachbereichstandard der DDR. Rohholz, Bootsbauholz, Zündwarenholz," ("Standard of the DDR").

Treu, A., and Hapla, F. (**2000**). "Untersuchung der Qualität von Fichten- und Tannenklangholz," ("Investigations on the quality of Norway spruce and Fir resonance wood"), Allg. Forst-Jagdztg. **171**, 215–222.

Wagenführ, A., Pfriem, A., and Eichelberger, K. (**2005**). "Einfluss einer thermischen Modifikation von Holz auf im Musikinstrumentenbau relevante Eigenschaften. Teil 1: Spezielle anatomische und physikalische Eigenschaften," ("Influence of thermal modification on wood properties relevant for musical instrument making-Part 1: Specific anatomical and physical properties"), Holztechnologie **46**, 36–42.

Wagenführ, R. (**1984**). *Anatomie des Holzes, (Wood Anatomy)*, (VEB Fachbuchverlag, Leipzig), pp. 153–154.

Wagenführ, R. (**1996**). *Holzatlas, (Wood Encyclopedia)*, (Fachbuchverlag, Leipzig). ISBN 3-446-00900-0, pp. 177–178.

Washusen, R., Ades, P., Evans, R., Ilic, J., and Vinden, P. (**2001**). "Relationship between density, shrinkage, extractives content and microfibril angle in tension from three provenances of 10-year-old Eucalyptus globules Labill.," Holzforschung **55**, 176–182.

Yano, H., Kajita, H., and Minato, K. (**1994**). "Chemical treatment of wood for musical instruments," J. Acoust. Soc. Am. **96**, 3380–3391.

Ziegenhals, G. (**1999**). "Ermittlung von Auswahlkriterien für Resonanzholz," ("Selection criteria for resonance wood"), Projekt Report - Institute für Musikinstrumentenbau Zwota.

Ziegenhals, G. (**2001**). "Resonanzholzmerkmale von Gitarrendecken," ("Resonance wood properties"), Fachausschuss Musikalische Akustik in der DEGA, Sept. 2001, 20–23.

Zieger, E. (**1960**). "Untersuchungen über äußere Merkmale, Holzeigenschaften und forstgeographische Vorkommen der Resonanzholzqualitäten bei Fichte und einigen anderen Holzarten," ("Investigations on outer characteristics, wood properties and provenances of resonance wood qualities of Norway spruce and some other species"), Mitt. aus der Staatsforstverwaltung Bayern. **31**, 285–298.

Ziegler, H., and Merz, W. (**1961**). "Der Haselwuchs; Über die Beziehung zwischen unregelmäßigem Dickenwachstum und Markstrahlverteilung," ("Hazel growth: Relation between irregular radial growth and xylem ray distribution"), Holz Roh-Werkst. **19**, 1–8.

Zimmermann, U. (**1996**). "Anforderungen an das Klangholz," ("Requirements for resonance wood"), Schweizer. Zeitschr. f. d. Forstwesen. **9**, 695–702.

Some roles of the vocal tract in clarinet breath attacks: Natural sounds analysis and model-based synthesis

Philippe Guillemain[a)]
Laboratoire de Mécanique et d'Acoustique, CNRS UPR 7051, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

(Received 26 June 2006; revised 16 January 2007; accepted 17 January 2007)

A simplified physical model mainly devoted to the reproduction of some transients of clarinet-like instruments is presented. From time-frequency analyses of natural clarinet sounds, it is shown that the vocal tract can play a significant role in some attacks as well as in the permanent regime. The model proposed consists in supplying a pressure source at the entrance of a cylindrical bore attached to the mouthpiece, allowing one to reach various vocal tract configurations. For real-time synthesis purposes, a digital scheme solving the physical problem is proposed. It is shown that this synthesis model is able to reproduce some of the complex features observed during the attacks of the natural sounds analyzed, as well as known effects of the vocal tract in permanent regime. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2642173]

PACS number(s): 43.75.Ef, 43.75.Pq [NHF] Pages: 2396–2406

I. INTRODUCTION

Synthesis models based upon simplified representations of the physical functioning of reed instruments have shown their ability to produce musically relevant timbre variations with respect to commands of the player. Moreover, such models are interesting from a real-time control point of view, since they minimize the mapping between the continuous command of the player and the synthesis parameters. In counterpart, direct measurement of controls on a player in performance situation is a task difficult to handle and the complex mechanisms responsible for the sound production make it difficult to know what the musician is doing from the analysis of the sound he produces. Moreover, inverting a model requires a deep but generally unavailable knowledge of some of its parts, such as the whole resonator. Indeed, though Backus[1] considered the vocal tract influence of the player to be negligible, it is now commonly stated that the resonator of reed instruments should be seen as an association of two coupled acoustic bores in series, one corresponding to the instrument and the other to the player (see, e.g., Benade[2]).

Sommerfeldt and Strong[3] and more recently Fritz[4] have studied experimentally and numerically, out of the real-time synthesis context, how the impedance peaks of the vocal tract may modify some features of the clarinet sound (spectrum, playing frequency). Scavone[5] proposed a real-time oriented synthesis model and discussed in another paper[6] several "physically informed" attack models including noisy components.

While these authors focus on the role of the vocal tract, modeled for some of them[3,5] with analogous electric circuits, during the permanent regime, the present paper investigates vocal tract influences during breath attacks without mechanical action on the reed and proposes a highly simplified distributed model for real-time sound synthesis. The aim of this paper is not to present a systematic study of vocal tract manipulations used in musical performances and their consequences on the sound produced but rather to use the classical "analysis by synthesis"[7] framework to propose a model capable of mimicking natural sounds displaying specific features and to determine the corresponding set of parameters relevant both from the physical and the synthesis points of view.

After a brief recall of Fritz's[4] model, Sec. II proposes a model in which the pressure is supplied at the entrance of an upstream bore rather than at the reed level. For sake of simplicity, this upstream bore is considered cylindrical and intends to provide a model of the first impedance peak of the whole respiratory airway.

In Sec. III, time-frequency analyses of transients of natural clarinet sounds are presented and hypotheses are made to explain the complex behaviors observed during the attacks.

Thanks to the presentation of a synthesis scheme that simulates the behavior of the physical model, Sec. IV shows that during attack transients, when the blowing pressure is supplied at the entrance of the upstream bore, self-oscillations tuned on its first impedance peak may start and decrease until the steady-state regime tuned on the instrument bore is reached.

It is concluded that the synthesis model generates transient and permanent regimes sharing many common features with the natural sounds analyzed in Sec. III.

II. PHYSICAL MODEL

The classical model describing the vocal tract by its impedance seen from the reed is first briefly recalled.

A. Classical model

The classical model (see, e.g., Wilson and Beavers[8]) used to represent the link between acoustic pressure and flow in the mouthpiece of a single reed instrument is based upon

[a)]Electronic mail: guillem@lma.cnrs-mrs.fr

 0001-4966/2007/121(4)/2396/11/$23.00

the steady Bernoulli equation. It is assumed that a jet of velocity $v_j(t)$ and pressure $p_j(t)$ is formed at the end of the reed channel, and that its kinetic energy is totally dissipated (in the mouthpiece for a jet entering the resonator, in the mouth for a jet entering the player's vocal tract). It is also assumed that the cross-section $S_j(t)$ of this jet is much smaller than the mouth cross-section S_m and the resonator cross-section S. It is finally assumed that the jet cross-section is proportional to the time varying reed channel opening $S_y(t)$, determined by the reed motion: $S_j(t)=\alpha S_y(t)$. For simplicity, it will be assumed in what follows that $\alpha=1$.

With these hypotheses:

$$p_m - p_r(t) = \frac{1}{2}\rho \frac{u_r(t)|u_r(t)|}{S_y(t)^2}, \tag{1}$$

where ρ is the mean air density. p_m is the mouth pressure. $p_r(t)$ and $u_r(t)$ are, respectively, the acoustic pressure and flow at the entrance of the resonator.

In permanent regime, by decomposing the mouth pressure p_m into an imposed static (DC) component p_0 and an oscillating component $p_u(t)$ created by the acoustic coupling between the vocal tract and the body of the instrument, Fritz[4] studied the self-oscillations by considering an equivalent impedance at the reed level including the upstream and downstream bores:

$$Z_{\text{tot}}(\omega) = \frac{P_r(\omega) - P_u(\omega)}{U_r(\omega)} = Z_r(\omega) + Z_u(\omega), \tag{2}$$

where the capital letters denote the Fourier transforms. $Z_u(\omega)=-P_u(\omega)/U_r(\omega)$ and $Z_r(\omega)=P_r(\omega)/U_r(\omega)$ denote the impedances of the upstream and downstream bores.

Nevertheless, this model can be questionable from a physical point of view in the case of a time varying blowing pressure, since the static pressure p_0 is imposed at the entrance of the reed channel. This makes the mouth pressure depend only on the acoustic coupling between the two bores and on the imposed pressure but not on the vocal tract itself, independently of the instrument.

B. Proposed model

We propose a model in which the mouth pressure p_m, from now on denoted $p_m(t)$, is the consequence of a blowing pressure $p_g(t)$ imposed at the entrance of an upstream bore surrounding the mouthpiece. Though Scavone[5] considered an equivalent lung flow source, simple calculations based on measurements by Mukai[9] show that in permanent regime, the cross-section area of the glottis is about three times larger for experienced musicians and twenty times larger for inexperienced musicians than the average reed channel opening of a clarinet. Moreover, during attacks the energy source can be located at the palatal constriction level rather than at the glottis or lungs level. For these reasons, supplying a pressure seems as realistic as supplying a flow and will be more convenient for real-time synthesis purposes.

1. Resonator model

The geometry of the system and its physical variables are depicted in Fig. 1.

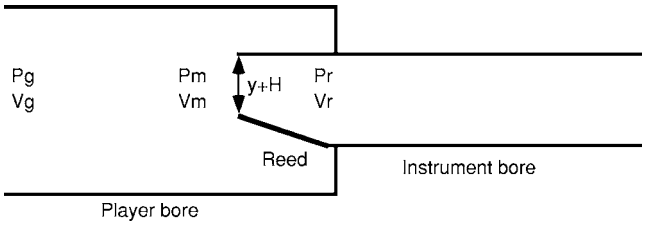


FIG. 1. Geometry of the system and physical variables.

The body of the instrument is assumed to be a perfect cylinder of equivalent (including only the imaginary part of the radiation impedance[10] as a length correction) length L and radius R, with linear propagation of plane waves. In this case, its input impedance linking the Fourier transforms of the pressure $p_r(t)$ and flow $u_r(t)$ in the mouthpiece of the instrument is

$$Z_r(\omega) = jZ_c \tan(k(\omega)L). \tag{3}$$

$Z_c=\rho c/S$ is the characteristic impedance of the bore. S is its input cross-section. $k(\omega)$ is the wave number including viscothermal losses:[11] $k(\omega)=\omega/c-j^{3/2}\eta c\sqrt{\omega}$, where $\eta=1/(Rc^{3/2})(\sqrt{l_v}+(c_p/c_v-1)\sqrt{l_t})$. Typical values of the physical constants, in mKs units, are: $c=340$, $l_v=4\times10^{-8}$, $l_t=5.6\times10^{-8}$, and $c_p/c_v=1.4$.

For sake of simplicity, digital efficiency and controllability issues in a real-time synthesis context, the "player bore" is also considered cylindrical, with radius R_m, cross-section S_m, and length L_m. It includes frequency dependent losses and models the low frequency behavior of the vocal tract. The losses model, taken into account in the wave number denoted k_m is also based upon the classical viscothermal losses rule[11] but the geometrical radius is replaced by a smaller and adjustable radius R_p in order to take into account additional losses corresponding to those of the human tissues. Indeed, a straightforward simulation shows that this method provides an approximation of Sondhi's[12] losses model sufficient for digital sound synthesis purposes. Moreover, it provides an additional degree of freedom to control the quality factor of the first impedance peak independently of the bore length and radius. For digital synthesis purposes, it is finally assumed that k_m equals zero at zero frequency. This way, the DC component of the blowing pressure $p_g(t)$ is fully transmitted to the mouth pressure $p_m(t)$.

Obviously, in real situations, these parameters are modified dynamically by the player during the attack and during the play, but for the purpose of this paper, which mainly aims at understanding and simulating the role of the vocal tract during the transient, it will be assumed that they remain constant. The three control parameters S_m, L_m, and R_p allow one to reach different configurations, running from the classical model in which S_m is large and L_m is small, to more realistic configurations in which L_m and R_p will determine, respectively, the frequency and the quality factor of the first impedance peak of the vocal tract and S_m will determine the level of acoustic coupling between the player bore and the instrument bore.

By denoting $U_m=S_mV_m$ and $U_g=S_mV_g$ the Fourier transforms of the acoustic flows associated to the acoustic veloci-

ties v_m and v_g, the pressure and flow propagation within the player bore is described by the following transmission line equations:

$$P_g(\omega) = \cos(k_m L_m) P_m(\omega) + jZ_m \sin(k_m L_m) U_m(\omega),$$

$$U_g(\omega) = \frac{j}{Z_m} \sin(k_m L_m) P_m(\omega) + \cos(k_m L_m) U_m(\omega),$$

where $Z_m = \rho c / S_m$ is the characteristic impedance of the upstream bore.

These equations are finally written as follows:

$$P_g(\omega) = Z_m U_g(\omega) + e^{-jk_m L_m}(P_m(\omega) - Z_m U_m(\omega)), \quad (4)$$

$$P_m(\omega) = -Z_m U_m(\omega) + e^{-jk_m L_m}(P_g(\omega) + Z_m U_g(\omega)). \quad (5)$$

It is worth noting that since $P_g(\omega)$ is imposed, Eqs. (4) and (5) can be combined to remove $U_g(\omega)$:

$$P_m(\omega) = \frac{2e^{-jk_m L_m}}{1 + e^{-2jk_m L_m}} P_g(\omega) - jZ_m \tan(k_m L_m) U_m(\omega), \quad (6)$$

which shows that the resonances induced by the vocal tract are those of a quarter wave resonator. In permanent regime ($P_g(\omega) = p_0 \delta(\omega)$), this model is equivalent to Fritz's[4] model. As soon as $p_g(t)$ varies, the two models are different, due to the filtering of the blowing pressure by $2\exp(-jk_m L_m)/(1 + \exp(-2jk_m L_m))$. Therefore, during attack transients, it can be expected that a large bandwidth excitation yields oscillations of $p_m(t)$ tuned on the poles of this filter, independent of the level of the acoustic coupling between the two bores occurring in permanent regime and determined by the value of the impedance $jZ_m \tan(k_m L_m)$.

2. Flow model

If it is no longer assumed that the jet cross-section is much smaller than the mouth cross-section, the acoustic velocity $v_m(t)$ in the player bore can no longer be ignored. In this case, for a jet entering the instrument, the Bernoulli flow model reads:

$$p_m(t) + \tfrac{1}{2}\rho v_m(t)^2 = p_j(t) + \tfrac{1}{2}\rho v_j(t)^2. \quad (7)$$

Assuming flow conservation ($u_m(t) = S_m v_m(t) = u_r(t) = S_y(t) v_j(t)$) and total dissipation of the kinetic energy of the jet in the mouthpiece ($p_r(t) = p_j(t)$), Eq. (7) becomes in terms of pressure and flow variables:

$$p_m(t) - p_r(t) = \frac{1}{2}\rho u_r(t)^2 \left(\frac{1}{S_y(t)^2} - \frac{1}{S_m^2} \right). \quad (8)$$

In the same way, for a jet entering the mouth, the Bernoulli flow model reads:

$$p_r(t) + \tfrac{1}{2}\rho v_r(t)^2 = p_j(t) + \tfrac{1}{2}\rho v_j(t)^2. \quad (9)$$

Assuming again flow conservation and total dissipation of the kinetic energy of the jet in the mouth ($p_m(t) = p_j(t)$), Eq. (9) becomes in terms of pressure and flow variables:

$$p_m(t) - p_r(t) = -\frac{1}{2}\rho u_r(t)^2 \left(\frac{1}{S_y(t)^2} - \frac{1}{S^2} \right). \quad (10)$$

The reed channel opening $S_y(t)$ is determined by the product of its width w and its height $H + y(t)$ (see Fig. 1):

$$S_y(t) = w\theta(H + y(t))(H + y(t)), \quad (11)$$

where $\theta(H + y(t))$ is the Heaviside function, the role of which is to keep the reed channel opening positive ($\theta(H + y(t)) = 0$ when $y(t) \leq -H$) and to model the beating-reed phenomenon. H denotes the position of the reed at equilibrium (without any blowing pressure).

The reed is modeled as a linear single degree of freedom system and its displacement $y(t)$ is given by the following dynamic equation:

$$\frac{1}{\omega_r^2}\frac{d^2 y(t)}{dt^2} + \frac{q_r}{\omega_r}\frac{dy(t)}{dt} + y(t) = -\frac{p_m(t) - p_r(t)}{\mu_r \omega_r^2}, \quad (12)$$

where $\omega_r = 2\pi f_r$, q_r^{-1} and μ_r are, respectively, the angular frequency, the quality factor of the reed resonance, and the reed mass per unit area.

The full physical model of the functioning of the instrument is made of:

(1) The blowing pressure $p_g(t)$.
(2) Equations (4) and (5) describing the propagation of pressure and flow between each termination of the player bore.
(3) The instrument bore impedance equation (3).
(4) The reed dynamics equation (12).
(5) The reed channel opening equation (11).
(6) The flow model equations (8) and (10).

III. EXPERIMENTAL OBSERVATIONS

This section first presents time-frequency analyses of transient parts of two natural clarinet sounds. Then, it proposes an interpretation of the phenomena, linked with general features of the impedance (seen from the reed) of the vocal tract, based on measurements by Fritz[4] and simulations by Sommerfeldt.[3]

A. Analysis of natural sounds

In order to study the role of the player bore during transients, two musicians were asked to play attacks without involving a mechanical contact between the tongue and the reed. Example 1 was performed by an inexperienced player. Example 2 was played by an experienced clarinet teacher and performer on its own instrument. This musician pays a lot of attention to what he calls his "internal phonation" and is, according to him, trained to control the movements and shape of its respiratory airway when playing.

All the spectrograms and spectrogram slices presented in what follows have been computed with a Gaussian window, the width of which at half-height is 25 ms.

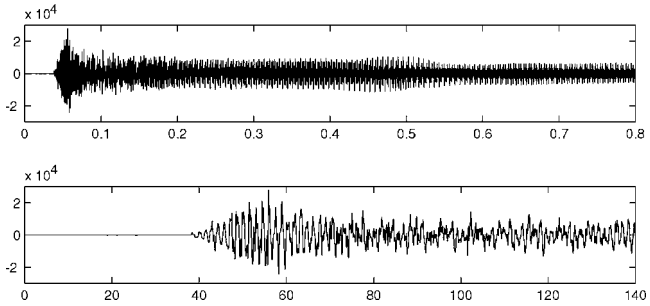


FIG. 2. Example 1. Top panel: First 0.8 s of the time signal. Bottom panel: Enlargement of the first 140 ms.

1. Example 1

The top of Fig. 2 shows the external pressure, recorded at a sampling frequency of 44.1 kHz, 1 m-away from the instrument with an omnidirectional microphone, of the first 0.8 s of an attack of a clarinet sound. The bottom shows an enlargement of the first 140 ms. The fundamental frequency, estimated in the steady-state part of the sound, is 149 Hz. The envelope of the attack shows an increasing phase, followed by a decreasing phase. The total duration of these two phases is about 40 ms. The main frequency of oscillation of the signal during these phases is around 700 Hz and is far above the fundamental frequency of the steady-state regime.

Figure 3 shows the spectrogram of the sound in the range [0–4 kHz]. During the first 0.1 s, the component of highest level is a transient component of short duration (around 40 ms), the frequency of which is around 690 Hz. This frequency does not correspond to the frequency of one of the harmonics of the sound (less than the frequency of the fifth harmonic) and the spectrogram does not show a smooth glissando between the frequency of this transient component and that of the fifth harmonic. This transient component appears before the beginning of the fifth harmonic and dies after it. One can notice around 2 kHz a similar component at a lower level. Its frequency corresponds to three times that of the first transient component. In the same way, it can be noticed around 3.4 kHz a transient component whose frequency corresponds to five times that of the main transient component. After the first 0.1 s, the fifth harmonic has the highest level. During the first 0.2 s, the spectral content of the sound is rich, with the presence of many components that seem to be neither in harmonic relationships between them nor subharmonic of the fundamental frequency. During the first 0.5 s, the amplitudes of the harmonics vary rapidly, which seems to indicate that a stable, steady state oscillation regime is not reached.

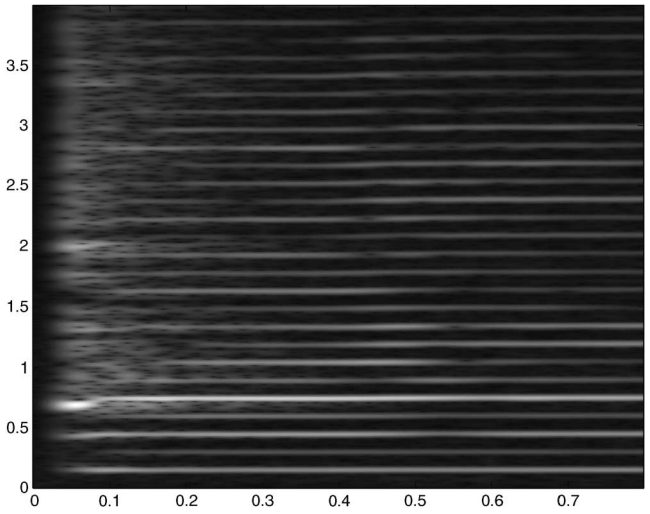


FIG. 3. Example 1. Spectrogram. Horizontal axis in seconds, vertical axis in kilohertz.

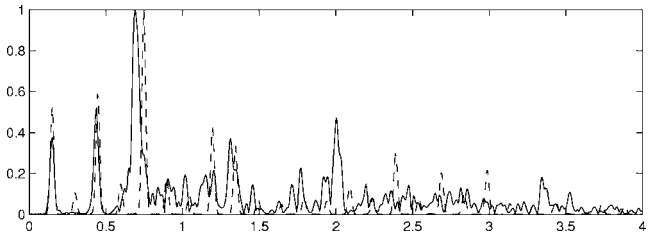


FIG. 4. Example 1. Spectrogram slices at $t=0.07$ s (solid line) and $t=0.7$ s. (dashed line). Horizontal axis in kilohertz.

This is confirmed by Fig. 4, which shows superimposed two vertical "slices" of the spectrogram at $t=0.07$ s (solid line) and $t=0.7$ s (dashed line). For clarity the curves have been normalized. The spectrogram slice corresponding to the attack shows clearly the component at 690 Hz, as well as its harmonics. Harmonic two emerges among other peaks. Harmonic three is clearly visible around 2 kHz. Harmonic four does not emerge and harmonic five is split into two components around 3.4 kHz. All these transient components are no longer visible on the slice corresponding to the permanent regime and their frequencies are different from those of the stable self-oscillations. In permanent regime, harmonics three and five have a higher level than the fundamental.

It is worth noting that other instances of the same note played by the same musician lead to similar results. It has been noticed that notes played with a lower blowing pressure lead to the death of the harmonics of the transient component at 690 Hz and to the simultaneous birth of the transient component and the fifth harmonic of the steady-state regime. In the same way, a different note (two semitones higher) played by the same musician on a different clarinet lead to a qualitatively similar behavior of the spectrogram during the transient, with the presence of a transient component around 700 Hz and a rich spectral content during the first 0.1 s.

2. Example 2

Figure 5 shows that in this example, the attack is slower than in the first example. Self-oscillations start around $t=0.2$ s. The enlargement shows that during the beginning of the transient, mainly noise seems to be present.

Figure 6 shows that the brightness of the sound is weaker than in the first sound example by considering the smaller level of high frequency harmonics in the steady state part of the sound. Between 0 and 1200 Hz, all the harmonics appear nearly simultaneously, except for harmonic three, appearing before the others, in the continuation of a noisy and low level transient component at 500 Hz. The levels of harmonics three and five are clearly higher than that of the fundamental. Similarly, harmonics seven and eight also have a

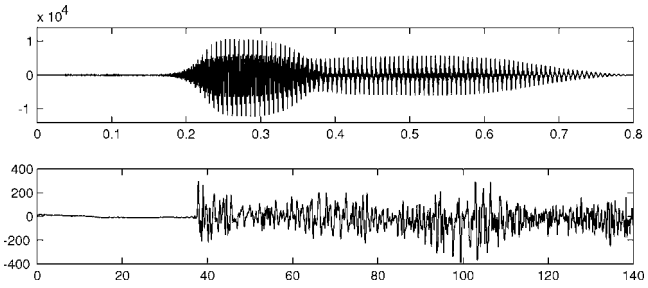


FIG. 5. Example 2. Top panel: First 0.8 s of the time signal. Bottom panel: Enlargement of the first 140 ms.

high level. From the beginning of the self-oscillations until $t=0.35$ s the components are frequency modulated and seem to remain in harmonic relationships. An inharmonicity in the whole resonator might be the responsible for this pitch variation. In the steady state regime, above the odd harmonic number seven (around 1200 Hz), the level of the even harmonics is high.

It is worth noting that a different note played by the same musician (ten semitones higher) also exhibits an early start of a high level third harmonic in the continuation of a transient component whose frequency is close to that of the third harmonic (around 850 Hz).

B. Discussion

According to musicians and measurements by Fritz,[4] two main vocal tract configurations are used in permanent regime.

The /*a*/ configuration (as in "father") is used in the lower register. In this configuration, the impedance of the player air column seen from the reed shows a peak around 300–500 Hz. The height of this peak remains small compared to those of the instrument bore.

The /*i*/ configuration (as in "see") is preferably used in the higher register. In this configuration, the impedance of the player air column seen from the reed shows a peak around 700–900 Hz. The height of this impedance peak is of the same order of magnitude as those of the instrument bore.

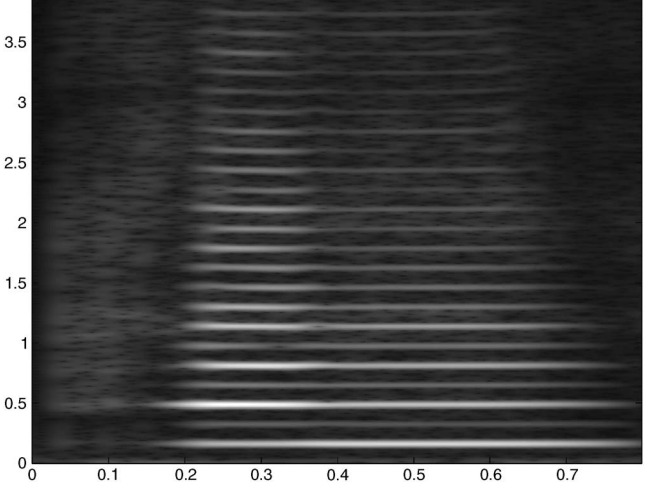


FIG. 6. Example 2. Spectrogram. Horizontal axis in seconds, vertical axis in kilohertz.

As suggested by Clinch[13] and measured by Wilson,[14] a trained player may align the first resonance of his vocal tract to an harmonic of the note played. This could correspond to the second sound example. Indeed, this sound is played by an experienced teacher, in the lower register. Thus, it can be hypothesized that the high level of harmonic three corresponds to an alignment of a resonance of the vocal tract in a configuration close to /*a*/ with a resonance of the instrument and that a strong coupling of the two bores at this frequency is responsible for the early birth of the third harmonic. Another argument in favor of this hypothesis is that a different note played by this musician shows a similar behavior.

On the contrary, the performer of the first example did not make any matching between his vocal tract configuration and the note played. The high level of harmonic five seems to indicate that he used a vocal tract configuration closer to /*i*/ than to /*a*/, with a strong resonance located between the harmonics four and five. Moreover, he supplied the pressure suddenly at a high level. Therefore, it can be hypothesized that a mechanism involving the vocal tract is responsible for the raising of a distinct transient component around 690 Hz accompanied by its harmonics and many inharmonic components. These hypotheses are sustained by the fact that a different note played by the same player on a different instrument exhibits a similar transient component at roughly the same frequency.

In order to verify these hypotheses, a synthesis model simulating the behavior of the proposed physical model is presented and studied.

IV. SYNTHESIS MODEL

This section presents a real-time oriented digital scheme allowing one to compute the variables of the physical model. This scheme is a straightforward modification of that proposed by Guillemain *et al.*[15] using digital impedance models (dimensionless Kirchhoff's variables instead of wave variables) to represent numerically the whole physical model.

The dimensionless variables, $\tilde{p}_g$, $\tilde{u}_g$, $\tilde{p}_m$, $\tilde{u}_m$ and $\tilde{p}_r$, $\tilde{u}_r$ are defined from the physical variables by the relation:

$$\tilde{p}_{g,m,r} = \frac{p_{g,m,r}}{p_M}, \quad \tilde{u}_{g,m,r} = Z_c \frac{u_{g,m,r}}{p_M},$$

where p_M is the static beating-reed pressure involving the physical parameters of the reed and is classically defined by: $p_M = \mu_r H \omega_r^2$.

A. Instrument bore

Using dimensionless variables, the input impedance $Z_r(\omega)$ given by Eq. (3) is written as

$$\tilde{Z}_r(\omega) = \frac{\tilde{P}_r(\omega)}{\tilde{U}_r(\omega)} = j \tan(k(\omega)L) = \frac{1 - e^{-2jk(\omega)L}}{1 + e^{-2jk(\omega)L}}. \tag{13}$$

The delay, dispersion, and dissipation contained in $\exp(-2jk(\omega)L)$ are modeled by a first order low-pass digital filter and an integer delay $D = E(2F_s L/c)$:

$$\exp(-2jk(\omega)L) \simeq \frac{b_0 z^{-D}}{1 - a_1 z^{-1}}, \tag{14}$$

where $E(x)$ denotes the integer part of x, F_s is the sampling frequency and $z=\exp(j\omega/F_s)$.

The coefficients b_0 and a_1 are expressed analytically[15] as functions of the length and radius of the bore by imposing that the height of the first two impedance peaks of the digital model matches those of the continuous model.

This finally leads to

$$\tilde{p}_r(n) = \tilde{u}_r(n) + V_r, \tag{15}$$

$$V_r = a_1(\tilde{p}_r(n-1) - \tilde{u}_r(n-1)) - b_0(\tilde{p}_r(n-D) + \tilde{u}_r(n-D)). \tag{16}$$

B. Player bore

The delay and losses contained in $\exp(-jk_mL_m)$ are also modeled by a first order low-pass digital filter and an integer delay:

$$\exp(-jk_mL_m) \simeq \frac{b_m z^{-D_m}}{1 - a_m z^{-1}}. \tag{17}$$

The delay D_m corresponds to $D_m=E(F_sL_m/c)$. The coefficients b_m and a_m are computed analytically so that the modulus of the digital model matches, for two given frequencies, that of the continuous model. For an easier control of the model, the first matched frequency is zero, so that the DC component of p_g can be entirely transmitted to p_m ($\tilde{P}_m(0)=\tilde{P}_g(0)$), yielding $b_m=1-a_m$. The second frequency is $c/(4L_m)$ and corresponds to the frequency of the first impedance peak of the player bore.

With this approximation, the time domain digital version of the system of equations (4) and (5) becomes

$$\tilde{p}_g(n) = \lambda\tilde{u}_g(n) + a_m(\tilde{p}_g(n-1) - \lambda\tilde{u}_g(n-1)) + b_m(\tilde{p}_m(n-D_m) - \lambda\tilde{u}_m(n-D_m)), \tag{18}$$

$$\tilde{p}_m(n) = -\lambda\tilde{u}_m(n) + a_m(\tilde{p}_m(n-1) - \lambda\tilde{u}_m(n-1)) + b_m(\tilde{p}_g(n-D_m) - \lambda\tilde{u}_g(n-D_m)), \tag{19}$$

where $\lambda=Z_m/Z_c=S/S_m$.

Since $\tilde{p}_g$ is imposed, Eq. (18) is modified so that $\tilde{u}_g$ can be calculated from $\tilde{p}_g$, $\tilde{p}_m$, and $\tilde{u}_m$.

This leads to the final set of equations describing the player bore:

$$\tilde{u}_g(n) = \frac{\tilde{p}_g(n) - V_g}{\lambda}, \tag{20}$$

$$\tilde{p}_m(n) = -\lambda\tilde{u}_m(n) + V_m, \tag{21}$$

where

$$V_g = a_m(\tilde{p}_g(n-1) - \lambda\tilde{u}_g(n-1)) + b_m(\tilde{p}_m(n-D_m) - \lambda\tilde{u}_m(n-D_m)), \tag{22}$$

$$V_m = a_m(\tilde{p}_m(n-1) + \lambda\tilde{u}_m(n-1)) + b_m(\tilde{p}_g(n-D_m) - \lambda\tilde{u}_g(n-D_m)). \tag{23}$$

C. Reed motion

The dimensionless reed model consists of replacing the reed displacement $y(t)$ by $x(t)=y(t)/H$. With this notation, the reed opening equation (11) becomes

$$S_y(t) = wH\theta(1 + x(t))(1 + x(t)) \tag{24}$$

and the reed dynamics equation (12) becomes

$$\frac{1}{\omega_r^2}\frac{d^2x(t)}{dt^2} + \frac{q_r}{\omega_r}\frac{dx(t)}{dt} + x(t) = e(t), \tag{25}$$

where $e(t)=\tilde{p}_r(t)-\tilde{p}_m(t)$ denotes the dimensionless reed excitation.

Equation (25) is discretized by the use of centered differentiation schemes: $j\omega \simeq F_s/2(z-z^{-1})$ and $-\omega^2 \simeq F_s^2(z-2+z^{-1})$. This yields the difference equation:

$$x(n) = b_{1_a}e(n-1) + a_{1_a}x(n-1) + a_{2_a}x(n-2), \tag{26}$$

where the coefficients b_{1_a}, a_{1_a}, and a_{2_a} are expressed analytically[15] as functions of ω_r and q_r.

D. Nonlinear characteristics

For the sake of digital efficiency, it is assumed that $wH/S_m \simeq 0$ and $wH/S \simeq 0$. In this case, from Eqs. (8) and (10) the dimensionless nonlinear characteristics reads:

$$\tilde{u}_r(n) = W\,\mathrm{sign}(\tilde{p}_m(n) - \tilde{p}_r(n))\sqrt{|\tilde{p}_m(n) - \tilde{p}_r(n)|}, \tag{27}$$

where W represents the reed channel opening:

$$W = \zeta\theta(1 + x(n))(1 + x(n)). \tag{28}$$

The parameter ζ corresponds to the definition by Kergomard:[16] $\zeta=wHZ_c\sqrt{2/(\rho p_M)}$.

E. Synthesis scheme

The synthesis scheme presented here calculates at any sample n the values of $\tilde{u}_g(n)$, $\tilde{p}_m(n)$, $\tilde{u}_m(n)$, $\tilde{p}_r(n)$, $\tilde{u}_r(n)$, and $x(n)$ as functions of their past values and the known blowing pressure $\tilde{p}_g(n)$:

(1) Calculate V_g, V_m, V_r with Eqs. (22), (23), and (16) and let $V=V_m-V_r$.
(2) Calculate $\tilde{u}_g(n)$ with Eq. (20).
(3) Calculate $x(n)$ with Eq. (26).
(4) Calculate W with Eq. (28).
(5) Let $\tilde{u}_m(n)=\tilde{u}_r(n)$, replace $\tilde{p}_m(n)$, $\tilde{p}_r(n)$ by their definitions from Eqs. (21), (15) into Eq. (27) and let $b_c=1+\lambda$.
(6) Solve analytically Eq. (27), yielding $\tilde{u}_r(n)=\frac{1}{2}\,\mathrm{sign}(V)\times(-b_cW^2+W\sqrt{b_c^2W^2+4|V|})$.
(7) Calculate $\tilde{p}_r(n)$, $\tilde{p}_m(n)$ with Eqs. (15) and (21).
(8) Calculate $e(n)=\tilde{p}_r(n)-\tilde{p}_m(n)$.

The external pressure $\tilde{p}_{\mathrm{ext}}(n)$ is calculated as the difference between the sum of mouthpiece pressure and flow at

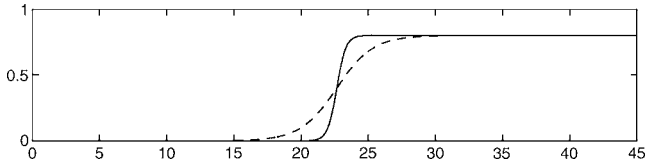


FIG. 7. Typical variations of the transient blowing pressure. Horizontal axis in milliseconds.

sample n and at sample $n-1$, corresponding to the simplest approximation of the derivative of $\tilde{p}_{\text{ext}}(t)$ since

$$\tilde{P}_{\text{ext}}(\omega) = j\omega e^{-jk(\omega)L}(\tilde{P}_r(\omega) + \tilde{U}_r(\omega)),$$

where $\exp(-jk(\omega)L)$ can be ignored from a perceptual point of view.

V. RESULTS OF SIMULATIONS

In order to validate the hypothesis made from the analysis of natural sounds, the synthesis model is used to generate sounds.

The transient variations of the blowing pressure $\tilde{p}_g(t)$ are controlled as follows:

$$\tilde{p}_g(t) = \frac{\gamma_c}{2}(1 + \tanh(\alpha(t - t_0))). \tag{29}$$

The parameter γ_c controls the pressure level and the parameter α controls the raising time from 0 to the maximum γ_c, therefore the excitation bandwidth. In the model by Kergomard,[16] γ_c corresponds to the ratio p_m/p_M. Figure 7 shows typical shapes of the transient pressure for two values of α when $\gamma_c=0.8$.

A. Upstream or downstream bore alone

These examples demonstrate that when the instrument bore is removed, corresponding to imposing $\tilde{p}_r(t)=0$, self-oscillations can start, tuned on the first peak of an impedance corresponding to the player bore closed by the reed. The reed resonance frequency is chosen high: $f_r=10$ kHz so that its role on the functioning of the model can be ignored and $q_r=0.3$. The values of the control parameters are: $\gamma_c=0.8$, corresponding to a beating-reed situation and $\zeta=0.35$. The raising of the blowing pressure is chosen fast: $\alpha=3000\ \text{s}^{-1}$.

1. Instrument bore alone

This simulation corresponds to the classical model ignoring the vocal tract. The length of the bore is $L=0.57$ m, its radius is $R=7$ mm.

The top of Fig. 8 shows the first 140 ms of the external pressure, the bottom its spectrogram over a duration of 0.8 s and on the frequency range [0–3 kHz]. Though the blowing pressure ($\tilde{p}_m(t)=\tilde{p}_g(t)$) is high and its raising is fast, the envelope of the external pressure is smooth during the attack. A permanent regime is reached at around $t=100$ ms, corresponding to a total transient duration of about 45 ms (the beginning of the sound is at $t=55$ ms).

The spectrogram shows that mostly odd harmonics are present. The birth of each harmonic is directly related to its rank. Its frequency seems constant during the transient and its amplitude decreases with respect to its rank.

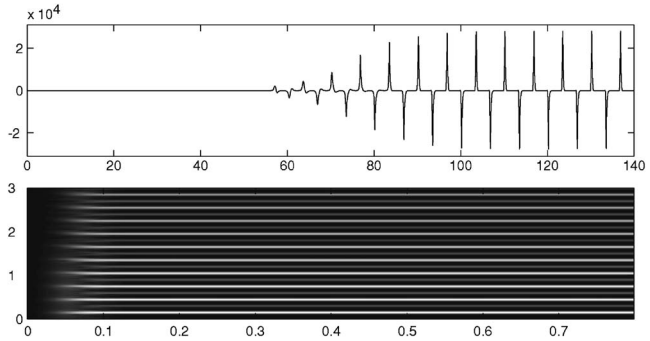


FIG. 8. Instrument bore alone. Top panel: First 140 ms the external pressure. Bottom panel: Spectrogram on 0.8 s in the range (0–3 kHz).

2. Player bore alone

This simulation corresponds to the removal of the instrument bore, obtained by setting $\tilde{p}_r(t)=0$. The length of the player bore is $L_m=0.125$ m, its radius is 2.5 times that of the instrument bore: $R_m=17.5$ mm. The equivalent radius R_p used in the calculation of the losses has been chosen five times smaller than the geometrical radius. These values have been adjusted so that the height of the first impedance peak of the player bore is close to that of the third impedance peak of the instrument bore.

The top of Fig. 9, shows the first 140 ms of the external pressure, the bottom its spectrogram over a duration of 0.8 s and on the frequency range [0–3 kHz]. Compared to Fig. 8, the raising of the external pressure level nearly follows that of the blowing pressure $\tilde{p}_g(t)$ since a steady state regime is reached after the first two periods of oscillations.

The spectrogram shows that mostly odd harmonics with constant frequencies are present. The fundamental frequency of the sound is 650 Hz and corresponds to that of the first peak of the digital impedance of the player bore. Since the delay D_m is quantified, this value differs slightly from that of the first impedance peak of the continuous impedance model which is $c/(4L_m)=680$ Hz.

These two examples show that the functioning of the player/reed and reed/instrument systems is comparable. Indeed, using the notations and results of Kergomard,[16] by set-

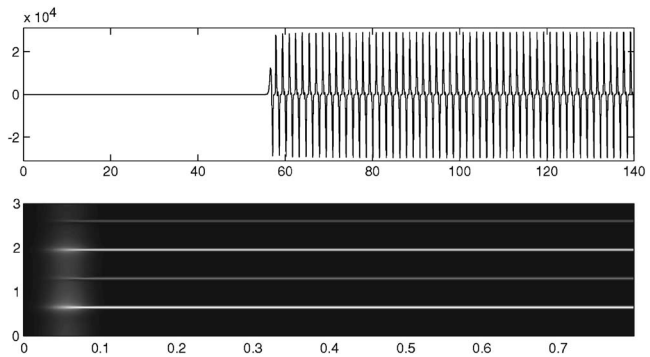


FIG. 9. Player bore alone. Top panel: First 140 ms of the external pressure. Bottom panel: Spectrogram on 0.8 s in the range (0–3 kHz).

ting $\tilde{p}_g=\gamma$ and $\tilde{p}_g-\tilde{p}_m=p$, it can be shown that for a lossless player bore and a massless reed, in permanent regime $\tilde{p}_m$ is a square signal oscillating between 0 and $2\tilde{p}_g$.

Moreover, in order to study the role of the spectral bandwidth of the blowing pressure on the attack duration, the same simulations have been performed with a slower raising of the blowing pressure: $\alpha=200\ \mathrm{s}^{-1}$. These simulations showed that the excitation bandwidth plays a little role on the raising of self-oscillations of the instrument bore alone since the attack duration remained close to 45 ms. On the contrary, a significant influence on the raising of the self-oscillations of the player bore alone was noticed since the steady-state regime was no longer reached instantaneously but after 300 ms.

B. Full model

This section presents simulations obtained with the full resonator model. The reed resonance frequency has been chosen as $f_r=2$ kHz in order to get closer to a normal playing condition, as it has been measured, e.g., by Thompson[17] and $q_r=0.3$. The values of the control and geometrical parameters were adjusted heuristically using the analysis by synthesis concept in order to generate two simulated signals corresponding to the natural examples displayed, respectively, in Figs. 3 and 6, according to the hypotheses made in Sec. III B.

1. Simulation 1

The first example has been computed with a fast raising of the blowing pressure ($\alpha=3000\ \mathrm{s}^{-1}$). The values of the control and geometrical parameters are: $\gamma_c=0.8$, $\zeta=0.35$, $L=0.57$ m, $R=7$ mm, $L_m=0.125$ m, $R_m=17.5$ mm, and $R_p=R_m/5$. This set of parameters leads to a height of the first impedance peak of the player bore comparable to that of the third impedance peak of the instrument bore and to a mistuning between the resonances of the player bore and the instrument.

Figure 10 shows from top to bottom: (1) the mouth pressure in solid line superimposed to the blowing pressure in dashed line; (2) the spectrum of the mouth pressure; and (3) the mouthpiece pressure. The transient duration is 90 ms.

Subplot (1) shows that after an instantaneous raising, the mouth pressure exhibits an oscillating behavior during the first 90 ms. During the first 50 ms after the beginning, the amplitude of the oscillations remains constant. Then it decays during 40 ms until, around $t=110$ ms a totally different regime is reached, made of stable oscillations around γ_c.

Subplot (2) shows that the frequency of the transient oscillation is tuned on the first impedance peak of the player bore and that the small oscillations of the permanent regime are due to the acoustic coupling of the two bores. Indeed, the spectrum of the mouth pressure clearly shows, at least at high frequency, the sharp (hence not localized in time) harmonics corresponding to self-oscillations of the instrument bore, and a large (hence localized in time) component at 650 Hz.

Subplot (3) shows that the mouthpiece pressure reaches its steady state level at around $t=110$ ms. Comparison between subplots (1) and (3) shows that from the beginning of self-oscillations at $t=20$ ms, the increasing phase of the oscillations of the mouthpiece pressure corresponds to a decreasing phase of the oscillations of the mouth pressure and that both pressures reach a permanent regime at the same time.

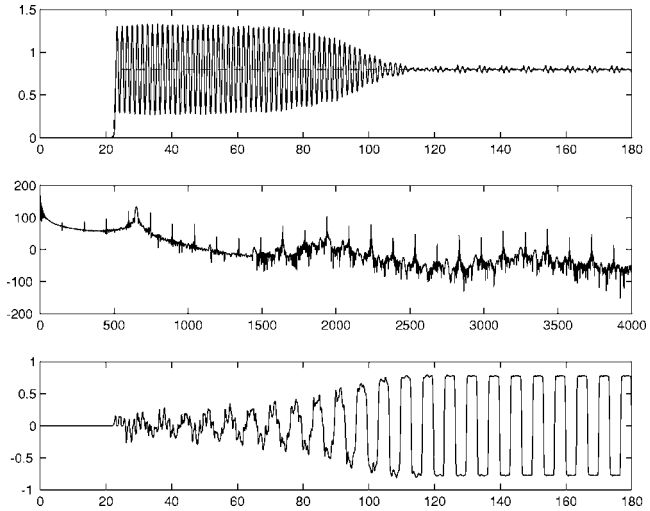


FIG. 10. From top to bottom: (1) First 180 ms of the mouth pressure (solid line) superimposed to the blowing pressure (dashed line). (2) Spectrum of the mouth pressure in decibels, horizontal axis in hertz. (3) First 180 ms of the mouthpiece pressure.

The transient behaviors observed here can be partly explained from the results of the previous subsection. While the mouthpiece pressure remains small, self-oscillations in the mouth, tuned on the first impedance peak of the player bore, starts. These oscillations die when the mouthpiece pressure becomes large enough and a stable regime, tuned on the first resonance of the instrument bore, is reached. Indeed, when the instrument bore is alone, the attack time is 45 ms while the steady-state regime is reached nearly instantaneously when the player bore is alone. The lengthening of the raising of the mouthpiece pressure (90 ms instead of 45 ms) could be caused by the oscillations of the mouth pressure, reaching periodically values (around 0.25) below the oscillation threshold[8] of the instrument bore alone (around 0.33).

Figure 11 shows, from top to bottom: (1) the transfer function between the mouthpiece pressure and flow; (2) the transfer function between the mouth pressure and flow; and (3) the equivalent impedance seen from the reed calculated as the transfer function between $\tilde{p}_r(t)-(\tilde{p}_m(t)-\tilde{p}_g(t))$ and the flow $\tilde{u}_r(t)$, corresponding to Eq. (2).

Subplots (1) and (2) show that the ratio $\tilde{P}_r(\omega)/\tilde{U}_r(\omega)$ corresponds to the input impedance of the instrument bore alone (the two curves superimpose perfectly) and that the ratio $\tilde{P}_m(\omega)/\tilde{U}_r(\omega)$ contains a peak at 0 Hz corresponding to the DC component of the supplied pressure and a peak at 650 Hz corresponding to the first impedance peak of the player bore. Up to this frequency, this behavior is similar to Scavone's[5] model and differs at high frequency, with the presence of other peaks, due to the use of a distributed element rather than lumped elements to model the vocal tract.

Subplot (3) shows that the first peak corresponding to

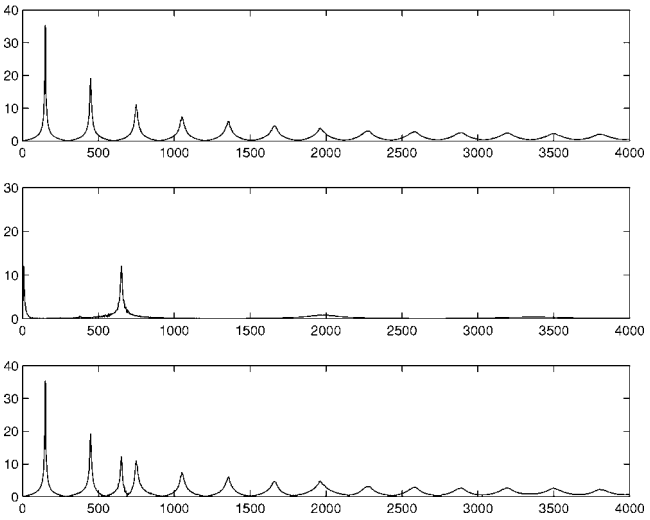


FIG. 11. From top to bottom: (1) Transfer function between mouthpiece pressure and flow. (2) Transfer function between mouth pressure and flow. (3) Equivalent impedance seen from the reed. Horizontal axes in hertz.

the player bore is slightly higher than the third peak corresponding to the instrument bore and that these peaks are not tuned.

The top of Fig. 12 shows the first 180 ms of the external pressure, the bottom two vertical slices of the spectrogram displayed in Fig. 13. Until $t=90$ ms, the external pressure exhibits a complex behavior. After $t=90$ ms, it becomes similar to the top of Fig. 8. The two vertical slices of the spectrogram, computed at $t=0.07$ s in solid line and at $t=0.7$ s in dashed line show that during the attack, the component at 650 Hz and its odd harmonics are visible and vanish totally in the steady state regime.

These behaviors can also be observed in Fig. 13, which shows that during the first 0.1 s of the attack, the spectral content of the sound is rich and inharmonic. Comparison with Fig. 3 shows many common features and most of the comments of Fig. 3 remains valid. On the natural sound, the transient component at 690 Hz appears before the fifth harmonic of the permanent regime while they appear simultaneously on the simulation. The difference in the levels of the transient components between natural and simulated sounds might be attributed to an additional radiation source. Simulations performed with a different reed resonance frequency

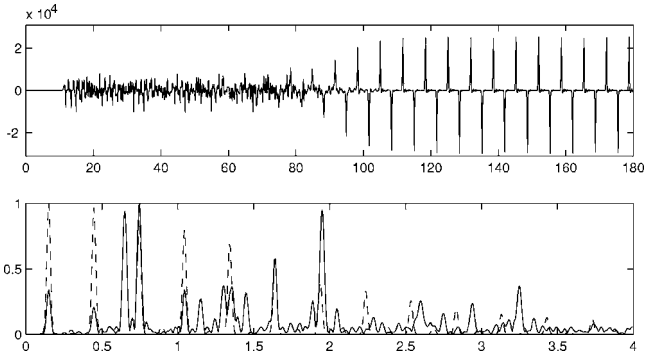


FIG. 12. Top panel: First 180 ms of the external pressure. Bottom panel: Spectrogram slices at $t=0.07$ s (solid line) and $t=0.7$ s (dashed line). Horizontal axis in kilohertz.

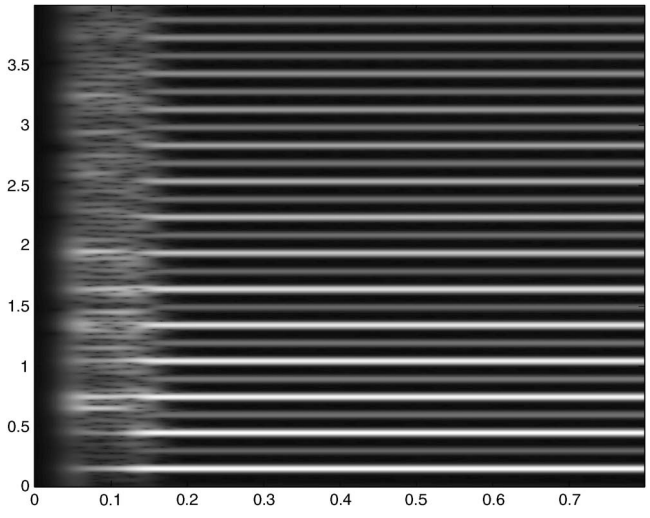


FIG. 13. Spectrogram on 0.8 s in the range (0–4 kHz) corresponding to a fast raising of the blowing pressure.

indicate that the reed does not play any role on the raising of the self-oscillations tuned on the player bore.

2. Simulation 2

This second example has been computed with a slower raising speed of the blowing pressure ($\alpha=900$ s^{-1}). The values of the control and geometrical parameters are: $\gamma_c=0.45$, $\zeta=0.3$, $L=0.52$ m, $R=7$ mm, $L_m=0.17$ m, $R_m=9.1$ mm, and $R_P=R_m/4$. The chosen length L_m is that of the vocal tract from the glottis to the mouth. Its first resonance frequency $(c/(4L_m))$ is 500 Hz and corresponds to that of the first impedance peak of the vocal tract in a neutral position (see, e.g., Mathur[18]). This set of parameters leads to a height of the first impedance peak of the player bore smaller than those of the instrument bore and to a strong coupling between the two bores, due to similar radii and to the tuning of the first player bore resonance to the second instrument bore resonance.

Figure 14 shows, from top to bottom: (1) the mouth pressure in solid line superimposed to the blowing pressure in dashed line; (2) the spectrum of the mouth pressure; and (3) the mouthpiece pressure. The chosen duration of 300 ms corresponds to that of the whole transient.

Subplot (1) shows that after a fast raising, the mouth pressure exhibits unstable oscillations until $t=150$ ms that turn into stable oscillations around γ_c. The amplitude of these oscillations is larger than in the first simulation, showing that the coupling between the two bores is more important.

Subplot (2) shows that all the harmonics corresponding to the instrument bore alone are visible. An increase of the level of the harmonics around 500 Hz can be noticed. The large baselines of all the peaks indicate nonstationary behaviors, such as frequency or amplitude modulations of the harmonics.

Subplot (3) shows that the mouthpiece pressure raises with a step-like shape and exhibits a long nonstationary part, until $t=300$ ms.

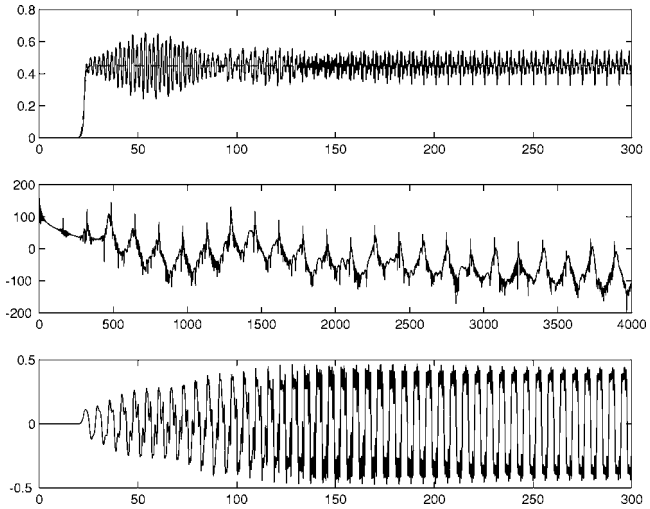


FIG. 14. From top to bottom: (1) First 300 ms of the mouth pressure (solid line) superimposed to the blowing pressure (dashed line). (2) Spectrum of the mouth pressure in decibels, horizontal axis in hertz. (3) First 300 ms of the mouthpiece pressure.

The top of Fig. 15 shows the first 300 ms of the external pressure, the bottom two vertical slices, computed at $t=0.07$ s in solid line and at $t=0.7$ s in dashed line, of the spectrogram displayed in Fig. 16. Though $\tilde{p}_g(t)$ remains constant after its raising, the amplitude of the external pressure shows a complex behavior and reaches a maximum around $t=130$ ms. During the transient, the level of harmonic three is higher than that of the others. In the permanent regime, a formant appears around 1200 Hz, with an increase of the level of the harmonics seven, eight, and nine. The fundamental frequency of the sound differs during the attack and permanent regimes.

Figure 16 shows significant frequency and amplitude modulations of all the components during the first 0.3 s as well as a high level of the harmonics five, seven, eight, and nine and an early birth of the harmonic three. Most of these features can be linked to those observed in the natural sound example 2 in Fig. 6. A simulation performed with a reed resonance frequency of 10 kHz indicates that the reed seems to be the main reason of the formant around 1200 Hz and the player bore responsible for the frequency modulation and the high level of the harmonics three and five.

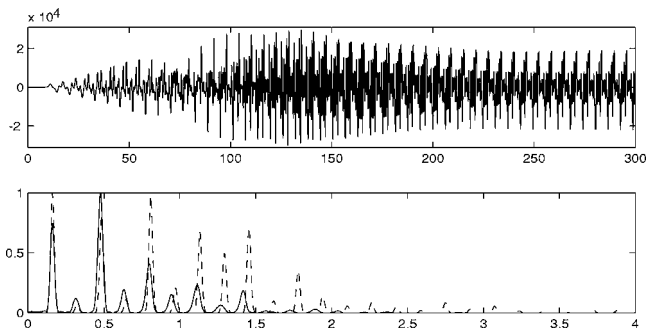


FIG. 15. Top panel: First 300 ms of the external pressure. Bottom panel: Spectrogram slices at $t=0.07$ s (solid line) and $t=0.7$ s (dashed line). Horizontal axis in kilohertz.

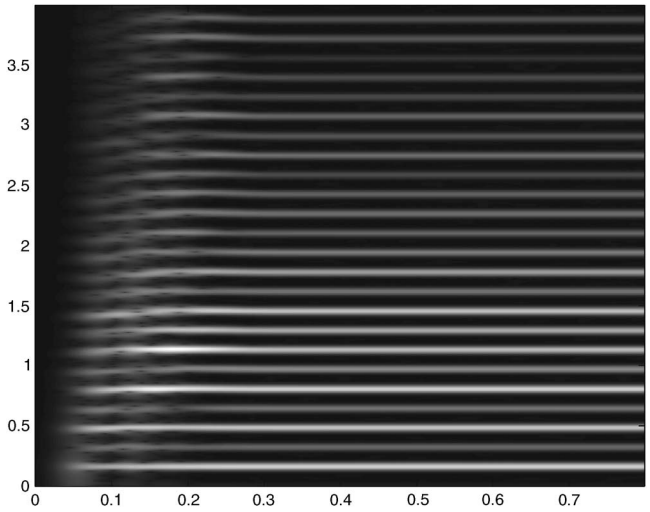


FIG. 16. Spectrogram on 0.8 s in the range (0–4 kHz) corresponding to a slow raising of the blowing pressure.

VI. CONCLUSIONS

Thanks to the introduction of a simple physical model of the association player-clarinet and a real-time synthesis scheme, it has been shown that the player's vocal tract might play an important role in some transient situations. Despite its simplicity and low computation cost, the synthesis model allows one to generate sounds sharing common features with the natural sounds considered here, both on the transient and the steady-state parts. The "analysis by synthesis" concept has been used in order to determine the parameters linked to the vocal tract from the analysis of natural sounds recorded in normal playing conditions. These analyses show that during some attacks, the spectral bandwidth of the pressure transient produced by the player can be high enough to excite a resonance of the respiratory airway, as is the case in the production of speech consonants.

Simulations using a five cylinders model of the vocal tract did not show any significant difference in the behavior of transient and permanent regimes and further works may rather take into account time-varying geometries or nonlinear effects induced by jets formation at vocal tract constriction levels and study the role of the vocal tract in the coloring of the breath noise.

Though this study is focused on the clarinet, the same effects would likely be observed and could be simulated on other reed and valve instruments. In particular, large members of the saxophone family for which the reed channel opening is much larger than that of the clarinet and the first impedance peak is low may be even more sensitive to acoustic coupling between the player tract and the instrument bore. In the same way, notes played fortissimo, yielding a high pressure level within the instrument bore, probably increase nonlinear losses in toneholes, yielding a significant lowering of the impedance peaks of the instrument, hence giving even more relative weight to those of the player tract.

Future works will also use a piloted artificial mouth to study experimentally the functioning of the instrument in calibrated transient situations and direct measurements on musicians.

Sound examples are available at: http://www.lma.cnrs-mrs.fr/~guillemain/JASA06/JASA06.htm.

ACKNOWLEDGMENTS

The author thanks Claude Crousier, professional clarinet performer and teacher, and Richard Kronland-Martinet, member of the Laboratoire de Mécanique et d'Acoustique, for playing the natural sounds presented in this paper. Jean Kergomard, member of the Laboratoire de Mécanique et d'Acoustique, and Claudia Fritz, presently at the Music Faculty of Cambridge University, are deeply thanked for their precious advice. This work is supported by the "Consonnes" project, funded by the French Agence Nationale de la Recherche.

[1]J. Backus, "The effect of the player's vocal tract on woodwind instrument tone," J. Acoust. Soc. Am. **78**, 17–20 (1985).
[2]A. Benade and P. Hoekje, "Vocal tract effects in wind instrument regeneration," J. Acoust. Soc. Am. **71**, S91 (1982).
[3]S. Sommerfeldt and W. Strong, "Simulation of a player-clarinet system," J. Acoust. Soc. Am. **83**, 1908–1919 (1988).
[4]C. Fritz, "La clarinette et le clarinettiste: Influence du conduit vocal sur la production du son," Ph.D. thesis, Univ. Paris 6 and New South Wales, France, 2005.
[5]G. Scavone, "Modeling vocal-tract influence in reed wind instruments," in *Proceedings of the 2003 Stockholm Music Acoustics Conference*, Stockholm, Sweden.
[6]G. Scavone, "Modeling and control of performance expression in digital waveguide models of woodwind instruments," in *Proceedings of the 1996 International Computer Music Conference*, Hong Kong.
[7]J. C. Risset, "Timber analysis by synthesis: Producing representations, imitations and variants for musical composition," in *Representations of Musical Signals*, edited by A. Picciali, G. de Poli, and C. Roads (MIT, Cambridge, MA, 1991), pp. 7–43.
[8]T. A. Wilson and G. S. Beavers, "Operating modes of the clarinet," J. Acoust. Soc. Am. **56**, 653–658 (1974).
[9]M. S. Mukai, "Laryngeal movement while playing wind instruments," in *Proceedings of the International Symposium of Musical Acoustics*, Tokyo, Japan, 1992, pp. 239–242.
[10]H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," Phys. Rev. **73**, 383–406 (1948).
[11]A. D. Pierce, *Acoustics* (McGraw-Hill, New York 1981), presently available from the Acoustical Society of America, New York (1990).
[12]M. M. Sondhi, "Model for wave propagation in a lossy vocal tract," J. Acoust. Soc. Am. **51**, 1070–1075 (1974).
[13]P. Clinch, G. Troup, and L. Harris, "The importance of the vocal tract resonance in clarinet and saxophone performance: A preliminary account," Acustica **50**, 280–284 (1982).
[14]T. Wilson, "The measured vocal tract impedance for clarinet performance and its role in sound production," J. Acoust. Soc. Am. **99**, 2455–2456 (1996).
[15]P. Guillemain, J. Kergomard, and T. Voinier, "Real-time synthesis of clarinet-like instruments using digital impedance models," J. Acoust. Soc. Am. **118**, 483–494 (2005).
[16]J. Kergomard, "Elementary considerations on reed-instruments oscillations," in *Mechanics of Musical Instruments*, edited by A. Hirschberg *et al.*, Lectures notes CISM (Springer, New York, 1995).
[17]S. C. Thompson, "The effect of the reed resonance on woodwind tone production," J. Acoust. Soc. Am. **66**, 1299–1307 (1979).
[18]S. Mathur, B. Story, and J. Rodriguez, "Vocal-tract modeling: Fractional elongation of segment lengths in a waveguide model with half-sample delays," IEEE Tran. on Audio, Speech and Language Processing **14**, 1754–1762 (2006).

Sound quality assessment of wood for xylophone bars

Mitsuko Aramaki[a)]
CNRS Laboratoire de Mécanique et d'Acoustique 31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France

Henri Baillères and Loïc Brancheriau
CIRAD-Forêt, TA 10/16, avenue Agropolis, 34398 Montpellier Cedex 5, France

Richard Kronland-Martinet and Sølvi Ystad
CNRS, Laboratoire de Mécanique et d'Acoustique 31, chemin Joseph Aiguier 13402 Marseille Cedex 20, France

(Received 15 March 2006; revised 22 January 2007; accepted 22 January 2007)

Xylophone sounds produced by striking wooden bars with a mallet are strongly influenced by the mechanical properties of the wood species chosen by the xylophone maker. In this paper, we address the relationship between the sound quality based on the timbre attribute of impacted wooden bars and the physical parameters characterizing wood species. For this, a methodology is proposed that associates an analysis-synthesis process and a perceptual classification test. Sounds generated by impacting 59 wooden bars of different species but with the same geometry were recorded and classified by a renowned instrument maker. The sounds were further digitally processed and adjusted to the same pitch before being once again classified. The processing is based on a physical model ensuring the main characteristics of the wood are preserved during the sound transformation. Statistical analysis of both classifications showed the influence of the pitch in the xylophone maker judgement and pointed out the importance of two timbre descriptors: the frequency-dependent damping and the spectral bandwidth. These descriptors are linked with physical and anatomical characteristics of wood species, providing new clues in the choice of attractive wood species from a musical point of view. [DOI: 10.1121/1.2697154]

PACS number(s): 43.75.Kk, 43.66.Jh, 43.60.Uv [NFH] Pages: 2407–2420

I. INTRODUCTION

The mechanical and anatomical properties of woods are of importance for the sound quality of musical instruments. Yet, depending on the role of the wooden elements, these properties may differ. Xylophone sounds are produced by striking wooden bars with a mallet, and thus the mechanical properties of the wood are important. This study is the first step towards understanding what makes the sound of an impacted wooden bar attractive for xylophone makers from a musical point of view. For this purpose, we recorded sounds from a wide variety of wood species to compare their sound quality and relate it to the wood properties. An original methodology is proposed that associates analysis-synthesis processes and perceptual classification analysis. Perceptual classification was performed by a renowned instrument maker.

The xylophone maker community agrees on the choice of wood species. This choice is driven by the sound quality, but other nonacoustically relevant properties are considered as well (e.g., robustness; esthetic aspects). The wood species most used in xylophone manufacturing is *Dalbergia* sp. Several authors have sought to determine which physical characteristics are of importance for the generated sound. In particular, Holz (1996) concluded that an "ideal" xylophone wood bar is characterized by a specific value range of density, Young modulus, and damping factors. Ono and Norimoto (1983) demonstrated that samples of spruce wood (*Picea excelsa, P. glehnii, P. sitchensis*)—considered a suitable material for soundboards—all had a high sound velocity and low longitudinal damping coefficient as compared to other softwoods. The cell-wall structure may account for this phenomenon. Internal friction and the longitudinal modulus of elasticity are markedly affected by the microfibril angle in the S2 tracheid cell layer, but this general trend does not apply to all species. For instance, pernambuco (*Guilandina echinata* Spreng.), traditionally used for making violin bows, has an exceptionally low damping coefficient relative to other hardwoods and softwoods with the same specific modulus (Bucur, 1995; Matsunaga *et al.*, 1996; Sugiyama *et al.*, 1994). This feature has been explained by the abundance of extractives in this species (Matsunaga and Minato, 1998). Obataya *et al.*(1999) confirmed the importance of extractives for the rigidity and damping qualities of reed materials. Matsunaga *et al.* (1999) reduced the damping coefficient of spruce wood by impregnating samples with extractives of pernambuco (*Guilandina echinata* Spreng.). The high sound quality conditions are met by the wood species commonly used by xylophone makers (like *Dalbergia* sp.), but other tropical woods may serve. We propose to focus on the perceptual properties of impacted wood bars as the basis for pointing out woods suitable for xylophone manufacturing. Several studies using natural or synthetic sounds have been conducted to point out auditory clues associated with geom-

[a)]Author to whom correspondence should be addressed. Electronic mail: aramaki@lma.cnrs-mrs.fr

etry and material properties of vibrating objects (Avanzini and Rocchesso, 2001; Giordano and McAdams, 2006; Lutfi and Oh, 1997; Klatzky *et al.*, 2000; McAdams *et al.*, 2004). These studies revealed the existence of perceptual clues allowing the source of the impact sound to be identified merely by listening. In particular, the perception of material correlated mainly with the internal friction (related to the damping factors of the spectral components) as theoretically shown by Wildes and Richards (1988). Nevertheless, it has not been determined whether the perceptual clues highlighted in the distinction of different materials are those used to establish the subjective classification of different species of wood.

The perceptual differences reported in the literature are linked with subtle changes in timbre, defined as "the perceptual attribute that distinguishes two tones of equal, pitch, loudness, and duration" (ANSI, 1973). This definition points out the importance of comparing sounds with similar loudness, duration, and pitch. Concerning loudness and duration, the sounds of interest can easily be adjusted in intensity by listening, and they have about the same duration since they correspond to the very narrow category of impacted wooden bars. Concerning pitch, the bars do not have the same values because the pitch depends on the physical characteristics of the wood, i.e., essentially of the Young modulus and the mass density. To tune the sounds to the same pitch, we propose to digitally process the sounds recorded on bars of equal length. Synthesis models can be used for this purpose, allowing virtual tuning by altering the synthesis parameters. Such an approach combining sound synthesis and perceptual analysis has already been proposed. Most of the proposed models are based on the physics of vibrating structures, leading to a modal approach of the synthesis process (Adrien, 1991; Avanzini and Rocchesso, 2001) or to a numerical method of computation (Bork, 1995; Chaigne and Doutaut, 1997; Doutaut *et al.*, 1998). Yet, although these models lead to realistic sounds, they do not easily allow for an analysis-synthesis process implicating the generation of a synthetic sound perceptually similar to an original one. To overcome this drawback, we propose an additive synthesis model based on the physics of vibrating bars, the parameters of which can be estimated from the analysis of natural sounds.

The paper is organized as follows: in Sec. II, we discuss the design of an experimental sound data bank obtained by striking 59 wooden bars made of different woods carefully selected and stabilized in a climatic chamber. In Sec. III, we then address the issue of digitally tuning the sounds without changing the intrinsic characteristics of the wood species. This sound manipulation provided a tuned sound data bank in which each sound was associated with a set of descriptors estimated from both physical experiments and signal analysis. The experimental protocol is described in Sec. IV. It consists of the classification carried by a professional instrument maker. The classification was performed with both the original and the tuned data banks to better understand the influence of pitch on the classification. These results are discussed in Sec. VII, leading to preliminary conclusions that agree with most of the knowledge and usage in both wood mechanics, xylophone manufacturing, and sound perception.

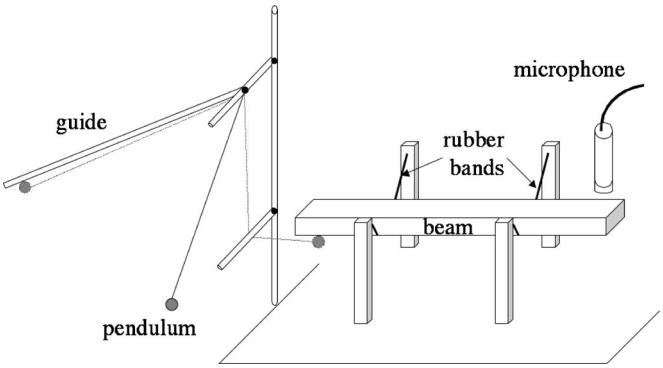


FIG. 1. Experimental setup used to strike the wood samples and record the impact sounds. The setup was placed in an anechoic room.

II. DESIGN OF AN EXPERIMENTAL SOUND DATA BANK

a. Choice of wood species. Most percussive instruments based on wooden bars are made of specific species (for example, *Dalbergia* sp. or *Pterocarpus* sp.). In this experiment, we used tropical and subtropical species, most of which were unknown to instrument makers. A set of 59 species presenting a large variety of densities (from 206 to 1277 kg/m^3) were chosen from the huge collection (about 8000) of the CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement, Montpellier, France). Their anatomical and physical characteristics have been intensely studied and are well described. The name and density of each species are in Table III.

b. Manufacturing wooden bars. Both geometry and boundary conditions govern the vibration of bars. By considering bars with the same geometry and boundary conditions, sounds can be compared to determine the intrinsic quality of the species. Hence, a set of bars was made according to the instrument maker recommendations. The bars were manufactured to be as prismatic as possible, with dimensions L $=350$ mm $\times$ W $=45$ mm $\times$ T $=20$ mm, without singularities and cut in the grain direction. We assume that the growth rings are parallel to the tangential wood direction and that their curvature is negligible. The longitudinal direction is collinear to the longitudinal axis of the bars. The bars were stabilized in controlled conditions.

c. Recording of impact sounds under anechoic conditions. An experimental setup was designed that combines an easy way to generate sounds with a relative precision ensuring the repeatability of the measurements, as shown in Fig. 1. In this way, impact excitation was similar for all the impacted bars. Moreover, to minimize the sound perturbations due to the environment, the measurements took place in an anechoic room.

The bar was placed on two rubber bands, ensuring free-free-type boundary conditions. The rubbers minimized perturbations due to suspension (see, for example, Blay *et al.*, 1971 for more details). Bars were struck with a small steel pendulum. The ball on the string was released from a constrained initial position (guide), and after the string wrapped around a fixed rod, the ball struck the bar from underneath. The robustness of this simple procedure showed the radiated

sounds were reproducible: the determination error was less than 0.1% for the fundamental frequency and 4.3% for the damping coefficient of the first mode (Brancheriau *et al.*, 2006a). To ensure broad spectral excitation, the ball was chosen to generate a sufficiently short pendulum/bar contact (to be as close as possible to an ideal Dirac source). The excitation spectrum is given by the Fourier transform of the impact force, so that the shorter the impact, the broader the spectrum excitation. For that, a steel ball was used since the modulus of elasticity of steel is much larger than that of wood (the ratio is about 200). This setup makes contact duration between the ball and the bar short (Graff, 1975). This duration was shortened because the impact point was underneath the bar, maximizing the reversion force. After several experiments, a good compromise between speed, short duration, and lack of deformation of the material was obtained with a steel ball of 12 g and a 14 mm diameter, tightened by a 30-cm-long string. The impact point played an important role in the generation of sounds. To prevent the first modes from vanishing, the bar was struck close to one of its extremities (at 1 cm), allowing high frequency modes to develop. An omni-directional microphone (Neumann KM183mt) was placed in the close sound field at the opposite end of the impact location to measure the sound-radiated pressure. This configuration obviates the contribution of the spectral peak generated by the ball, peak which was at about 10 kHz. The sounds were digitally recorded at 48 kHz sampling frequency.

d. Signal characteristics. Figure 2 shows the temporal signal, the spectral representation, and the time-frequency representation of a typical sound obtained experimentally. The temporal signals are characterized by a short onset and a fast decay. Consequently, their durations generally do not exceed 1 s. Their spectra are composed of emergent resonances that do not overlap much. As shown by the time-frequency representation, the damping of these spectral components is frequency dependent, the high frequency components being more heavily damped than the low frequency ones.

III. DESIGN OF TUNED SOUND DATA BANK FOR TIMBRE STUDY

To facilitate comparison of the timbre of sounds generated striking different wood species, their pitch was equalized. In practice, this could have been possible using the same procedure adopted by percussive instrument makers, where the bar geometry is modified removing some substance around the center of the bar to be tuned (Fletcher and Rossing, 1998). This approach comes, however, with the risk of making irreversible mistakes, for example, removing an excessive amount of wood. As an alternative, we propose to digitally tune the pitch of sounds generated striking bars of equal length. Such an approach relies on the plausible assumption that the pitch of our recorded signals is primarily determined by the frequency of the first vibrational mode. In particular, we use a sound synthesis model which allows for sound transformations that are accurate relative to the physical phenomena, as compared to other signal processing approaches such as pitch shifting.

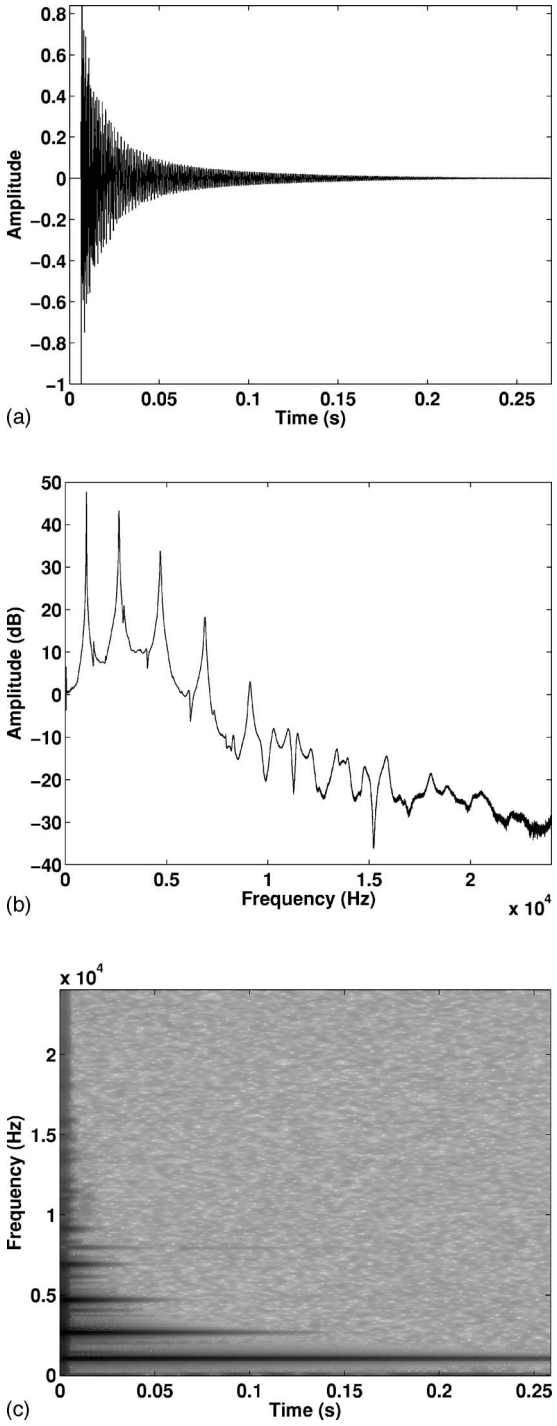


FIG. 2. (a) Wave form, (b) spectral representation, and (c) spectrogram (amplitude in logarithmic scale) of a typical sound obtained by impacting a wooden bar.

A. Synthesis model based on physical approach

To tune the sounds, we propose to use an additive synthesis model. This model simulates the main characteristics of the vibrations produced by an impacted bar to exhibit the principal properties of the radiated sound.

1. Simplified mechanical model

Numerous mechanical models of bar vibrations are available in the literature, but the relevant information can be

pointed out using a simple model based on assumptions that are coherent with our experimental design. According to the manufacturing of the bars, one can assume that the fiber orientation follows the axis of the bar and that the ratio length/width is large. Consequently, one can neglect the anisotropy property of the wood and the contribution of the longitudinal and torsional modes (which are few, weak, and of little influence on the radiated sound). These assumptions allow for the consideration of a one-dimensional mechanical model depending only on the longitudinal Young modulus. Such a model can be described by the well-known Euler-Bernoulli equation

$$EI\frac{\partial^4 y(x,t)}{\partial x^4} + \rho S\frac{\partial^2 y(x,t)}{\partial t^2} = 0, \tag{1}$$

where E is the longitudinal Young modulus, I the quadratic moment, ρ the mass density, and S the cross section area. The general solution of the equation is given by

$$y(x,t) = \sum_n Y_n(x)e^{i\gamma_n t} \tag{2}$$

with

$$Y_n(x) = A\cosh(k_n x) + B\sinh(k_n x) + C\cos(k_n x) + D\sin(k_n x). \tag{3}$$

By injecting Eq. (2) and Eq. (3) into the Eq. (1), one obtains

$$\gamma_n = \pm\sqrt{\frac{EI}{\rho S}}k_n^2. \tag{4}$$

Our experimental setup corresponds to free-free boundary conditions written

$$\frac{\partial^2 Y(0)}{\partial x^2} = \frac{\partial^2 Y(L)}{\partial x^2} = \frac{\partial^3 Y(0)}{\partial x^3} = \frac{\partial^3 Y(L)}{\partial x^3} = 0$$

leading to

$$k_n = (2n+1)\frac{\pi}{2L}. \tag{5}$$

To take into account viscoelastic phenomena, E is considered as complex valued, see, for example (Valette and Cuesta, 1993)

$$E = E_d(1 + i\eta), \tag{6}$$

where E_d is the dynamical Young modulus, and η a dimensionless material loss factor. By injecting relations (5) and (6) into relation (4) and assuming that $\eta \ll 1$, one obtains the following important expressions:

$$\gamma_n = \omega_n + i\alpha_n \tag{7}$$

with

$$\begin{cases} \omega_n \approx \sqrt{\frac{E_d I}{\rho S}}(2n+1)^2\frac{\pi^2}{4L^2} \\ \alpha_n \approx \frac{\eta}{2}\omega_n \end{cases}. \tag{8}$$

Thus, one can rewrite the relation (2):

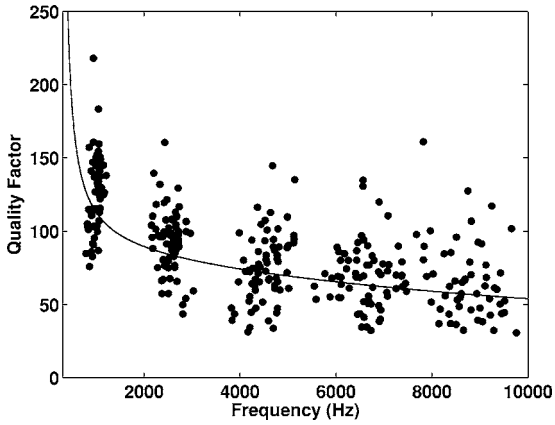


FIG. 3. Quality factor Q_n estimated on original sounds. The Q_n values are fitted, in a least squares sense, by a rational function (black curve) corresponding to Eq. (11).

$$y(x,t) = \sum_n Y_n(x)e^{i\omega_n t}e^{-\alpha_n t}. \tag{9}$$

It is accepted (Chaigne and Doutaut, 1997; McAdams *et al.*, 2004; Ono and Norimato, 1985) that the damping factors in case of wooden bars are described by a parabolic form:

$$\alpha(f) = a_0 + a_2 f^2 \tag{10}$$

where the constants a_0 and a_2 depend on the wood species. This corresponds to a quality factor Q_n given by

$$Q_n = \frac{\pi f_n}{\alpha_n} = \frac{\pi f_n}{a_0 + a_2 f_n^2}. \tag{11}$$

This behavior was experimentally verified, as shown in Fig. 3.

These expressions show that the vibrations of the bar, which are correlated with the radiated sound pressure, can be described by a sum of elementary components consisting of exponentially damped monochromatic signals. The frequency of these elementary components is inversely proportional to the square of the length of the bar, and their damping is proportional to the square of the frequency.

2. Additive synthesis model

The synthesis model aims at simulating the analytical solutions written in Eq. (9), which are expressed as a sum of exponentially damped sinusoids

$$s(x,t) = \theta(t)\sum_{n=1}^{N} A_n(x)\sin(\omega_n t)e^{-\alpha_n t}, \tag{12}$$

where N is the number of components, $\theta(t)$ the Heaviside function, A_n the amplitude, ω_n the frequency and α_n the damping coefficient of the nth component. The choice of either sine or cosine functions has no perceptual influence on the generated sounds but sine functions are often used in sound synthesis since they avoid discontinuities in the signal at $t=0$. Hence, the signal measured at a fixed location is considered to be well represented by the expression (12). Its spectral representation is given by

$$S(\omega) = \sum_{n=1}^{N} \frac{A_n}{2i}\left(\frac{1}{\alpha_n + i(\omega - \omega_n)} - \frac{1}{\alpha_n + i(\omega + \omega_n)}\right)$$

and the z transform by

$$S(z) = \sum_{n=1}^{N} \frac{A_n}{2i}\left(\frac{1}{1 - e^{(i\omega_n - \alpha_n)}z^{-1}} - \frac{1}{1 - e^{(-i\omega_n - \alpha_n)}z^{-1}}\right).$$

B. Estimation of synthesis parameters

Before the tuning process, the recorded sounds described in Sec. II are equalized in loudness, analyzed, and then resynthesized with the synthesis model described above. The loudness was equalized by listening tests. For that, the synthesis parameters are directly estimated from the analysis of the recorded sounds. The estimation of the parameters defining the sound is obtained by fitting the recorded signal with the expression given in relation (12). To do so, we used a signal processing approach that consists of identifying the parameters of a linear filter by auto regressive and moving average (ARMA) analysis. We model the original signal as the output of a generic linear filter whose z transform is written

$$H(z) = \frac{\sum_{m=0}^{M} a_m z^{-m}}{1 + \sum_{n=1}^{N} b_n z^{-n}} = a_0 z^{N-M} \frac{\prod_{m=1}^{M} (z - z_{0m})}{\prod_{n=1}^{N} (z - z_{pn})},$$

where z_{0m} are the zeros and z_{pn} are the poles of the system. Only the most prominent spectral components were modeled by $H(z)$. These spectral components were determined within a 50 dB amplitude dynamic, the reference being the amplitude of the most prominent spectral peak. Hence, the number of poles N and zeros M of the linear ARMA filter is determined by the number of spectral components taken into account. The coefficients a_m and b_n are estimated using classical techniques such as Steiglitz-McBride (Steiglitz and McBride, 1965). The synthesis parameters corresponding to the amplitudes, frequencies, and damping coefficients of the spectral components are thus determined:

$$\begin{cases} A_n = |H(z_{pn})|, \\ \omega_n = \arg(z_{pn}) f_s, \\ \alpha_n = \log(|z_{pn}|) f_s, \end{cases} \tag{13}$$

where f_s is the sampling frequency. In addition to the synthesis model described above, we have taken into account the attack time. Actually, even though the rising time of the sounds is very short, it does influence the perception of the sounds. These rising times were estimated on the original sounds and were reproduced by multiplying the beginning of the synthetic signal by an adequate linear function. Synthesis sounds were evaluated by informal listening tests confirming that their original sound qualities were preserved. The synthesis quality was further confirmed by results from the professional instrument maker showing a similar classification of original and synthetic sounds (classifications Cl and C2, see Sec. VI A 1).

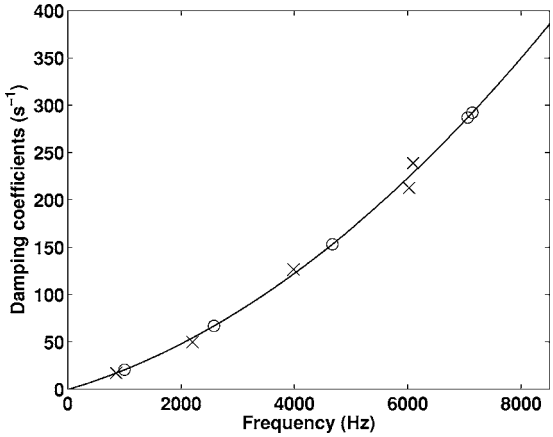


FIG. 4. The damping coefficients of the original sound (×) are fitted by a parabolic function (solid curve). The damping coefficients of the tuned sound (◦) are determined according to this damping law.

C. Tuning the sounds

The processing of tuning the sounds at the same pitch was based on some assumptions specific to the investigated stimulus set and consistent with the vibratory behavior of the bar. For the kind of sounds we are dealing with (impacted wooden bars), we assume the pitch to be related to the frequency of the first vibration mode, which is correlated with the length of the bar [cf. Eq. (8)]. Actually, if the length L changes to βL, then ω_n changes to ω_n/β^2. As a consequence, a change in pitch corresponds to a dilation of the frequency components. These assumptions made it possible to virtually equalize the pitches of the recorded bank of sounds. To minimize the pitch deviation, the whole set of sounds was tuned by transposing the fundamental frequencies to 1002 Hz, which is the mean fundamental frequency of all the sounds. The amplitude of the spectral components was kept unchanged by the tuning process. Once again, no precise listening test was performed, but our colleagues found the synthesis sounds preserved the specificity of the material.

According to the discussion in III A 1, the damping is proportional to the square of the frequency. Thus, from the expression (10), a damping law can be defined by a parabolic function that can be written in a general form:

$$\alpha(\omega) = D_A \omega^2 + D_B \omega + D_C. \tag{14}$$

As a consequence, when the pitch is changed, the damping coefficient of each tuned frequency component has to be evaluated according to the damping law measured on the original sound (cf. Fig. 4).

Figure 5 shows the comparison between the spectrum of a measured signal and the spectrum of a tuned signal after the resynthesis process. The entire sound data bank is available at http://www.lma.cnrs-mrs.fr/~kronland/JASA_Xylophone/sounds.html.

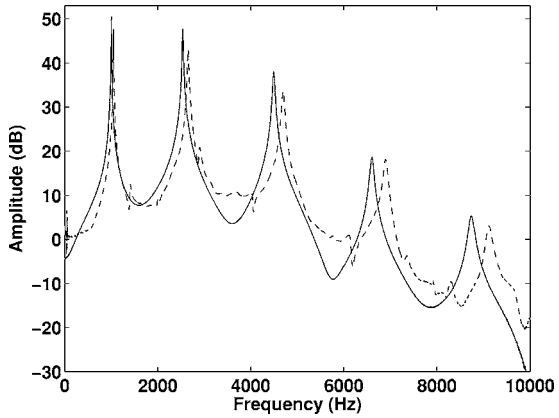


FIG. 5. Comparison between a spectrum of a measured signal (dashed trace) and the spectrum of the associated tuned signal (solid trace).

IV. EXPERIMENTAL PROTOCOL

Sounds from different wooden bars were evaluated focusing on the perceived musical quality of the wood samples. The participant was placed in front of a computer screen on which the sounds (all represented as identical crosses) were randomly distributed. The participant was asked to place the sounds on a bidimensional computer display. In particular, he was told that the horizontal dimension of the display represented an axis of musical quality so that sounds judged as having the worst/best quality were to be placed on the leftmost/rightmost part of the display. The participant could listen to the sounds as often as he wanted by simply clicking on the cross. The tests were carried on a laptop Macintosh equipped with a Sony MDR CD550 headset.

For this study, one instrument maker specialized in xylophone manufacture carried the task. For a complete perceptual study, more participants would, of course, be needed. As a first step we aimed at presenting a new methodology for an interdisciplinary approach uniting instrument makers and specialists within acoustics, signal processing, and wood sciences.

Three tests were conducted using this experimental protocol. The instrument maker carried the classification with the original sounds (recorded sounds with different pitches), called C1 (Brancheriau *et al.*, 2006a; Brancheriau *et al.*, 2006b). A second classification, called C2, using the synthesized sounds (resynthesis of the original sounds with different pitches) was done two years later. The comparison of C1 and C2 allowed us to check the quality of the resynthesis as well as the reliability of our experimental participant. The third test (C3) was carried on the signals tuned to the same pitch. The xylophone maker was not aware of the synthetic nature of sounds in C2 and C3. In particular, he was told that, for classification C3, the same pieces of wood had been sculpted in order to tune the sounds to the same fundamental frequency. Classification C3 is presented in Table III.

V. DESCRIPTORS

A. Mechanical descriptors

The wood species used for this study have been intensively examined at CIRAD and their anatomical and physical characteristics are well known. Thus, the mechanical descriptors are defined by the mass density, ρ, the longitudinal modulus of elasticity, E_ℓ, and the transverse shear modulus, G_t. The descriptors E_ℓ and G_t can be calculated using Timoshenko's model and the Bordonné solutions (Brancheriau and Baillères, 2002). We have also considered the specific longitudinal modulus, E_ℓ/ρ, and the specific shear modulus, G_t/ρ.

B. Signal descriptors

To characterize the sounds from an acoustical point of view, we calculated the following timbre descriptors (Caclin *et al.*, 2005; McAdams *et al.*, 1995): attack time, AT (the way the energy rises during the onset of the sound), spectral bandwidth, SB (spectrum spread), spectral centroid, SCG (brightness), and spectral flux, SF (the way the sound vanishes).

The attack time, AT, a temporal descriptor, characterizes the signal onset and describes the time it takes for the signal to reach its maximum. It is generally estimated as the time it takes the signal to deploy its energy from 10% to 90% of the maximum. The spectral timbre descriptors characterize the organization of the spectral peaks resulting from the modal behavior of the bar vibration. One of the most well known is the spectral centroid, SCG, which is correlated with the subjective sensation of brightness (Beauchamps, 1982):

$$\mathrm{SCG} = \frac{\sum_k f(k)|\hat{s}(k)|}{\sum_k |\hat{s}(k)|}, \tag{15}$$

where $\hat{s}$ is the discrete Fourier transform of the signal $s(t)$ and f the frequency. The spectral bandwidth, SB, measures the spread of the spectral components around the spectral centroid and is defined as (Marozeau, de Cheveigné, McAdams and Winsberg, 2003)

$$\mathrm{SB} = \sqrt{\frac{\sum_k |\hat{s}(k)|(f(k) - \mathrm{SCG})^2}{\sum_k |\hat{s}(k)|}}. \tag{16}$$

Finally, the fourth classical timbre descriptor called the spectral flux, SF, is a spectro-temporal descriptor that measures the deformation of the spectrum with respect to time. In practice, the spectral flux is given by a mean value of the Pearson correlation calculated using the modulus of local spectral representations of the signal (McAdams *et al.*, 1995):

$$\mathrm{SF} = \frac{1}{N}\sum_{n=1}^{N} \frac{\langle s_n, s_{n-1}\rangle}{s_n^2 s_{n-1}^2}, \tag{17}$$

where N represents the number of frames, s_n the modulus of the local spectrum at the discrete time n, and $\langle,\rangle$ the discrete scalar product.

In addition to these well-known timbre descriptors, we propose to consider various acoustical parameters chosen as function of the specificities of the impact sounds, i.e., the

amplitude ratio between the first two frequency components of the sound, noted $A_{2/1}$, and the damping and the inharmonicity descriptors. The last two parameters are described below in more detail. The damping descriptor is defined from the Eq. (14) by the set of coefficients $\{D_A, D_B, D_C\}$ traducing the sound decrease. As the damping is the only parameter responsible for the variation of the spectral representation of the signal with respect to time, this descriptor is related to the spectral flux, SF. In addition, the damping coefficients α_1 and α_2 of components 1 and 2 have been included in the list of signal descriptors. The inharmonicity characterizes the relationship between the partials and the fundamental mode. This parameter is linked with the consonance, which is an important clue in the perceptual differentiation of sounds. For each spectral component, inharmonicity is defined by

$$I(n) = \frac{\omega_n}{\omega_0} - n. \tag{18}$$

From this expression, we propose an inharmonicity descriptor defined by a set of coefficients $\{I_A, I_B, I_C\}$ obtained by fitting $I(n)$ with a parabolic function, as suggested by the calculation $I(n)$ from Eq. (8):

$$I(n) = I_A n^2 + I_B n + I_C. \tag{19}$$

VI. RESULTS

Collected behavioral data could be considered as ordinal. Nevertheless, since the task consisted in placing the sounds on a quality axis "as a function of its musical quality," the relative position of the sounds integrates a notion of perceptual distance. Moreover, the classifications do not contain two sounds with the same position and do not show categories (see Table III). It was thus decided to consider the data as providing a quantitative estimate of perceived musical quality for the wood samples, the value associated with each species being given by its abscissa from 0 (worst quality) to 10 (best quality) on the quality axis. The main interest in using quantitative scales is the possibility of constructing an arithmetic model for perceived wood quality which can be easily used to estimate the musical quality of woods (and sounds) not considered in our experiments. All the statistical analyses were conducted with SPSS software (Release 11.0.0, LEAD Technologies).

A. Qualitative analysis—Choice of the variables

1. Resynthesis quality—Robustness of the classification

Only one participant performed the classifications on the basis of his professional skill, and his judgment of sound quality was used to build reference quality scales. The xylophone maker is thus considered as a "sensor" for measuring the acoustical wood quality. The raw classifications C1 and C2 were compared using the Wilcoxon signed rank test to evaluate the resynthesis quality of the model. Moreover, this comparison allowed us to evaluate the robustness of the xylophone maker classification. No particular distribution was assumed for the classifications. The Wilcoxon test is thus appropriate for comparing the distributions of the two classifications (C1, C2). The significance value of the Wilcoxon test ($p=0.624$) for (C1, C2) indicates that classification C1 equals classification C2. There was no significant difference in the xylophone maker responses between C1 and C2.

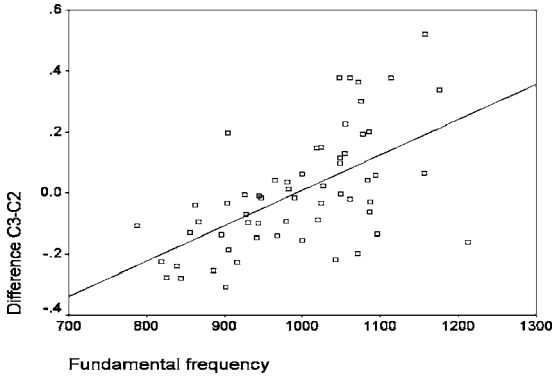


FIG. 6. Linear relationship between fundamental frequency and arithmetic difference C3-C2 ($R=0.59$, $N=59$).

2. Influence of the tuning process

The same Wilcoxon signed rank test was performed with classification C2 and classification C3 of tuned sounds. The hypothesis of equal distribution is rejected considering classifications C2 and C3. A significant difference between C2 and C3 ($p=0.001$) is due to the tuning process of sounds, which altered the sound perception of the xylophone maker. The arithmetic difference (C3-C2) was thus computed and related to the value of the fundamental frequency by using the Pearson correlation coefficient (Fig. 6). This coefficient value was found significant at the 1% level ($R=0.59$).

B. Quantitative analysis

1. Descriptor analysis

The 18 parameters presented in Table I were estimated for the tuned sounds and using standard mechanical calibrations. They are grouped into mechanical/physical descriptors and signal descriptors. In practice, for the spectral descriptors, the Fourier transform was estimated using a fast Fourier transform (FFT) algorithm. The length of the FFT was chosen so that it matches the longest sound, i.e., 2^{16} samples. For the SF calculation, the number of samples was 256 with an overlap of 156 samples. A Hamming window was used to minimize the ripples. Mechanical descriptors are linked with the intrinsic behavior of each sample but also linked with signal descriptors, as shown in Fig. 7. Indeed, the bivariate coefficients of determination matrix calculated on the basis of the 18 characteristic parameters revealed close collinearity between the parameters. Considering the strong relationship between the parameters, the statistical analyses were conducted by grouping the mechanical/physical descriptors and the signal descriptors in order to find those that best explain the classification C3.

A principal component analysis was thus conducted (Table II). Principal components analysis finds combinations of variables (components) that describe major trends in the data. This analysis generated a new set of parameters derived from the original set in which the new parameters (principal

TABLE I. Mechanical and signal descriptors computed from dynamic tests.

	No.	Variable	Signification
Mechanical descriptors	1	ρ	Mass density (kg/m^3)
	2	E_ℓ	Longitud. modulus of elasticity (MPa)
	3	G_t	Shear modulus (MPa)
	4	E_ℓ/ρ	Specific longitudinal modulus
	5	G_t/ρ	Specific shear modulus
Signal descriptors	6	$A_{2/1}$	Amplitude ratio of mode 2 and 1
	7	α_1	Temporal damping of mode 1 (s^{-1})
	8	α_2	Temporal damping of mode 2 (s^{-1})
	9	SCG	Spectral centroid (Hz)
	10	SB	Spectral bandwidth (Hz)
	11	SF	Spectral flux
	12	AT	Attack time (ms)
	13	D_A	Coefficient D_A of $\alpha(\omega)$
	14	D_B	Coefficient D_B of $\alpha(\omega)$
	15	D_C	Coefficient D_C of $\alpha(\omega)$
	16	I_A	Coefficient I_A of $I(n)$
	17	I_B	Coefficient I_B of $I(n)$
	18	I_C	Coefficient I_C of $I(n)$

components) were not correlated and closely represented the variability of the original set. Each original parameter was previously adjusted to zero mean and unit variance so that eigenvalues could be considered in choosing the main factors. In this case, the eigenvalues sum the number of variables, and eigenvalues can be interpreted as the number of original variables represented by each factor. The principal components selected thus corresponded to those of eigenvalue superior or equal to unity. Table II shows that six principal components accounted for 87% of all information contained in the 18 original parameters.

The relationships between original variables and principal components are presented in Figs. 8(a) and 8(b). These figures display the bivariate coefficient of determination between each principal component and each original parameter; the bivariate coefficient corresponds to the square loading coefficient in this analysis. The variance of the inharmonicity coefficients $\{I_A, I_B, I_C\}$ and the damping coefficients $\{D_A, D_B, D_C\}$ are captured by the first principal component and to a lesser degree by the third component [Fig. 8(a)]. The damping coefficients (α_1 and α_2), however, are mainly linked with the second component. This component is also linked with the amplitude ratio $A_{2/1}$ and with the timbre descriptors (SCG, SB, SF, AT). The variance of the mechanical/physical descriptors is scattered between all the principal components (parameter 1 is linked with PC1 and 2; parameter 2 with PC1 and 4; parameter 3 with PC3 and 5; parameter 4 with PC2, 3, and 4; and parameter 5 with PC3 and 5).

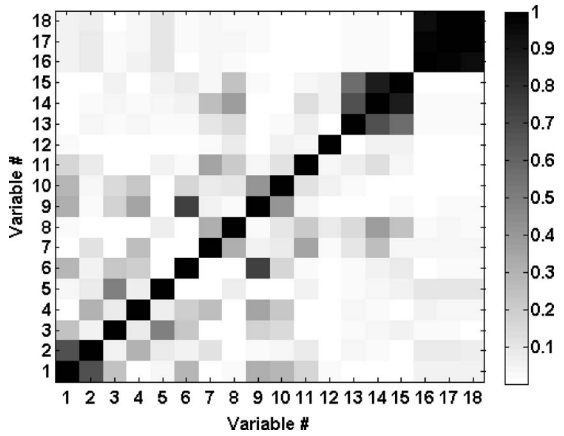


FIG. 7. Bivariate coefficients of determination for characteristic parameters (N=59).

2. Relationship between the descriptors and the acoustic classification of tuned sounds

a. Bivariate analysis. Figure 9 presents the results of bivariate analysis between characteristic parameters and classification C3. Assuming a linear relationship, the parameter α_1 (temporal damping of mode 1) appeared to be the best individual predictor with a R^2 value of 0.72. The second most significant predictor was the spectral flux, SF, with a R^2 value of 0.38. The other parameters were of minor importance considering classification C3. Note that the only mechanical parameter of interest was E_ℓ/ρ (specific longitudinal modulus) with a relatively low R^2 value of 0.25. Furthermore, the mass density, ρ, was not reflected in the acoustic classification (no significant R^2 value at the 1% level). Light woods and heavy woods were thus not differ-

TABLE II. Variance explained by the principal components (number of initial variables=18, number of samples=59).

Component	Eigen val.	% of Var.	Cumul. (%)
I	4.0	22.5	22.5
II	3.9	21.9	44.3
III	3.5	19.3	63.7
IV	1.8	10.1	73.8
V	1.2	6.7	80.5
VI	1.1	6.1	86.6

TABLE III. Botanical names of wood species (N=59), their density (kg/m^3), α_1 the temporal damping of mode 1(s^{-1}), SB the spectral bandwidth (Hz) and classification C3 by the xylophone maker (normalized scale from 0 to 10).

Botanical name	Density (kg/m^3)	α_1(s^{-1})	SB (Hz)	C3
Pericopsis elata Van Meeuw	680	21.76	2240	5.88
Scottellia klaineana Pierre	629	23.97	2659	6.38
Ongokea gore Pierre	842	26.07	2240	5.15
Humbertia madagascariensis Lamk.	1234	28.84	3820	0.48
Ocotea rubra Mez	623	23.47	2521	5.42
Khaya grandifoliola C.DC.	646	33.02	2968	0.95
Khaya senegalensis A. Juss.	792	33.98	3101	0.33
Coula edulis Baill.	1048	27.6	2674	2.1
Tarrietia javanica Bl.	780	20.33	2198	9.15
Entandrophragma cylindricum Sprague	734	30.6	2592	1.12
Afzelia pachyloba Harms	742	20.56	2048	8.24
Swietenia macrophylla King	571	20.99	1991	9.22
Aucoumea klaineana Pierre	399	32.17	2275	1.81
Humbertia madagascariensis Lamk	1277	23.36	3171	3.48
Faucherea thouvenotii H. Lec.	1061	20.18	2512	6.05
Ceiba pentandra Gaertn.	299	29.16	2396	2.57
Letestua durissima H. Lec.	1046	19.56	2770	3.87
Monopetalanthus heitzii Pellegr.	466	23.98	2344	5.57
Commiphora sp.	390	16.52	1269	9.77
Dalbergia sp.	916	14.29	2224	9.79
Hymenolobium sp.	600	20.58	2402	7.86
Pseudopiptadenia suaveolens Brenan	875	20.8	1989	6.53
Parkia nitida Miq.	232	26.86	1440	5.75
Bagassa guianensis Aubl.	1076	20.68	2059	6.82
Discoglypremna caloneura Prain	406	34.27	1506	1.38
Brachylaena ramiflora Humbert	866	21.85	2258	4.71
Simarouba amara Aubl.	455	21.26	1654	9.37
Gossweilerodendron balsamiferum Harms	460	35.26	1712	1.08
Manilkara mabokeensis Aubrev.	944	23.89	1788	3.25
Shorea-rubro squamata Dyer	569	23.9	1604	6.75
Autranella congolensis A. Chev.	956	38.97	3380	0.35
Entandrophragma angolense C. DC.	473	22.79	1612	7.67
Distemonanthus benthamianus Baill.	779	19.77	2088	8.75
Terminalia superba Engl. & Diels	583	21.89	2004	9.32
Nesogordonia papaverifera R.Cap.	768	27.96	2097	2.37
Albizia ferruginea Benth.	646	24.71	2221	4.32
Gymnostemon zaizou. Aubrev. & Pellegr.	380	30.15	2130	1.83
Anthonotha fragrans Exell & Hillcoat	777	24.87	1926	4.2
Piptadeniastrum africanum Brenan	975	22.41	3226	3.68
Guibourtia ehie J. Leon.	783	26.36	2156	4.05
Manilkara huberi Standl.	1096	35.11	2692	0.77
Pometia pinnata Forst.	713	25.5	1835	6.23
Glycydendron amazonicum Ducke	627	20.41	2292	7.91
Cunonia austrocaledonica Brong. Gris.	621	31.05	3930	0.59
Nothofagus aequilateralis Steen.	1100	37.76	3028	0.18
Schefflera gabriellae Baill.	570	28.16	1872	1.42
Gymnostoma nodiflorum Johnst.	1189	33	3013	1.26
Dysoxylum sp.	977	23.85	2106	4.49
Calophyllum caledonicum Vieill.	789	19.82	2312	8.66
Gyrocarpus americanus Jacq.	206	38.39	1982	0.6
Pyriluma sphaerocarpum Aubrev.	793	30.83	2318	1.23
Cedrela odorata L.	512	30.45	2070	3
Moronobea coccinea Aubl.	953	21.67	1781	4.92
Goupia glabra Aubl.	885	45.61	2525	0.22
Manilkara huberi Standl.	1187	22.6	2917	2.78
Micropholis venulosa Pierre	665	22.51	3113	7.12
Cedrelinga catenaeformis Ducke	490	22.5	1626	7.31
Vouacapoua americana Aubl.	882	23.18	1986	6.88
Tarrietia Densiflora Aubrev & Normand	603	29.76	2326	1.62

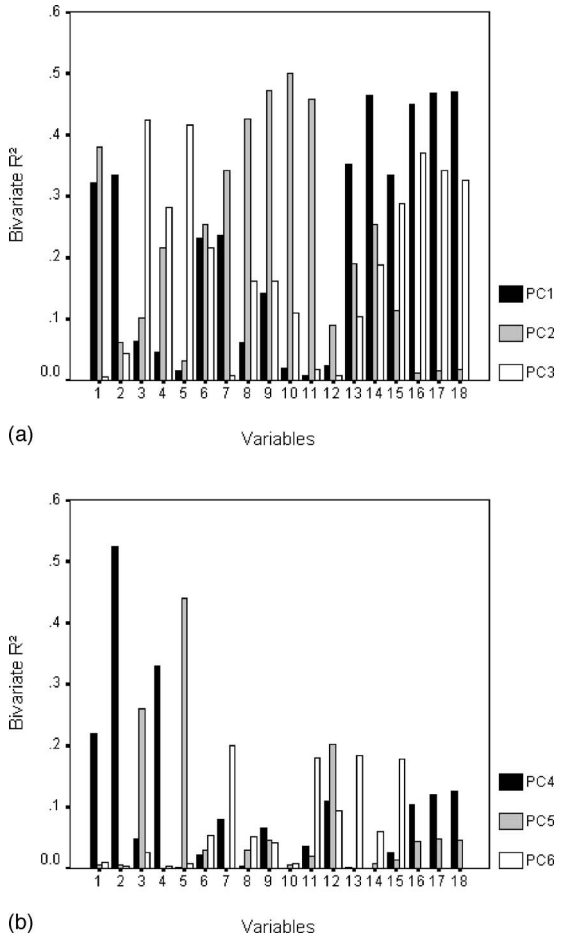


FIG. 8. Bivariate determination coefficient between original variables and principal components: (a) for PC1, PC2 and PC3; (b) for PC4, PC5 and PC6.

entiated by the xylophone maker in the acoustic classification.

b. Multivariate linear regression analysis. The second step of the analysis was to build a robust linear model to take into account the most significant predictors. The robustness of the model assumes that no multicollinearity among the variables exists (Dillon and Goldstein, 1984). The stepwise selection method was thus used to perform multivariate

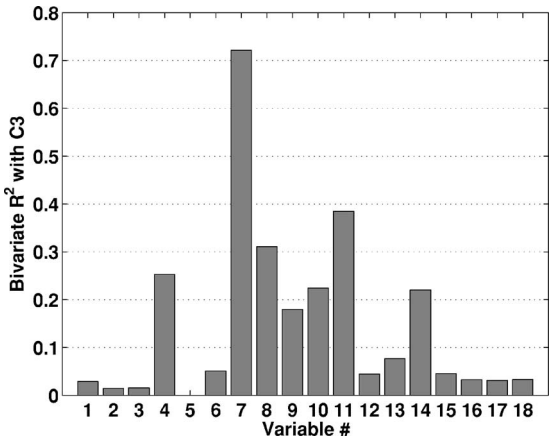


FIG. 9. Bivariate coefficients of determination between characteristic parameters and classification C3 (N=59).

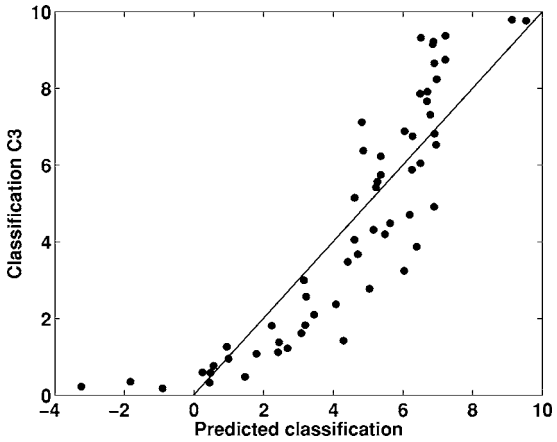


FIG. 10. Predicted vs observed C3 classification (linear predictors α_1 and SB, R^2=0.77, N=59).

analysis. This method enters variables into the model one by one and tests all the variables in the model for removal at each step. Stepwise selection is designed for the case of correlations among the variables. Other automatic selection procedures exist (forward selection and backward elimination, for example), and the models obtained by these methods may differ, especially when independent variables are highly intercorrelated. Because of the high correlation between variables, several regression models almost equally explain classification C3. However, stepwise selection was used to build one of the most significant models with noncorrelated variables relating to different physical phenomena.

The final linear model obtained by stepwise variable selection included the two predictors, α_1 and SB. The predicted classification is given by:

$$\hat{C}3_{\text{Linear}} = -3.82 \times 10^{-1}\alpha_1 - 1.32 \times 10^{-3}SB + 17.52. \tag{20}$$

The multiple coefficient of determination was highly significant (R^2=0.776 and Adjusted R^2=0.768, Fig. 10) and each regression coefficient was statistically different from zero (significance level: 1%). The predictor α_1 was predominant in the model with a partial coefficient value of $R_{\alpha 1}$=−0.84 (R_{SB}=−0.44). The negative sign of $R_{\alpha 1}$ showed that samples with high damping coefficients were associated with a poor acoustic quality.

Partial least squares regression showed that the damping coefficient α_1 was predominant in the model (Brancheriau *et al.*, 2006b). However, the physical significance of the partial least squares model was difficult to explain because the original variables were grouped in latent variables. The stepwise procedure was thus used to better understand the regression results.

The multivariate analysis differed from the bivariate analysis by the replacement of SF by SB, because the selected set of predictors was formed by noncorrelated variables. SB was thus selected because of the low correlation between α_1 and SB with a coefficient value of $R_{\alpha 1/\text{SB}}$=0.29 instead of SF with a value of $R_{\alpha 1/\text{SF}}$=−0.60.

Principal components regression (PCR) was another way to deal with the problem of strong correlations among the variables. Instead of modeling the classification with the variables, the classification was modeled on the principal component scores of the measured variables (which are orthogonal and therefore not correlated). The PCR final model was highly significant with a multiple R^2 value of 0.741 and Adjusted R^2 value of 0.721. Four principal components were selected and the resulting scatter plot was similar to the one in Fig. 10. Comparing the two multivariate models, we found the PCR model to be less relevant than the stepwise one. The R^2 of the PCR model was indeed lower than the R^2 of the stepwise model. Furthermore, the PCR model included four components while only two independent variables were included in the stepwise model. The difference between these two models was explained by the fact that the whole information contained in the characteristic parameters (Table I) was not needed to explain the perceptual classification. The PCR procedure found components that capture the greatest amount of variance in the predictor variables, but did not build components that both capture variance and achieve correlation with the dependent variable.

c. Multivariate nonlinear regression analysis. The configuration of points associated with the linear model (C3, α_1 and SB) in Fig. 10 indicated a nonlinear relationship. This was particularly true for samples of poor acoustic quality (negative values of the standardized predicted classification). As a final step of the analysis, we built a nonlinear model of the behavioral response. In particular, we transformed the values predicted by the linear model $\hat{C}3_{\text{Linear}}$ using a sigmoidal transform. Such transform was consistent with the relationship between C3 and $\hat{C}3_{\text{Linear}}$ (see Fig. 10). The fitting coefficients were extracted via the Levenberg-Marquardt optimization procedure by minimizing the residual sum of squares (dependent variable C3 and independent variable $\hat{C}3_{\text{Linear}}$: predicted classification with the linear modeling). The final equation is written as follows:

$$\hat{C}3_{\text{sigmoid}} = \frac{10}{1 + e^{-\frac{\hat{C}3_{\text{Linear}} - 5}{1.64}}} \tag{21}$$

with $\hat{C}3_{\text{Linear}}$ defined by Eq. (20). The multiple coefficient of determination was highly significant (R^2=0.82) and each nonlinear regression coefficient was statistically different from zero (significance level: 1%). The nonlinear model provided a better fit than the linear model; moreover no apparent systematic feature appeared, indicating that residuals were randomly distributed (Fig. 11).

VII. DISCUSSION

In this section, we discuss the main results presented above, attempting to better understand the sound descriptors' influence on the xylophone maker classification. Further on, we discuss the influence of the pitch and the relationship between the wood anatomy and the produced sounds.

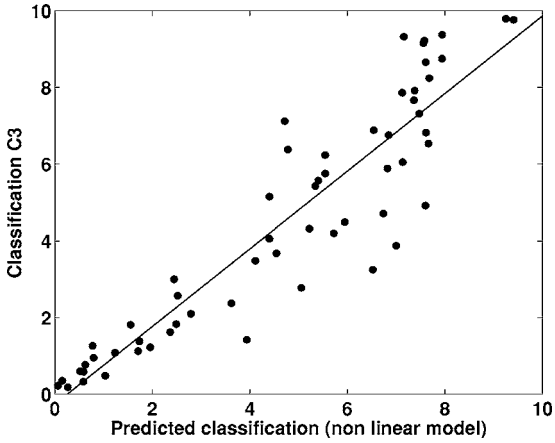


FIG. 11. Predicted vs observed C3 classification (nonlinear predictors α_1 and SB, R^2=0.82, N=59).

A. On the reliability of the xylophone maker

As we pointed out in the introduction, this paper does not aim to give categorical clues for choosing interesting species of wood for xylophone manufacturing. Nevertheless, note that these first conclusions probably accurately reflect what xylophone makers look for. Although we tested our methodology with only one renowned xylophone maker, the results show that:

- In accordance with the xylophone maker community, our maker chose *Dalbergia* sp. as the best species. Moreover, this choice was confirmed on both tuned and original sound classifications.
- The comparison of classifications C1 and C2 showed no significant differences according to the Wilcoxon test.

These observations confirm the good reliability of our xylophone maker and the accuracy of the results, which were further informally confirmed by both instrument makers and musicians.

B. Relation between descriptors and wood classification

The classification by the xylophone maker is correlated with several descriptors. Those that play an important role are three descriptors related to the time course of the sound (α_1, α_2 and SF) and two descriptors related to the spectral content of the sound (SCG and SB). Note that the physical descriptors linked with the wood properties do not explain by themselves the classification of the instrument maker, even though E_ℓ/ρ seems to be the most pertinent one. The relatively low importance of the specific modulus regarding classification C3 could be explained by its high correlation with the fundamental frequency (R^2=0.91) and its low correlation with the temporal damping coefficient α_1 (R^2=0.26). Most of the descriptors are correlated; these correlations are coherent with the physics and are shown in a qualitative way in Fig. 7. Both coefficients of the polynomial decomposition of $\alpha(\omega)$ are strongly correlated. So are the coefficients of the polynomial decomposition of $I(n)$. This finding points out the relative consistency in the behavior of the damping and

the inharmonicity laws with respect to the species. Parameters α_1 and α_2 are also correlated, showing the monotonic behavior of the damping with respect to the frequency: the higher the frequency, the higher the damping. As a consequence, both α_1 and α_2 are correlated with the spectral flux, SF, since these descriptors are the only ones that relate to the time course of the sound.

Both global spectral descriptors, SCG and SB, are also correlated, showing that their increase is strongly related to the adjunction of high frequency energy. These descriptors are in addition correlated with the ratio $A_{2/1}$ and with the physical descriptors ρ and E_ℓ/ρ. This correlation can be explained by the way the energy is distributed through the excited modes. Actually, assuming that the bars are impacted identically (good reproducibility of the impact in the experimental setup), the initial energy injected depends on the impedance of each bar. Since the bars were impacted in the transversal direction, one can assume that the transversal Young modulus of elasticity together with the mass density are the main parameters in the difference of amplitudes of modes 1 and 2.

The multivariate linear regression analysis highlighted two main descriptors: α_1 and SB. These descriptors are noncorrelated and give rise to a linear predictor of the classification $\hat{C}3_{\text{Linear}}$ that explains 77% of the variance. This model is of great importance in the choice of species. Actually, it emphasizes the fact that the xylophone maker looks for a highly resonant sound (the coefficient of α_1 is negative) containing a few spectral components (the coefficient of SB is also negative). Such a search for a crystal-clear sound could explain the general choice of *Dalbergia* sp., which is the most resonant species and the most common in xylophone bars. Indeed, the predominance of α_1 agrees with the first rank of *Dalbergia* sp., for which $\alpha_1 = 14.28\ \text{s}^{-1}$ is the smallest in the data bank ($14.28\ \text{s}^{-1} < \alpha_1 < 45.61\ \text{s}^{-1}$) and SB $= 2224$ Hz is medium range in the data bank ($1268\ \text{Hz} < \text{SB} < 3930\ \text{Hz}$). Holz (1996) showed that the damping factor value α_1 should be lower than about $30\ \text{s}^{-1}$ for a fundamental frequency value of 1000 Hz, which corresponds to the mean value of the study. The average value of α_1 is indeed $26.13\ \text{s}^{-1}$ with a standard deviation of $6.18\ \text{s}^{-1}$. Actually, xylophone makers use a specific way of carving the bar by removing substance in the middle (Fletcher and Rossing, 1998). This operation tends to minimize the importance of partial 2, decreasing both the SCG and the SB. The importance of α_1 in the model is in line with several studies showing that the damping is a pertinent clue in the perception of impacted materials (Klatzky *et al.*, 2000; Wildes and Richards, 1988). Concerning parameter SB, the spectral distribution of energy is also an important clue, especially for categorization purposes.

The linear classification prediction has been improved by taking into account nonlinear phenomena. The nonlinear model then explains 82% of the variance. The nonlinear relationship between the perceptual classification and predictors (α_1 and SB) was explained by the instrument maker's strategy during the evaluation of each sample. The xylophone maker proceeded by first identifying the best samples and then the worst samples. This first step gave him the upper and lower bounds of the classification. The final step was to sort the samples of medium quality and place them between the bounds. One could deduce that three groups of acoustic quality (good, poor, and medium quality) were formed before the classification and that inside these groups the perceptual distance between each sample was different. The sigmoid shape indicated that the perceptual distance was shorter for good and poor quality groups than for medium quality groups. As a consequence, the nonlinear model is probably linked with the way the maker proceeded and cannot be interpreted as an intrinsic model for wood classification. Another explanation for the nonlinear relationship can also be found in the nonlinear transform relating physical and perceptual dimensions.

Note finally that there was no correlation between the classification and the wood density. However it is known that the wood density is of great importance for instrument makers. Holz (1996) suggested that the "ideal" xylophone wood bars would have density values between 800 and 950 kg/m^3. This phenomenon is due to the way we designed our experimental protocol, focusing on the sound itself and minimizing multi-sensorial effects (avoiding the access to visual and tactile information). Actually, in a situation where the instrument maker has access to the wood, bars with weak density are rejected for manufacturing and robustness purposes, irrespective of their sound quality.

C. Influence of the fundamental frequency (pitch) on the classification

As discussed previously, timbre is a key feature for appreciating sound quality and it makes it possible to distinguish tones with equal pitch, loudness, and duration (ANSI, 1973). Since this study aims at better understanding which timbre descriptor is of interest for wood classification, one expected differences in the classification of the tuned and the original sound data banks. The difference between classifications C2 (various pitches) and C3 (same pitches) shows a clear linear tendency; it is represented in Fig. 6 as a function of the original fundamental frequency of the bars. The difference is negative (respectively positive) for sounds whose fundamental frequencies are lower (respectively higher) than the mean frequency. The Pearson coefficient associated with the linear relationship between the arithmetic difference of the classification and the fundamental frequency leads to the important observation that *a wooden bar with a low fundamental frequency tends to be upgraded while a wooden bar with a high fundamental frequency tends to be downgraded*. This finding agrees with our linear prediction model, which predicts weakly damped sounds would be better classified than highly damped ones. Actually, sounds with low (respectively high) fundamental frequencies were transposed toward high (respectively low) frequencies during the tuning process, implying α_1 increase (respectively decrease), since the damping is proportional to the square of the frequency (cf. Sec. III C). As an important conclusion, one may say that the instrument maker cannot judge the wood itself independently of the bar dimensions, since the classification is influenced by the pitch changes, favoring wood samples generating low fundamental frequency sounds.

Once again, note the good reliability of our instrument maker, who did not change the classification of sounds whose fundamental frequency was close to the mean fundamental frequency of the data bank (i.e., sounds with nearly unchanged pitch). Actually, the linear regression line passes close to 0 at the mean frequency 1002 Hz. Moreover, the *Dalbergia* sp. was kept at the first position after the tuning process, suggesting that no dramatic sound transformations had been made. In fact, this sample was transposed upwards by 58 Hz, changing α_1 from 13.6 s^{-1} to 14.28 s^{-1}, which still was the smallest value of the tuned data bank.

D. Relationship between wood anatomy and perceived musical quality

The damping α_1 of the first vibrational mode was an important descriptor explaining the xylophone maker classification. Equation (11) shows that this descriptor is related to the quality factor Q, and consequently to the internal friction coefficient tan ϕ (inverse of the quality factor Q), which depends on the anatomical structure of the wood. An anatomical description of the best classified species has been discussed in a companion article (Brancheriau *et al.*, 2006b). We briefly summarize the main conclusions and refer the reader to the article for more information. A draft anatomical portrait of a good acoustic wood could be drawn up on the basis of our analysis of wood structures in the seven acoustically best and seven poorest woods. This portrait should include a compulsory characteristic, an important characteristic, and two or three others of lesser importance. The key trait is the axial parenchyma. It should be paratracheal, and not very abundant if possible. If abundant (thus highly confluent), the bands should not be numerous. Apotra-cheal parenchyma can be present, but only in the form of well-spaced bands (e.g., narrow marginal bands). The rays (horizontal parenchyma) are another important feature. They should be short, structurally homogeneous but not very numerous. The other characteristics are not essential, but they may enhance the acoustic quality. These include:

- Small numbers of vessels (thus large);
- A storied structure;
- Fibers with a wide lumen (or a high flexibility coefficient, which is the ratio between the lumen width and the fiber width; it is directly linked with the thickness of the fiber).

These anatomical descriptions give clues for better choosing wood species to be used in xylophone manufacturing. They undoubtedly are valuable for designing new musical materials from scratch, such as composite materials.

VIII. CONCLUSION

We have proposed a methodology associating analysis-synthesis processes and perceptual classifications to better understand what makes the sound produced by impacted wooden bars attractive for xylophone makers. This methodology, which focused on timbre-related acoustical properties, requires equalization of the pitch of recorded sounds. Statistical analysis of the classifications made by an instrument maker highlighted the importance of two salient descriptors: the damping of the first partial and the spectral bandwidth of the sound, indicating he searched for highly resonant and crystal-clear sounds. Moreover, comparing the classifications of both the original and processed sounds showed how the pitch influences the judgment of the instrument maker. Indeed, sounds with originally low (respectively high) fundamental frequency were better (lesser) classified before the tuning process than after. This result points to the preponderance of the damping and reinforces the importance of the pitch manipulation to better dissociate the influence of the wood species from that of the bar geometry. Finally, the results revealed some of the manufacturers' strategies and pointed out important mechanical and anatomical characteristics of woods used in xylophone manufacturing. From a perceptual point of view, the internal friction seems to be the most important characteristic of the wood species. Nevertheless, even though no correlation has been evidenced between the classification and the wood density, it is well known that this parameter is of great importance for instrument makers as evidence of robustness. As mentioned in the introduction, this work was the first step towards determining relations linking sounds and wood materials. Future works will aim at confirming the results described in this paper by taking into account classifications made by other xylophone makers in the statistical analysis. We plan to use this methodology on a new set of wood species having mechanical and anatomical characteristics similar to those well classified in the current test. This should point out unused wood species of interest to musical instrument manufacturers and will give clues for designing new musical synthetic materials.

ACKNOWLEDGMENTS

The authors thank Robert Hébrard, the xylophone maker who performed the acoustic classification of the wood species. They are also grateful to Pierre Détienne for useful advice and expertise in wood anatomy. They also thank Bloen Metzger and Dominique Peyroche d'Arnaud for their active participation in the experimental design and the acoustical analysis, and Jérémy Marozeau who provided the graphical interface for the listening test. We would also thank the reviewers for useful suggestions.

American National Standards Institute (**1973**). *American National Standard Psychoacoustical Terminology* (American National Standards Institute, NY).

Adrien, J. M. (**1991**). *The Missing Link: Modal Synthesis* (MIT Press, Cambridge, MA), Chap. 8, pp. 269–297.

Avanzini, F., and Rocchesso, D. (**2001**). "Controlling material properties in physical models of sounding objects," in *Proceedings of the International Computer Music Conference 2001*, 17–22 September 2001, Hawana, pp. 91–94.

Beauchamps, J. W. (**1982**). "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," J. Audio Eng. Soc. **30**(6), 396–406.

Blay, M., Bourgain , and Samson (**1971**). "Application des techniques électroacoustiques à la détermination du module d'élasticité par un procédé nondestructif (Application of electroacoustic techniques to determine the elasticity modulus by nondestructive procedure)," Technical Review to Advance Techniques in Acoustical, Electrical and Mechanical Measurement **4**, 3–19.

Bork, I. (**1995**). "Practical tuning of xylophone bars and resonators," Appl. Acoust. **46**, 103–127.

Brancheriau, L., and Baillères, H. (**2002**). "Natural vibration analysis of

clear wooden beams: A theoretical review," Wood Sci. Technol. **36**, 347–365.
Brancheriau, L., Baillères, H., Détienne, P., Gril, J., and Kronland-Martinet, R. (**2006a**). "Key signal and wood anatomy parameters related to the acoustic quality of wood for xylophone-type percussion instruments," J. Wood Sci. **52**(3), 270–274.
Brancheriau, L., Baillères, H., Détienne, P., Kronland-Martinet, R., and Metzger, B. (**2006b**). "Classifying xylophone bar materials by perceptual, signal processing and wood anatomy analysis," Ann. Forest Sci. **62**, 1–9.
Bucur, V. (**1995**). *Acoustics of Wood* (CRC Press, Berlin).
Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (**2005**). "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," J. Acoust. Soc. Am. **118**(1), 471–482.
Chaigne, A., and Doutaut, V. (**1997**). "Numerical simulations of xylophones. I. Time-domain modeling of the vibrating bars," J. Acoust. Soc. Am. **101**(1), 539–557.
Dillon, W. R., and Goldstein, M. (**1984**). *Multivariate Analysis—Methods and Applications* (Wiley, New York).
Doutaut, V., Matignon, D., and Chaigne, A. (**1998**). "Numerical simulations of xylophones. II. Time-domain modeling of the resonator and of the radiated sound pressure," J. Acoust. Soc. Am. **104**(3), 1633–1647.
Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, Berlin).
Giordano, B. L., and McAdams, S. (**2006**). "Material identification of real impact sounds: Effects of size variation in steel, wood, and Plexiglass plates," J. Acoust. Soc. Am. **119**(2), 1171–1181.
Graff, K. F. (**1975**). *Wave Motion in Elastic Solids* (Ohio State University Press), pp. 100–108.
Holz, D. (**1996**). "Acoustically important properties of xylophon-bar materials: Can tropical woods be replaced by European species?" Acust. Acta Acust. 82(6), 878–884.
Klatzky, R. L., Pai, D. K., and Krotkov, E. P. (**2000**). "Perception of material from contact sounds," Presence: Teleoperators and Virtual Environments **9**(4), 399–410.
Lutfi, R. A., and Oh, E. L. (**1997**). "Auditory discrimination of material changes in a struck-clamped bar," J. Acoust. Soc. Am. **102**(6), 3647–3656.
Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.
Matsunaga, M., and Minato, K. (**1998**). "Physical and mechanical properties required for violin bow materials II. Comparison of the processing properties and durability between pernambuco and substitutable wood species," J. Wood Sci. **44**(2), 142–146.
Matsunaga, M., Minato, K., and Nakatsubo, F. (**1999**). "Vibrational property changes of spruce wood by impregnating with water-soluble extractives of pernambuco (*Guilandina echinata Spreng*.)," J. Wood Sci. **45**(6), 470–474.
Matsunaga, M., Sugiyama, M., Minato, K., and Norimoto, M. (**1996**). "Physical and mechanical properties required for violin bow materials," Holzforschung **50**(6), 511–517.
McAdams, S., Chaigne, A., and Roussarie, V. (**2004**). "The psychomechanics of simulated sound sources: Material properties of impacted bars," J. Acoust. Soc. Am. **115**(3), 1306–1320.
McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. D., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.
Obataya, E., Umewaza, T., Nakatsubo, F., and Norimoto, M. (**1999**). "The effects of water soluble extractives on the acoustic properties of reed (*Arundo donax* L.)," Holzforschung **53**(1), 63–67.
Ono, T., and Norimoto, M. (**1983**). "Study on Young's modulus and internal friction of wood in relation to the evaluation of wood for musical instruments," Jpn. J. Appl. Phys., Part 1 **22**(4), 611–614.
Ono, T., and Norimoto, M. (**1985**). "Anisotropy of Dynamic Young's Modulus and Internal Friction in Wood," Jpn. J. Appl. Phys., Part 1 **24**(8), 960–964.
Steiglitz, K., and McBride, L. E. (**1965**). "A technique for the identification of linear systems," IEEE Trans. Autom. Control **AC-10**, 461–464.
Sugiyama, M., Matsunaga, M., Minato, K., and Norimoto, M. (**1994**). "Physical and mechanical properties of pernambuco (*Guilandina echinata* Spreng.) used for violin bows," Mokuzai Gakkaishi **40**, 905–910.
Valette, C., and Cuesta, C. (**1993**). *Mécanique de la Corde Vibrante (Mechanics of Vibrating String)*, Traité des Nouvelles Technologies, série Mécanique (Hermès, Paris).
Wildes, R. P., and Richards, W. A. (**1988**). *Recovering Material Properties from Sound* (MIT Press, Cambridge, MA), Chap. 25, pp. 356–363.

Optical and acoustic monitoring of bubble cloud dynamics at a tissue-fluid interface in ultrasound tissue erosion

Zhen Xu[a] and Timothy L. Hall
Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

J. Brian Fowlkes
Department of Radiology and Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109

Charles A. Cain
Department of Biomedical Engineering and Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109

(Received 27 July 2006; revised 19 January 2007; accepted 28 January 2007)

Short, high-intensity ultrasound pulses have the ability to achieve localized, clearly demarcated erosion in soft tissue at a tissue-fluid interface. The primary mechanism for ultrasound tissue erosion is believed to be acoustic cavitation. To monitor the cavitating bubble cloud generated at a tissue-fluid interface, an optical attenuation method was used to record the intensity loss of transmitted light through bubbles. Optical attenuation was only detected when a bubble cloud was seen using high speed imaging. The light attenuation signals correlated well with a temporally changing acoustic backscatter which is an excellent indicator for tissue erosion. This correlation provides additional evidence that the cavitating bubble cloud is essential for ultrasound tissue erosion. The bubble cloud collapse cycle and bubble dissolution time were studied using the optical attenuation signals. The collapse cycle of the bubble cloud generated by a high intensity ultrasound pulse of 4–14 μs was ~40–300 μs depending on the acoustic parameters. The dissolution time of the residual bubbles was tens of ms long. This study of bubble dynamics may provide further insight into previous ultrasound tissue erosion results. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2710079]

PACS number(s): 43.80.Gx, 43.35.Ei, 43.80.Sh [FD] Pages: 2421–2430

I. INTRODUCTION

Tissue disintegration using ultrasound induced cavitation[1–9] and shockwave generated tissue destruction[10,11] have been observed by many researchers. Our recent studies have shown that short, high-intensity pulses delivered at certain pulse repetition frequencies (PRFs) can achieve complete mechanical tissue fragmentation.[12–14] This technique can be considered as a form of soft tissue lithotripsy ("histotripsy"). At a tissue-fluid interface, histotripsy results in tissue erosion with sharply demarcated boundaries.[12] Acoustic cavitation is the hypothesized primary mechanism for the erosion process, which is supported by a high amplitude and temporally changing acoustic backscatter observed during erosion. This highly variable backscatter is likely due to the sound reflection from a dynamically changing bubble cluster generated by high intensity ultrasound pulses. The bubbles behave as one dynamic unit, growing and shrinking together, and therefore this bubble cluster can be called a "bubble cloud" as defined by previous cavitation researchers.[15–18]

The temporally varying acoustic backscatter does not occur immediately at the onset of the ultrasound exposure. It takes a certain period of time (or certain number of pulses) to initiate this backscatter pattern, depending on the acoustic pulse parameters.[19] The time to initiation is shorter at higher pressures. Initiation of this backscatter pattern is an excellent indicator of erosion (98% positive predication rate).[19] After initiation, sometimes when the ultrasound pulses are still being delivered, the variable backscatter stops, which we label as extinction.[19] When extinction occurs, tissue erosion ceases. The variable backscatter can be re-initiated again without changing the pulse parameters.[19] The extinction and the re-initiation are both stochastic events. In this paper, we study the initiation and extinction by recording an optical attenuation signal and correlating it to the acoustic backscatter. The optical attenuation method detects intensity loss of the transmitted light through the bubble cloud. Further, we study the bubble cloud dynamics including cloud collapse cycle and bubble dissolution time using the optical attenuation method and high speed imaging.

The acoustic pressures effective for histotripsy are similar to those found in lithotripter shockwave pulses. Our histotripsy pulses are several acoustic cycles in duration instead of the one cycle pulses sometimes used in lithotripsy. The selection of pulse pressure, pulse duration and PRF affects the extent and efficiency of the mechanical tissue disruption induced by histotripsy.[12,13,20,21] Shockwave lithotripsy studies have shown that positive pressure can compress existing bubbles, while following negative pressure can cause bubble growth and collapse. The growth and collapse cycles are

[a]Author to whom correspondence should be addressed. Electronic mail: zhenx@umich.edu

TABLE I. Acoustic parameters used in all figures.

Fig. No.	Pulse duration	P–[a] (MPa)	P+[b] (MPa)	I_{SPPA} (W/cm^2)	PRF (Hz)	No. Pulse	PO_2	W/O Tissue	Att.[d] Duration (μs)
2	4 μs (3 cycles)	>21[c]	>76[c]	>29.3 k[c]	N/A	1	98–100%	Water	N/A
3	8 μs (6 cycles)	>21[c]	>76[c]	>29.3 k[c]	N/A	1	98–100%	Water	187
4,5	14 μs (10 cycles)	21	76	29.3 k	10	200	98–100%	Water	243.2±153.3
6	4 μs (3 cycles)	17.1	52.9	19.3 k	2 k	910	33–40%	Water	N/A
		21	76	29.3 k					
7	4 μs (3 cycles)	15.5	28.4	10.9 k	2 k	200	22–24%	Tissue	40.1±6.1
8,9	4 μs (3 cycles)	13.9	25.1	9.5 k	200	910	98–100%	Tissue	53.3±24.8
10	4 μs (3 cycles)	13.9	25.1	9.5 k	200	910	98–100%	Tissue	87.0
		15.5	28.4	10.9 k					95.4
		17.1	52.9	19.3 k					113.1
11,12	8 μs (6 cycles)	>21[c]	>76[c]	>29.3 k[c]	N/A	1	22–24%	Tissue	288.2
		5.2	6.6	1.1 k					

[a]P–: peak negative pressure.
[b]P+: peak positive pressure.
[c]In Figs. 2, 3, 11, and 12 the pressure levels could not be successfully measured due to instantaneous cavitation. At a lower power level, P– was measured to be 21 MPa and P+ 76 MPa.
[d]Attenuation duration in mean ± standard deviation.

long (hundreds of μs) compared to the lithotripsy pulse length (several μs). The bubble radius-time curve has been modeled[22] and confirmed experimentally *in vitro*[23–25] and *in vivo*.[26–28] In this paper, we used the transmitted light signals to trace the growth and collapse cycle of the bubble cloud generated by histotripsy pulses.

The dissolution time of residual bubbles from the collapse has also been of great interest in cavitation studies.[24,29] The residual bubbles can provide seeds for subsequent cavitation events.[30,31] Therefore, the bubble dissolution time can be critical for the timing of the next pulse for controlling cavitational effects. We studied the bubble dissolution time by delaying a lower amplitude pulse which follows the high amplitude histotripsy pulse. The lower amplitude pulse generates a bubble cloud only when residual bubbles remained from the previous pulse to serve as cavitation nuclei. Dissolution time of residual bubble cavitation nuclei generated by the histotripsy pulse was determined by changing the delay time between the two pulses and monitoring subsequent bubble generation by the lower amplitude pulse.

II. METHODS

A. Ultrasound generation and calibration

Ultrasound pulses were generated by an 18-element piezocomposite spherical-shell therapeutic array (Imasonic, S.A., Besançon, France) with a center frequency of 750 kHz and a geometric focal length of 100 mm. The therapy array has an annular configuration with outer and inner diameters of 145 and 68 mm, respectively, yielding a radiating area of ~129 cm^2. All the array elements were excited together in phase. The array driving system, maintained under PC control, consists of channel driving circuitry, associated power supplies (Model 6030A, HP, Palo Alto, CA), and a software platform to synthesize driving patterns. A PC console also provided control of a motorized three-dimensional positioning system (Parker Hannifin, Rohnert Park, CA) to position the array at each exposure site.

The pressure wave form at the focus of the 18-element array in the acoustic field was measured in degassed water (12–25% concentration) (i.e., free-field conditions) using a fiber-optic probe hydrophone developed in house.[32] The lateral and axial pressure profiles of the focused beam were measured to be 2.2 mm × 12.6 mm in width (full width at half maximum), at peak negative pressure of 14 MPa and 1.8 × 11.9 mm at 19 MPa. The beam width decreased with increasing pressure. The peak negative and positive pressures and spatial-peak pulse-average intensity (I_{SPPA})[33] used in experiments depicted in Figs. 4–10 were measured for free-field conditions and reported in Table I. The pressure levels used in other experiments (Figs. 2, 3, 11, and 12) could not be calibrated successfully due to the instantaneous cavitation. However, we were able to measure a peak negative pressure of 21 MPa and a peak positive pressure of 76 MPa at a lower power level without generation of bubbles during measurement.

B. Tissue sample preparation

Fresh porcine atrial wall tissue (1–2 mm thick) was obtained from a local abattoir and used within 24 h of harvesting. All tissue specimens were preserved in a 0.9% sodium chloride solution at 4 °C. Tissue was wrapped over ring-shaped tube fitting (2 cm in diameter), so that no tissue interfered with the laser beam. The tissue was degassed by submerging in degassed water for 1 h prior to experimentation.

C. Optical attenuation detection

The optical attenuation method monitors the light transmission and detects the light beam reduction caused by bubbles. The schematic diagram of the experimental setup is shown in Fig. 1. Bubble clouds were produced in a 30-cm-wide × 60-cm-long × 30-cm-high water tank designed to enable optical observations. A 1 mW helium-neon gas laser (Model. 79245, Oriel, Stratford, CT) was placed on one

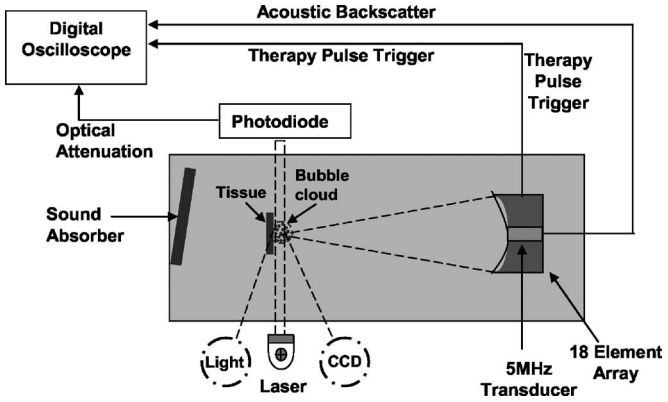


FIG. 1. Experimental setup for bubble cloud monitoring at a tissue-water interface using optical attenuation and acoustic backscatter methods. Light source and CCD camera in dashed circle are set up for high speed imaging in water. However, at a tissue-water interface, the light source was blocked by the tissue and the imaging could not be used with the optical attenuation detection.

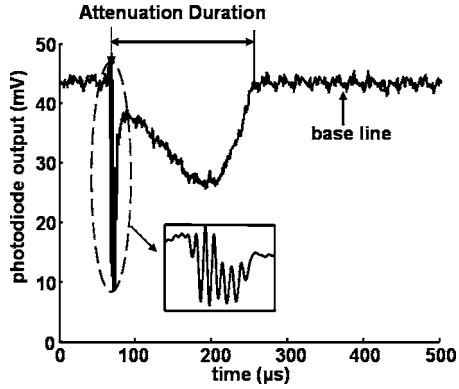


FIG. 3. Light attenuation signal caused by a bubble cloud recorded as a photodetector output. The bubble cloud was generated by a 6-cycle (8 μs) histotripsy pulse in water. The top left arrow indicates the arrival of the histotripsy pulse at the therapy transducer focus where the laser beam is projected. The inset is an expanded view (compressed in vertical direction and expanded in horizontal direction) of the optical signal tracking the ultrasound pulse wave form.

side of the tank to pass a laser beam through the ultrasound focus (and in front of the tissue at a tissue-water interface). The light intensity was monitored continuously by a photodiode (Model DET100, ThorLabs, Newton, NJ) aligned with the laser beam at the other side of the tank.

To direct the laser beam through the ultrasound focus, the therapy transducer was first pulsed in free water to create a visible bubble cloud at its focus. A photo of the bubble cloud taken by a high speed camera is shown in Fig. 2. The position of the transducer was adjusted to direct the laser beam through the center of the bubble cloud. To form a tissue-water interface, a piece of porcine atrial wall was placed parallel and right behind the laser beam. The laser beam width (0.48 mm @$1/e^2$) was smaller than the bubble cloud, so the photodiode measured the light transmitted through a portion of the bubble cloud and not the whole cloud.

The attenuated light signal was recorded as the voltage output of the photodiode. The photodiode output was connected to a digital oscilloscope (Model 9384L, LeCroy Chestnut, NY) using a 1 MΩ dc coupling in parallel with a 250 Ω resistor. An impedance of 250 Ω was chosen to achieve a good signal to noise ratio (30–35 dB) and a wide enough dynamic range (60 dB) for attenuation detection, while still maintaining good temporal resolution (–3 dB width response time of 15 ns). The acquisition of the optical signals was synchronized with the acoustic therapy pulse.

FIG. 2. An image of the bubble cloud taken by a high speed camera. The bubble cloud was generated by a single pulse of three cycles in gas saturated free water. It was used for the laser beam alignment. The ultrasound pulse propagated from the left to the right side of the image.

An example of a light attenuation signal is demonstrated in Fig. 3, in which the light intensity began to decrease at the arrival of a single 8 μs histotripsy pulse (produced by a driving signal of 6 cycle pulse at 750 kHz) in water. During the time window when the ultrasound pulse propagated through the laser beam, the light attenuation signal seemed to track the ultrasound therapy pulse wave form (inset in Fig. 3). This acousto-optic "artifact" is most likely due to changing in the index of refraction of water during the therapy pulse. It provides a convenient timing indicator locating the therapy pulse with respect to the generated bubble cloud and will be discussed further in Results and Discussion.

Researchers have used optical scattering and reflection signals from bubbles to effectively trace radius-time curves generated by the lithotripsy pulse.[24,26,34] Based on these studies, duration of the light attenuation (attenuation duration) is believed to indicate the bubble cloud collapse time. Attenuation duration is defined as the period of time when the light intensity (photodiode output) falls below a threshold of base line –3 times the noise level. The base line and noise level are mean and standard deviation (SD) values, respectively, of the photodiode output receiving the laser light when no bubbles are present. Initiation of light attenuation occurs when the light attenuation duration exceeds the pulse duration for five consecutive pulses. Extinction of light attenuation occurs when the light attenuation duration drops below the pulse duration for five consecutive pulses. The purpose of using pulse duration as a threshold is to overcome the light attenuation increase due to the ultrasound induced water index of refraction change. Although the threshold for attenuation reestablishing base line is arbitrary, varying the threshold slightly did not affect the detection of initiation and extinction events.

D. Acoustic backscatter

To receive the acoustic backscatter of therapy pulses, a 5 MHz, 2.5-cm-diam single element focused transducer

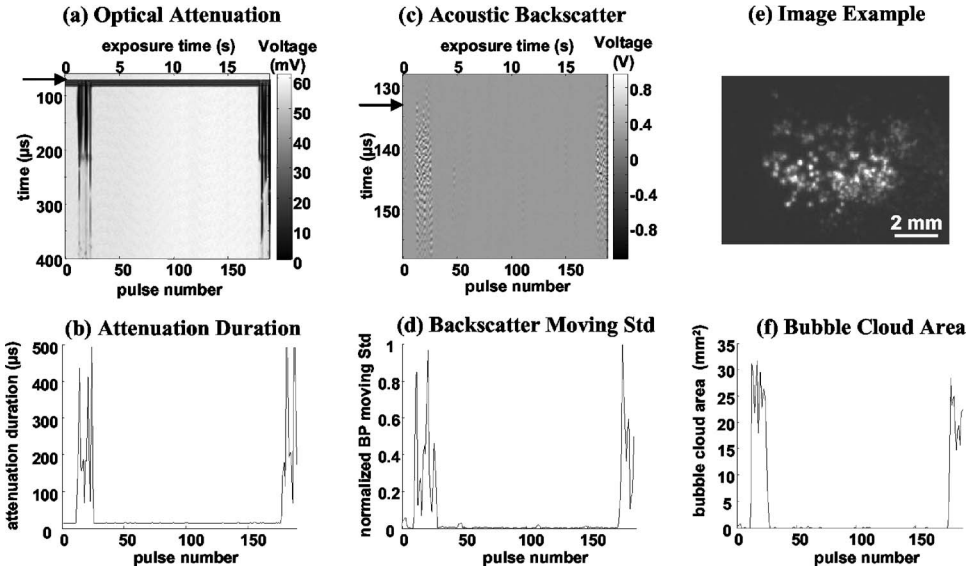


FIG. 4. Light attenuation, variable acoustic backscatter and corresponding bubble cloud imaging in water. (a) Light attenuation. Each vertical line is a light intensity signal corresponding to one ultrasound pulse, with light intensity encoded in gray color. Dark color indicates a decrease in the light intensity. The horizontal axis is pulse number. (b) The light attenuation duration for each pulse vs pulse number. (c) Acoustic backscatter in fast time-slow time display. Each vertical line is an *A*-line acoustic backscatter signal where the signal amplitude is encoded in gray color. (d) The normalized backscatter power moving SD vs pulse number. (e) Example bubble cloud image. Each image was taken 10 μs after the arrival of a 14 μs pulse. (f) Bubble cloud cross-sectional area vs pulse number. Horizontal arrows indicate the arrival of the histotripsy pulse.

(Valpey Fisher Corporation, Hopkinton, MA) with a 10 cm focal length was mounted confocally with the therapy array inside its inner hole. Acoustic backscatter signals were recorded and displayed as range-gated temporal voltage traces by a digital oscilloscope (Model 9384L, LeCroy, Chestnut Ridge, NY). The recorded wave forms were then transferred through general purpose interface bus and processed by the MATLAB program (Mathworks, Natick, MA).

Normalized acoustic backscatter power moving standard deviation (SD) was used to characterize the variability of backscatter which is described in our previous paper.[19] As the acoustic backscatter was due to reflected therapy pulses, the backscatter power was first normalized to a reference proportional to the therapy pulse power, which was determined by reflection from a stainless steel reflector.[35] Normalized backscatter power moving SD (moving window size =3) at a time point i was calculated as the standard deviation of backscatter power at point i, $i-1$, and $i-2$. The initiation and extinction of the temporally variable acoustic backscatter, relating to the beginning and suspension of erosion, were detected when the moving SD exceeds and falls below a threshold for five consecutive pulses, respectively.[19]

E. High-speed imaging

Images of a bubble cloud were captured by a fast gated, intensified 640×480 pixel, 12 bit, 11 frame/s charge coupled device (ICCD) camera (Picostar HR, La Vision, Goettingen, Germany).[36] A shutter speed of 100 ns was used. The imaging system could store up to 200 images at one time. The bubble cloud image was taken when the ultrasound pulse was propagating through the transducer focus. The bubble cloud was illuminated at a forward 30° angle using a xenon arc lamp (Model. 60069 Q Series, Oriel, Stratford, CT). We were able to use high speed imaging and optical attenuation monitoring simultaneously in water. However, at a tissue-water interface, the imaging setup could not be used simultaneously with the optical attenuation setup, because the light illumination for imaging was blocked by tissue as indicated in Fig. 1.

The ICCD camera captured images by detecting and recording a count proportional to the photon number at each pixel. The bubble presence at each pixel was determined when the photon count exceeded a threshold of mean +3 SD of the photon count at this pixel with no bubbles. The camera captured the image of the bubble cloud along the axial direc-

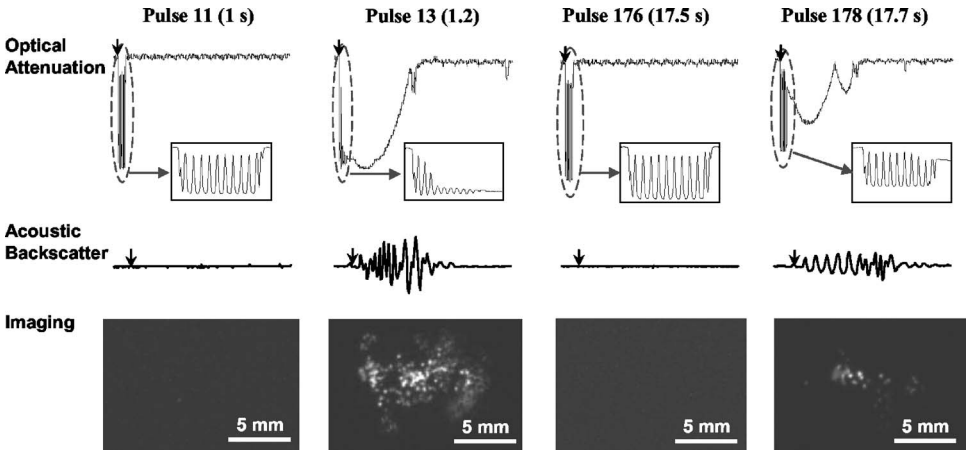


FIG. 5. Wave forms of synchronized light attenuation, acoustic backscatter signals and corresponding bubble cloud images from Fig. 4. The ultrasound pulse propagated from the left to the right side of the image. The short arrows in the top two rows indicate the arrival of the histotripsy pulse at the focus of the therapy transducer. The light attenuation (after the pulse) and variable acoustic backscatter signals were only detected when a bubble cloud was observed using high speed imaging. The insets in the top row are expanded views of the optical signal tracking ultrasound pulse wave form during the pulse.

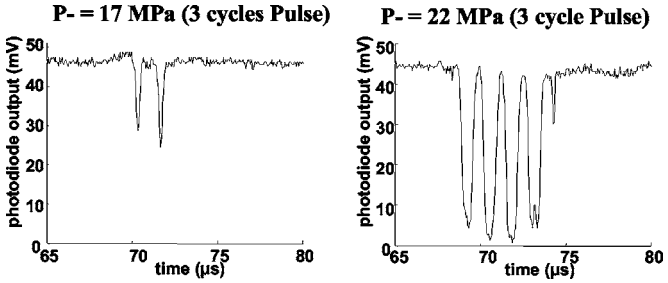


FIG. 6. Optical attenuation signal tracking the ultrasound pulse wave form without the formation of a bubble cloud. The peak amplitude of the signal increased with increasing pressure.

tion of the ultrasound beam (Figs. 4 and 5). We estimated the axial cross-sectional area of the bubble cloud by integrating the area of pixels with bubbles.

III. RESULTS

A. Light attenuation signals and bubble cloud imaging

Figures 4 and 5 show light attenuation signals synchronized with images of bubble clouds generated by histotripsy pulses in water. Each bubble cloud image was captured when an ultrasound pulse was propagating through its focus. When a bubble cloud was observed using high speed imaging, the light attenuation was also observed during and after the ultrasound pulse propagated through the laser beam. When no bubble cloud was seen, no light attenuation was observed after the ultrasound pulse, although the light intensity signal tracked the ultrasound pulse wave form during the pulse. The correspondence between light attenuation signals and bubble clouds observed by high speed imaging suggests that the light attenuation was caused by the bubble cloud.

The imaged bubble clouds all appear to consist of multiple bubbles. The sizes of bubble clouds generated in water varied from pulse to pulse. In Fig. 4, the short diameter of the bubble cloud (along the lateral beam of the transducer) was 2–6 mm, and the long diameter (along the axial beam of the transducer) was 3–13 mm. The total cross-sectional area of bubble cloud ranged from 15 to 30 mm^2.

In the absence of an observable bubble cloud, the light attenuation signal tracked the ultrasound wave form during the ultrasound pulse (Figs. 5 and 6). This signal was relatively constant for a given pulse pressure (Fig. 5) and its peak amplitude increased with increasing pressure (Fig. 6). These observations suggest that this signal is due to the water index of refraction change caused by pulse pressure fluctuations. The light signal which tracked the ultrasound wave form was also observed when a bubble cloud was seen (Fig. 5). In the presence of bubbles, it varied from pulse to pulse, even under the same pulse parameters (Fig. 5).

B. Initiation and extinction

Initiation of a temporally varying acoustic backscatter signal corresponded well to initiation of light attenuation both in water (Fig. 4) and at a tissue-water interface (Fig. 7). As light attenuation was only detected when the bubble cloud was generated, this result provides evidence that initiation of the variable acoustic backscatter was due to bubble cloud formation. Figure 4 shows an example of the initiation of the variable acoustic backscatter and light attenuation signals in water. The variable backscatter was initiated at 1.1 s (12th pulse) after the onset of insonation and the light attenuation was initiated at 1.2 s (13th pulse). Figure 7 presents an example of initiations of both signals at a tissue-water interface. The variable backscatter was initiated at 10.5 ms (21st pulse) after the onset of insonation, while light attenuation began at 18 ms (36th pulse). The slight difference between the two initiations could be due to: (1) the acoustic backscatter being more sensitive for bubble detection than the light attenuation; and/or (2) the cavitating bubble cloud initiated outside the laser beam first and later grew larger or moved into the laser beam. The latter is supported by the temporal shift of the acoustic backscatter in Fig. 7(c) showing the bubble cloud growing or moving towards the source transducer after the initiation of the variable backscatter.

The timing of extinction and re-initiation of the variable backscatter also corresponded well to the extinction and re-initiation of light attenuation. An example of extinction and re-initiation of the variable acoustic backscatter signal and light attenuation at a tissue-water interface is shown in Fig. 8. The corresponding wave forms are shown in Fig. 9. The light attenuation was only detected when the acoustic backscatter was of high amplitude and temporally, spatially changing. Parsons, Fowlkes, and Cain[37] have shown that temporal, spatial variation in the backscatter pattern was required for producing significant mechanical breakdown of bulk tissue using high-intensity pulsed ultrasound. Sometimes a relatively high amplitude but temporally spatially constant acoustic backscatter was detected when no light attenuation was observed at a tissue-water interface (Fig. 8). This high amplitude but static backscatter was likely the reflection from bubbles trapped on the uneven surface of the tissue and not actively cavitating. The optical system could not detect these bubbles due to their position on the tissue surface where the light beam could not reach them, or alternatively, the optical beam was not aligned close enough to the tissue to detect them.

We used a lower acoustic intensity for the tissue-water interface than in water alone, as the intensity threshold to initiate a bubble cloud detectable by the optical attenuation method appears to be lower at a tissue-water interface. The arrival of each therapy pulse at the transducer focus is indicated by an arrow in figures of optical attenuation and acoustic backscatter signals (Figs. 4–9). As the focal length of the therapy transducer was 100 mm, it took ~67 μs for the ultrasound therapy pulse to reach the laser beam and ~134 μs for acoustic backscatter detection.

C. Bubble cloud collapse cycle

The bubble cloud collapse cycle was detected by the duration of light attenuation (attenuation duration). The attenuation duration calculated from Figs. 3–12 ranged from 40–300 μs (Table I) which was 10–40× the ultrasound pulse duration (4–14 μs). In Fig. 3, a bubble cloud started to form at the arrival of a 6-cycle (8 μs) histotripsy pulse. It

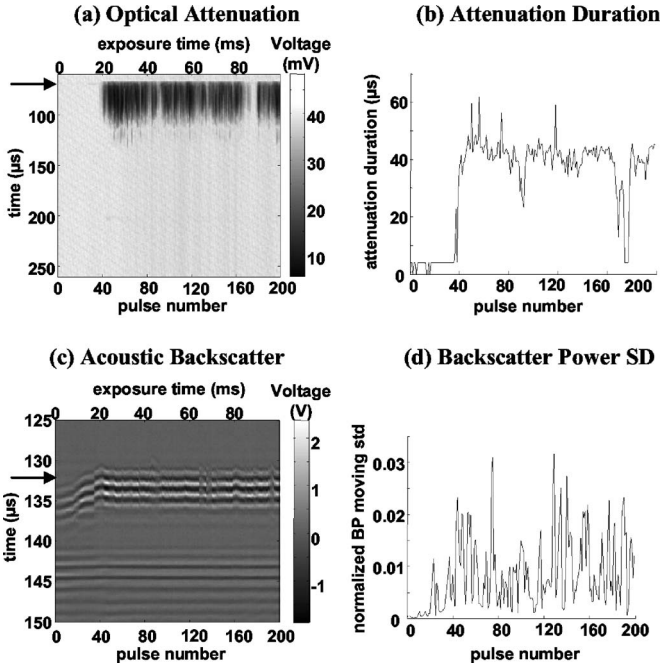


FIG. 7. Initiations of the light attenuation and the variable acoustic backscatter at a tissue-water interface. Panels (a)–(d) display the light attenuation and acoustic backscatter signals in the format described in Fig. 4. Initiation of the variable backscatter corresponded well to initiation of the light attenuation.

grew for over 100 μs as indicated by the light intensity decrease and collapsed when the light intensity returned to the base line level. The bubble cloud lasted for 187 μs before it collapsed to below the sensitivity level of the optical attenuation system ($\sim$20 times the duration of the histotripsy pulse). The pressure levels used in Fig. 3 could not be calibrated successfully due to the instantaneous cavitation. We measured a peak negative pressure of 21 MPa and a peak

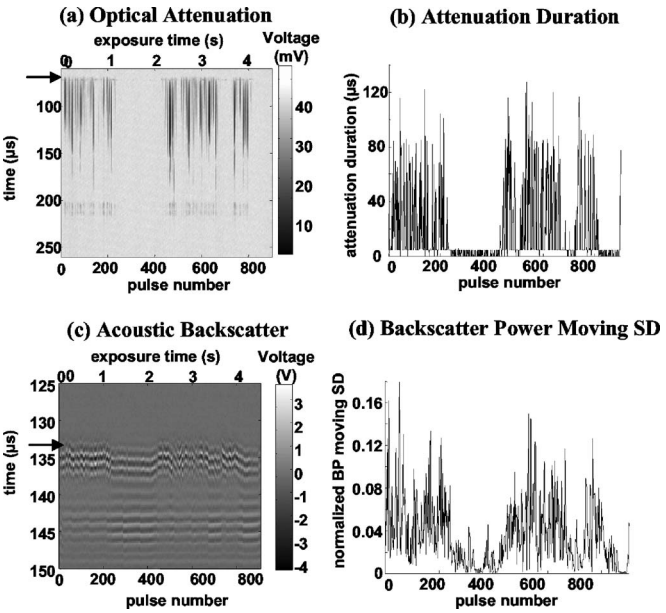


FIG. 8. Extinctions and re-initiations of the light attenuation and the variable acoustic backscatter at a tissue-water interface. Panels (a)–(d) display the light attenuation and acoustic backscatter signals in the format described in Fig. 4.

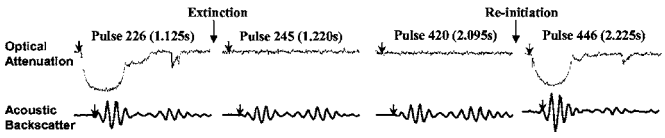


FIG. 9. Wave forms of light attenuation and acoustic backscatter signals in Fig. 8. The short arrows indicate the arrival of the histotrispy pulse at the focus of the therapy transducer. Variable backscatter was observed with light attenuation. Without light attenuation, a high amplitude but stable backscatter was observed.

positive pressure of 76 MPa at a lower power level. The attenuation duration varied with different pulse parameters (e.g., pressure, pulse duration and PRF) and gas content in the fluid. For example, the attenuation duration was longer with higher pressure (Fig. 10), indicating a longer collapse cycle.

D. Bubble dissolution time

Although the optical attenuation method was not sensitive enough to detect the presence of bubbles following the

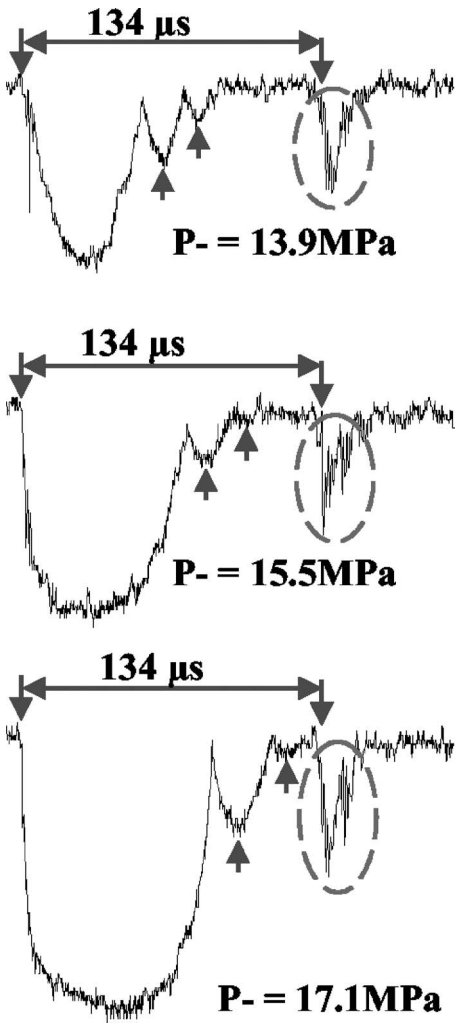


FIG. 10. Bubble cloud rebounds detected as the additional light attenuations (pointed by arrows) after the main light attenuation signal. The durations of the main light attenuation and the second peak are longer with higher pressure. Interestingly, the additional light attenuation peak in dashed circle always occurred at $\sim$134 μs after the ultrasound pulse even as the pulse pressure increased.

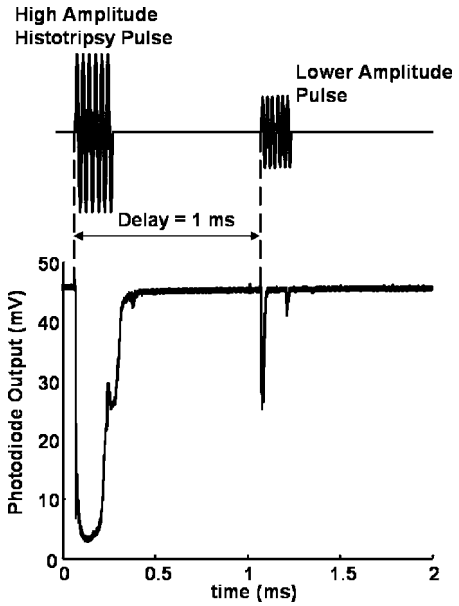


FIG. 11. A bubble cloud was first generated by a high amplitude histotripsy pulse ($p->21$ MPa, $p+>76$ MPa) which caused the light attenuation. One ms after the onset of the histotripsy pulse, a lower amplitude pulse ($p-=5.2$ MPa, $p+=6.6$ MPa) was delivered. The lower amplitude pulse, which could not produce a bubble cloud by itself, regenerated the bubble cloud resulting in another light attenuation. A pulse duration of 6 cycles and a gas concentration of 22–24% were employed.

collapse of the bubble cloud, residual bubbles survived long after the collapse. We measured the bubble dissolution time by delaying a lower amplitude pulse following the initial high amplitude histotripsy pulse. The lower amplitude pulse could not generate a bubble cloud itself. It had to use the residual bubbles produced by the previous high amplitude histotripsy pulse to recreate a bubble cloud. The dissolution time was determined by changing the delay time between the two pulses and monitoring the light attenuation caused by the subsequent lower amplitude pulse.

In Fig. 11, a high amplitude histotripsy pulse first created a bubble cloud at a tissue-water interface, causing light attenuation. The pressure levels of the high amplitude histotripsy pulse could not be calibrated successfully due to the instantaneous cavitation. The lower amplitude pulse ($p-=5.2$ MPa, $p+=6.6$ MPa) was delivered 1 ms later and regenerated the bubble cloud using the residual cavitation nuclei from the previous pulse (Fig. 11).

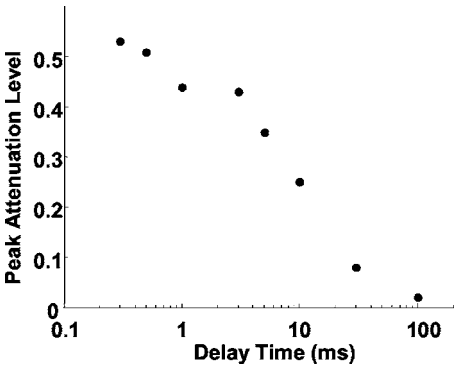


FIG. 12. The peak attenuation level of the light attenuation caused by the lower amplitude pulse as a function of the delay time. The peak attenuation level has a linearly decreasing trend on a log scale x axis. On a linear scale, the peak attenuation level displayed an exponential-like decay.

The delay time between the two pulses was varied from 300 μs to 100 ms. Results show that the lower amplitude pulse caused the light attenuation up to 30 ms, suggesting an approximately 30 ms dissolution time of the residual bubbles after the cloud collapse. The peak light attenuation level produced by the lower amplitude pulse showed an exponential-like decrease versus delay time (Fig. 12). This decrease is probably the result of an overall decreasing number and/or size of residual bubbles (resulting cavitation nuclei) over time due to dissolution.

E. Bubble cloud rebounds

Besides the primary light attenuation following the arrival of ultrasound pulses, additional drops in the light intensity were detected. Indicated by arrows in Fig. 10, there were two additional attenuation peaks immediately following the primary peak. The duration of the second peak was longer at higher pulse pressure, as was the third. These subsequent light intensity peaks were likely due to bubble cloud rebounds. The number of rebounds changed with acoustic parameters (e.g., pressure, PRF, pulse duration).

However, the light intensity peak marked by dashed circle in Fig. 10 always arrived at ~134 μs after the ultrasound pulse, and remained the same even as the pressure increased. The value of 134 μs was the round trip travel time between the source transducer and its focus, which suggests that these peaks were caused by the bubble cloud being driven by the sound reverberation between the source transducer surface and the initial cloud.

IV. DISCUSSION

In this paper, we monitored the intensity attenuation of light transmitted through cavitating bubble clouds both at a tissue-fluid interface and in water. The light attenuation was only detected when a bubble cloud was observed using high speed imaging. There were two types of bubbles studied: large macroscale bubbles and small microscale bubbles. The optical attenuation signals recorded activity of macroscale bubbles or bubble clouds large enough to cause light intensity loss. However, it was not sensitive enough to detect microscale bubbles including cavitation nuclei and bubble fragments following the collapse of the bubble cloud. Therefore, we used the optical signal to measure the cloud collapse cycle, and delayed a second pulse for bubble regeneration to indirectly measure the microscale bubble dissolution time. Based on the bubble cloud images (Figs. 2, 4, and 5), the optical signal was most likely measuring light through multiple bubbles along the laser beam, not the whole cloud (the laser beam was covering part of the whole cloud). However, since the bubbles within the cloud were growing and shrinking together, the dynamics of a portion of the bubble cloud may indicate the dynamics of the overall bubble cloud.

Our previous studies have shown tissue erosion was only seen with initiation of the variable acoustic backscatter signal.[19] The time to initiation is shorter at higher pressures.

After initiation, sometimes extinction of the variable acoustic backscatter occurs with the presence the ultrasound pulses, and correspondingly, erosion ceases.[19] The initiation and extinction of the variable acoustic backscatter correlated well with initiation and extinction of the light attenuation. This result suggests that initiation of the variable backscatter is due to the formation of a cavitating bubble cloud. This provides additional evidence that the cavitating bubble cloud is essential for histotripsy induced tissue erosion. Though the acoustic backscatter here only observed the sound reflection of the therapy pulses, it is possible to detect acoustic emission at the collapse of the bubble cloud. However, it requires an acoustic detector of higher sensitivity and lower frequency than the one used in this study.

The collapse cycle of the bubble cloud was measured by the duration of the light attenuation. The collapse cycle of a bubble cloud produced by a histotripsy pulse of 4–14 μs was ~40–300 μs. The bubble cloud underwent growth and collapse long after the ultrasound pulse ended and the collapse cycle varied with different acoustic parameters. These results are consistent with collapse cycles of bubble clouds generated by lithotripsy shockwave pulses, where pressure levels are similar to those produced by histotripsy pulses. Previous passive acoustic detection studies[23,28] and optical monitoring[24–26] have shown the bubbles generated by a lithotripter pulse collapsing several hundred μsed after the pulse. Church's model[22] predicted the same trend.

Our results also show that residual bubbles (resulting cavitation nuclei) survived long (~30 ms) after the bubble cloud collapse. The number of residual cavitation nuclei decreased over time due to the dissolution of bubbles. Using the residual cavitation nuclei from the previous pulse as seeds, the bubble cloud can be re-initiated at a much lower pressure than what is required to initiate it. This result provides an explanation for our previous observations that erosion can be maintained at intensities significantly lower than the intensity threshold required to initiate erosion.[20] The reduction in the cavitation threshold after the initial sonication pulse has been shown *in vivo*.[38] Similarly, Parsons *et al.*[13] found that higher amplitude pulses interleaved with lower amplitude pulses (applied several ms after) can maintain sufficient cavitation activity to generate significant mechanical tissue disruption, while higher-amplitude pulses alone mainly formed thermally mediated lesions. Moreover, previous PRF studies[26,31] have demonstrated residual bubbles generated by a lithotripter pulse can last for several ms. The number of cavitation bubbles generated was greater with higher PRF, as residual bubbles from the previous pulse can serve as cavitation nuclei for the subsequent pulse. We do not exclude the possibility that tissue fragments from erosion may also serve as cavitation nuclei[39] for subsequent pulses.

The long dissolution of residual cavitation nuclei is consistent with our working hypothesis for the histotripsy mechanism. Namely, each pulse generates a bubble cloud with two functions: (1) a subset of bubbles of the right size (cavitation nuclei) collapses to remove a portion of tissue fragments; and (2) the residual bubbles and bubble fragments from the collapse undergo dynamic changes between pulses and provide cavitation nuclei for subsequent pulses. We hypothesize that if the next pulse arrives at the tissue interface when the population and/or density of the cavitation nuclei is "just right," optimal tissue erosion occurs. The timing, pressure, and duration of acoustic pulses are all expected to affect the dynamics of the bubble cloud and, subsequently, the erosion produced. We have achieved almost an order of magnitude faster erosion rate with the same amount of energy only by adjusting pulse duration.[12] Further studies on monitoring of the bubble cloud dynamics are needed, especially on how acoustic parameters affect the bubble cloud.

The optical attenuation signals also detected the bubble cloud rebounds following the primary collapse cycle of the bubble cloud. The time interval between the main cycles and the duration and number of subsequent rebounds increased with increasing pressure, which has been consistently observed in radius-time curve measurements of bubble clouds generated by lithotripsy pulses.[34] Interestingly, an additional attenuation peak always occurred at ~134 μs after the histotripsy pulse, which was the round trip travel time between the transducer and its focus. The timing of this additional light attenuation remained the same even at different pressures. A possible explanation for this peak is that the sound reverberation pressure from the source transducer surface was sufficient to drive the bubble cloud. This illustrates the greatly increased sensitivity of bubble cloud generation some time after a proceeding bubble cloud. This effect has important consequences for initiation and maintenance of the histotripsy treatment process and was studied previously.[40]

Optical signal tracking the ultrasound pulse wave form is another interesting observation. It is known that acoustic pressure fluctuations in the sound field can cause spatial-temporal variations in water density, resulting in changes to the water's index of refraction and the light intensity pattern transmitted through the water, which is called the "Schlieren effect."[41] The observed optical signal at the instance of acoustic exposure was most likely caused by the Schlieren effect because it was observed when no bubbles were generated, and its peak amplitude increased with increasing acoustic pressure. When a bubble cloud was generated, this optical signal tracking ultrasound wave form varied widely even with the same acoustic parameters. In this case, we do not exclude the effect of bubble oscillation, but we were not able to extract the bubble oscillation effect from the optical signal mixed with the Schlieren effect.

The current optical system is not sensitive enough to detect small individual bubbles. In addition, the laser beam was not wide enough to cover a whole bubble cloud and bubbles outside the laser beam were not detected. The sensitivity of the current optical system can be improved by widening the laser beam width using optical lenses, higher laser power, use of the photodetector array, and increasing the sensitivity of the photodetector. High speed imaging is the most direct means to resolve the absolute sizes and spatial distribution of bubbles in the bubble cloud, and will be used in a future study.

V. CONCLUSIONS

We monitored the loss of light transmitted through bubble clouds generated by histotripsy pulses (4–14 μs

long). The light attenuation correlated well to the bubble cloud formation observed by high speed imaging. The collapse cycle of a bubble cloud was detected to be 40–300 μs in duration. Residual bubbles survived 30 mes after the bubble cloud collapse. The number of the residual bubbles decreased over time due to bubble dissolution. Using the residual bubbles, the intensity threshold to regenerate bubble cloud is lower than that required to originally generate bubble cloud.

ACKNOWLEGMENTS

The authors thank Dr. Mary-Ann Mycek and Dr. Steve Ceccio for generously providing their lab resources to this work. We would like to thank Mekhala Raghavan, Ching-Wei Chang and Dhruv Sud for their help with high speed imaging. We want to thank Jessica Parsons for her help with acoustic calibration. This research has been funded in part by grants from the National Institutes of Health R01-HL077629, and Hitachi Central Research laboratory.

[1]F. J. Fry, G. Kossoff, R. C. Eggleton, and F. Dunn, "Threshold ultrasound dosages for structural changes in the mammalian brain," J. Acoust. Soc. Am. **48**, 1413–1417 (1970).
[2]F. Dunn and F. J. Fry, "Ultrasonic threshold dosages for the mammalian central nervous system," IEEE Trans. Biomed. Eng. **18**, 253–256 (1971).
[3]L. A. Frizzell, C. S. Lee, P. D. Aschenbach, M. J. Borrelli, R. S. Morimoto, and F. Dunn, "Involvement of ultrasonically induced cavitation in hind limb paralysis of the mouse neonate," J. Acoust. Soc. Am. **74**, 1062–1065 (1983).
[4]G. R. ter Haar, S. Daniels, and K. Morton, "Evidence for acoustic cavitation in vivo: threshold for bubble formation with 0.75 MHz continuous-wave and pulsed beam," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 162–164 (1986).
[5]J. B. Fowlkes, P. L. Carson, E. H. Chiang, and J. M. Rubin, "Acoustic generation of bubbles in excised canine urinary bladders," J. Acoust. Soc. Am. **89**, 2740–2744 (1991).
[6]K. Hynynen, "Threshold for thermally significant cavitation in dog's thigh muscle in vivo," Ultrasound Med. Biol. **17**, 157–169 (1991).
[7]J. Y. Chapelon, J. Margonari, F. Vernier, F. Gorry, R. Ecochard, and A. Gelet, "In vivo effects of high-intensity ultrasound on prostatic adenocarcinoma Dunning R3327," Cancer Res. **52**, 6353–6357 (1992).
[8]N. B. Smith and K. Hynynen, "The feasibility of using focused ultrasound for transmyocardial revascularization," Ultrasound Med. Biol. **24**, 1045–1054 (1998).
[9]B. C. Tran, J. Seo, T. L. Hall, J. B. Fowlkes, and C. A. Cain, "Microbubble-enhanced cavitation for noninvasive ultrasound surgery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 1296–1304 (2003).
[10]J. Debus, P. Peschke, E. W. Hahn, W. J. Lorenz, A. Lorenz, H. Ifflaender, H. J. Zabel, G. Van Kaick, and M. Pfeiler, "Treatment of the Dunning prostate rat tumor R3327-AT1 with pulsed high energy ultrasound shock waves (PHEUS): Growth delay and histomorphologic changes," J. Urol. (Baltimore) **146**, 1143–1146 (1991).
[11]A. J. Coleman, T. Kodama, M. J. Choi, T. Adams, and J. E. Saunders, "The cavitation threshold of human tissue exposed to 0.2 MHz pulsed ultrasound: Preliminary measurements based on a study of clinical lithotripsy," Ultrasound Med. Biol. **21**, 405–417 (1995).
[12]Z. Xu, A. Ludomirsky, L. Y. Eun, T. L. Hall, B. C. Tran, J. B. Fowlkes, and C. A. Cain, "Controlled ultrasound tissue erosion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 726–736 (2004).
[13]J. E. Parsons, C. A. Cain, G. D. Abrams, and J. B. Fowlkes, "Pulsed cavitational ultrasound therapy for controlled tissue homogenization," Ultrasound Med. Biol. **32**, 115–129 (2006).
[14]W. W. Roberts, T. J. Hall, K. Ives, J. J. S. Wolf, J. B. Fowlkes, and C. A. Cain, "Pulsed cavitational ultrasound: A noninvasive technology for controlled tissue ablation (histotripsy) in the rabbit kidney," J. Urol. (Baltimore) **175**, 734–738 (2006).
[15]Y. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, J. C. J. Williams, R. O. Cleveland, T. Colonius, L. A. Crum, A. P. Evan, and J. A. McAteer, "Cavitation bubble cluster activity in the breakage of kidney stones by lithotripter shockwaves," **17**, 435–446 (2003).
[16]E. A. Zabolotskaya, Y. A. Ilinskii, G. D. Meegan, and M. F. Hamilton, "Bubble interactions in clouds produced during shock wave lithotripsy," *2004 IEEE Ultrasonics Symposium*, 23–27 Aug, Montreal, Canada, pp. 890–893 BN - 890 7803 8412 7801 (2004).
[17]M. Arora, L. Junge, and C. D. Ohl, "Cavitation cluster dynamics in shock-wave lithotripsy: Part 1. Free field," Ultrasound Med. Biol. **31**, 827–839 (2005).
[18]Y. Matsumoto, "Bubble and bubble cloud dynamics," *15th International Symposium on Nonlinear Acoustics*, Gottingen, Germany, pp. 65–74 (1999).
[19]Z. Xu, J. B. Fowlkes, E. D. Rothman, A. M. Levin, and C. Cain, "Controlled ultrasound tissue erosion: The role of dynamic interaction between insonation and microbubble activity," J. Acoust. Soc. Am. **117**, 424–435 (2005).
[20]Z. Xu, J. B. Fowlkes, A. Ludomirsky, and C. A. Cain, "Investigation of intensity threshold for ultrasound tissue erosion," Ultrasound Med. Biol. **31**, 1673–1682 (2005).
[21]K. Kieran, T. L. Hall, J. E. Parsons, J. S. Wolf, J. B. Fowlkes, C. A. Cain, and W. W. Roberts, "Exploring the acoustic parameter space in ultrasound therapy: Defining the threshold for cavitational effects," *Sixth International Symposium on Therapeutic Ultrasound*, S01 (Oxford, UK, 2006).
[22]C. C. Church, "A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter," J. Acoust. Soc. Am. **86**, 215–227 (1989).
[23]A. J. Coleman, M. Whitlock, T. Leighton, and J. E. Saunders, "The spatial distribution of cavitation induced acoustic emission, sonoluminescence and cell lysis in the field of a shock wave lithotripter," Phys. Med. Biol. **38**, 1545–1560 (1993).
[24]K. Jochle, J. Debus, W. J. Lorenz, and P. Huber, "A new method of quantitative cavitation assessment in the field of a lithotripter," Ultrasound Med. Biol. **22**, 329–338 (1996).
[25]T. J. Matula, P. R. Hilmo, M. R. Bailey, and L. A. Crum, "In vitro sonoluminescence and sonochemistry studies with an electrohydraulic shock-wave lithotripter," Ultrasound Med. Biol. **28**, 1199–1207 (2002).
[26]P. Huber, J. Debus, P. Peschke, E. W. Hahn, and W. J. Lorenz, "In vivo detection of ultrasonically induced cavitation by a fibre-optic technique," Ultrasound Med. Biol. **20**, 811–825 (1994).
[27]A. J. Coleman, M. J. Choi, and J. E. Saunders, "Detection of acoustic emission from cavitation in tissue during clinical extracorporeal lithotripsy," Ultrasound Med. Biol. **22**, 1079–1087 (1996).
[28]P. Zhong, I. Cioanta, F. H. Cocks, and G. M. Preminger, "Inertial cavitation and associated acoustic emission produced during electrohydraulic shock wave lithotripsy," J. Acoust. Soc. Am. **101**, 2940–2950 (1997).
[29]W. S. Chen, T. J. Matula, and L. A. Crum, "The disappearance of ultrasound contrast bubbles: Observations of bubble dissolution and cavitation nucleation," Ultrasound Med. Biol. **28**, 793–803 (2002).
[30]P. Huber, K. Jochle, and J. Debuss, "Influence of shock wave pressure amplitude and pulse repetition frequency on the lifespan, size and number of transient cavities in the field of an electromagnetic lithotripter," Phys. Med. Biol. **43**, 3113–3128 (1998).
[31]O. A. Sapozhnikov, V. A. Khokhlova, M. R. Bailey, J. C. Williams, Jr., J. A. McAteer, R. O. Cleveland, and L. A. Crum, "Effect of overpressure and pulse repetition frequency on cavitation in shock wave lithotripsy," J. Acoust. Soc. Am. **112**, 1183–1195 (2002).
[32]J. E. Parsons, C. A. Cain, and J. B. Fowlkes, "Cost-effective assembly of a basic fiber-optic hydrophone for measurement of high-amplitude therapeutic ultrasound fields," J. Acoust. Soc. Am. **119**, 1432–1440 (2006).
[33]AIUM *Acoustic Output Measurement Standard for Diagnostic Ultrasound Equipment*, *UD2-98* (AIUM/NEMA, 1998).
[34]T. J. Matula, P. R. Hilmo, B. D. Storey, and A. J. Szeri, "Radial response of individual bubbles subjected to shock wave lithotripsy pulses in vitro," Phys. Fluids **14**, 913–921 (2002).
[35]J. F. Chen, J. A. Zagzebski, and E. L. Madsen, "Tests of backscatter coefficient measurement using broadband pulse," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 603–607 (1993).
[36]P. K. Urayama, W. Zhong, J. A. Beamish, F. K. Minn, R. D. Sloboda, K. H. Dragnev, E. Dmitrovsky, and M.-A. Mycek, "A UV-visible-NIR fluorescence lifetime imaging microscope for laser-based biological sensing with picosecond resolution," Appl. Phys. B: Lasers Opt. **B76**, 483–496 (2003).
[37]J. E. Parsons, J. B. Fowlkes, and C. A. Cain, "Acoustic backscatter features associated with production of tissue homogenate using pulsed cavitational ultrasound therapy," *International Symposium on Therapeutic Ultrasound*, 323–327 (Boston, MA, 2005).

[38]D. L. Miller, "The effects of ultrasonic activation of gas bodies in Elodea leaves during continuous pulsed irradiation at 1 MHz," Ultrasound Med. Biol. **3**, 221–240 (1977).
[39]E. L. Carstensen and H. Flynn, "The potential for transient cavitation with microsecond pulses of ultrasound," Ultrasound Med. Biol. **8**, 720–724 (1982).
[40]Z. Xu, J. B. Fowlkes, and C. A. Cain, "A new strategy to enhance cavitational tissue erosion by using a high intensity initiating sequence," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1412–1424 (2006).
[41]T. A. Pitts, J. F. Greenleaf, Y.-Y. Lu, and R. R. Kinnick, "Tomographic Schlieren imaging for measurement of beam pressure and intensity," *Proceedings of IEEE Ultrasonics Symposium* **3**, 1665–1668 (Cannes, France, 1994).

Group velocity, phase velocity, and dispersion in human calcaneus *in vivo*

Keith A. Wear[a)]
U.S. Food and Drug Administration, Center for Devices and Radiological Health, HFZ-142 12720 Twinbrook Parkway, Rockville, Maryland 20852

(Received 17 November 2006; revised 23 January 2007; accepted 23 January 2007)

Commercial bone sonometers measure broadband ultrasonic attenuation and/or speed of sound (SOS) in order to assess bone status. Phase velocity, which is usually measured in frequency domain, is a fundamental material property of bone that is related to SOS, which is usually measured in time domain. Four previous *in vitro* studies indicate that phase velocity in human cancellous bone decreases with frequency (i.e., negative dispersion). In order to investigate frequency-dependent phase velocity *in vivo*, through-transmission measurements were performed in 73 women using a GE Lunar Achilles Insight® commercial bone sonometer. Average phase velocity at 500 kHz was 1489±55 m/s (mean ± standard deviation). Average dispersion rate was −59±52 m/sMHz. Group velocity was usually lower than phase velocity, as is expected for negatively dispersive media. Using a stratified model to represent cancellous bone, the reductions in phase velocity and dispersion rate *in vivo* as opposed to *in vitro* can be explained by (1) the presence of marrow instead of water as a fluid filler, and (2) the decreased porosity of bones of living (compared with deceased) subjects. © *2007 Acoustical Society of America.*
[DOI: 10.1121/1.2697436]

PACS number(s): 43.80.Qf, 43.20.Hq [CCC] Pages: 2431–2437

I. INTRODUCTION

Commercial bone sonometers measure broadband ultrasonic attenuation (BUA) and/or speed of sound (SOS) in order to assess bone status (Laugier, 2004). Many studies document the utility of SOS for this purpose (Rossman *et al.*, 1989; Tavakoli and Evans, 1991; Zagzebski *et al.*, 1991; Schott *et al.*, 1995; Turner *et al.*, 1995; Njeh *et al.*, 1996; Hans *et al.*, 1996; Glüer *et al.*, 1996; Bouxsein and Radloff, 1997; Bauer *et al.*, 1997; Laugier *et al.*, 1997; Strelitzki and Evans, 1996; Strelitzki *et al.*, 1997; Nicholson *et al.*, 1996, 1998, Thompson *et al.*, 1998; Hans *et al.*, 1999; Trebacz, and Natali, 1999; Hoffmeister *et al.*, 2000, 2002; Chaffai *et al.*, 2002; Lee *et al.*, 2003; Glüer *et al.*, 2004; Yamoto *et al.*, 2006).

Measures of wave velocity include *phase* velocity (velocity of a single-frequency component), *group* velocity (velocity of the center of a pulse), and *signal* velocity (velocity of the front of a pulse) (Morse and Ingard, 1986). These quantities are all closely related to time-domain SOS measurements performed by commercial bone sonometers. Several investigators have reported that phase velocity usually (but not always) decreases with frequency in human calcaneus samples *in vitro* (Strelitzki and Evans, 1996; Nicholson *et al.*, 1996; Droin *et al.*, 1998; Wear, 2000). This is unusual for biologic tissue and is contrary to what one might expect based on models relating frequency-dependent attenuation and frequency-dependent phase velocity (O'Donnell *et al.*, 1981; Waters *et al.*, 2000; Mobley *et al.*, 2003). However, the so-called "restricted-bandwidth form" of the Kramers-Kronig relations has been shown to accurately predict negative dispersion in bovine cancellous bone (Waters and Hoffmeister, 2005). Negative dispersion can also be explained by interference between fast and slow longitudinal modes in cancellous bone even when each mode is positively dispersive (Marutyan *et al.*, 2006a, b).

Another model that can predict negative dispersion is the so-called "stratified model." The simplest example of a stratified medium consists of alternating parallel layers of two materials as shown in Fig. 1. Fundamental theory of stratified media was developed by Bruggeman (1935), Tarkov (1940), Riznichenko (1949), Postma (1955), and Rytov (1956). Brekhovskikh (1980) wrote a nice summary. Plona and co-workers (1987) demonstrated good agreement between theory and experiment for negative dispersion in aluminum/water and plexiglass/water stratified media. While the stratified model is based on a simplistic geometric model for cancellous bone, it can be useful for understanding the dependencies of phase velocity on various structural and material parameters of the two components: (1) the trabecular bone material, and (2) the fluid filler, which is either marrow (*in vivo*) or water (*in vitro*). The stratified model has been shown to be useful for predicting angular dependence of fast and slow compressional waves in bovine cancellous bone *in vitro* (Hughes *et al.*, 1999; Padilla and Laugier, 2000), negative dispersion in human cancellous bone *in vitro* (Wear, 2001; Lee and Yoon, 2006), and negative dispersion in cancellous-bone-mimicking phantoms (Lee, 2006).

Most publications of phase velocity and dispersion in bone report measurements performed on bone samples *in vitro*. One exception provides measurements on a single human volunteer, and shows gradual negative dispersion (Chen and Chen, 2006). The present study offers the first *in vivo* measurements of phase velocity and dispersion in a large

[a)]Electronic mail: kaw@cdrh.fda.gov

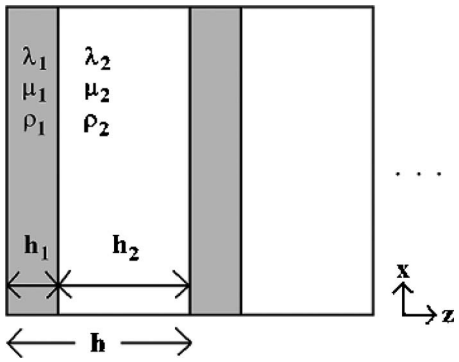


FIG. 1. The stratified model represents a composite medium as a one-dimensional structure of alternating layers of two media with density ρ and Lamé constants λ and μ.

human population (73 women). The stratified model is used to provide context for a comparison between these *in vivo* measurements and previous *in vitro* measurements.

II. METHODS

A. Data acquisition

A GE Lunar Achilles Insight® clinical bone sonometer was used for data acquisition (http://www.gehealthcare.com/usen/bone_densitometry/products/achilles.html). This system used a circular (25.4 mm diameter), broadband, piston transducer (center frequency=500 kHz) for transmission of ultrasound medio-laterally through the foot. Radio-frequency data were acquired using 52 central elements (corresponding to a 25.4-mm-diam circular area) of the Insight's 590-element two-dimensional receiver array. The element spacing was 3.175 mm. The element size was about 2 mm. The beam propagation distance was 10 cm. Data were digitized at 10 MHz. Signals from the 52 central elements were summed in order to form a single output signal. The Achilles Insight displays a real-time attenuation image in order to permit accurate positioning of the foot prior to data acquisition. Calibration spectra were obtained by performing measurements with only a temperature-controlled water path between the transmitting and receiving transducers.

In order to validate the data acquisition and analysis hardware/software, phase velocity measurements were performed on a 25.8-mm-thick polycarbonate plate that had previously been interrogated using a laboratory setup consisting of two coaxially aligned, focused 500 kHz Panametrics (Waltham, MA) 25.4-mm-diam transducers, a Panametrics 5800 pulser/receiver, and a LeCroy 9310C digitizing oscilloscope (Wear, 2000).

B. Clinical protocol

Measurements were performed in the nondominant feet of 73 ambulatory, nonpregnant women who were free of conditions that may be associated with altered bone metabolism including chronic renal disease, chronic liver disease, hyperparathyroidism, active hyperthyroidism, hypothyroidism, diabetes (type I or type II), cancer, Paget's disease, anorexia, bulimia, lactose intolerance, milk allergy, malabsorption syndrome, osteomalacia (Rickets), Vitamin D deficiency, or rheumatoid arthritis. Each subject removed shoes, socks, and/or stockings. A mist of isopropyl alcohol was sprayed on the foot in order to enhance coupling between the ultrasound beam and the foot. Table I gives demographic information. The set of volunteers included 5 African Americans, 9 Asian Americans, and 59 Caucasians. A GE Lunar PIXI DEXA calcaneal bone densitometer was used to measure areal BMD, which was reported in absolute units (g/cm^2) and also as a T-score (the number of standard deviations above or below the mean value for the normal reference population).

TABLE I. Demographic information for clinical study.

	Mean	Standard deviation	Range
Age (years)	47	13	(21, 78)
Height (in.)	64	2	(59, 68)
Weight (lb s)	136	26	(95, 220)
BMD (g/cm^2)	0.467	0.092	(0.241, 0.700)
DEXA T-score	−0.4	1.1	(−3.2, 2.8)

C. Data analysis

Phase velocity was computed using

$$c_p(\omega) = \frac{c_w}{1 + \frac{c_w \Delta\phi(\omega)}{\omega d}}, \tag{1}$$

where $\omega = 2\pi f$, and f is frequency. The calcaneal thickness, d, was assumed to be 2.5 cm, which corresponds to the average value reported by Nicholson *et al.* (1997) based on post-mortem measurements from 28 female calcanea. The speed of sound in water, c_w, was assumed to be 1525 m/s (at 37° in the Achilles temperature-controlled water reservoirs) (Kaye and Laby, 1973). The phase difference, $\Delta\phi(\omega)$, which is the difference between the phases of the water-path-only-measurement and the *in vivo* measurement, was computed as follows. The fast Fourier transform (FFT) of each digitized received signal was taken. The frequency-dependent phase of the signal at each frequency was computed from the inverse tangent of the ratio of imaginary to real parts of the FFT. Since the inverse tangent function yields principal values between $-\pi$ and π, the phase had to be unwrapped by adding an integer multiple of 2π to all frequencies above each frequency where a discontinuity appeared. Another method for dispersion measurement is the split-spectrum processing technique described by Chen and Chen (2006).

In order to provide a useful comparison for phase velocity measurements, group velocity, c_g, was also measured, using both time-domain and frequency-domain methods. For the time-domain method, the following formula was used:

$$c_g = \frac{c_w}{1 + \frac{c_w \Delta t}{d}}, \tag{2}$$

where Δt is the difference in arrival times of envelope maxima between the water-path-only-measurement and the *in vivo* measurement. In order to suppress low frequency

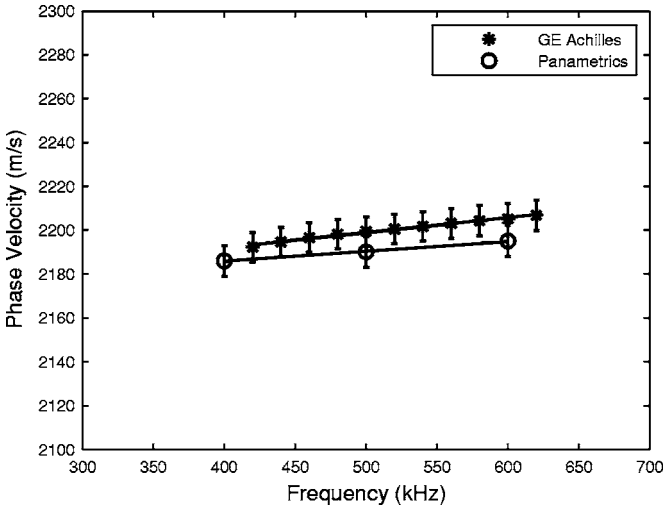


FIG. 2. Measurements of frequency-dependent phase velocity on polycarbonate plate obtained using (1) the GE Achilles Insight, and (2) Panametrics transducers in a water tank.

noise, digitized pulses were bandpass filtered (Gaussian filter with center frequency of 500 kHz and standard deviation of 200 kHz) prior to envelope detection. Signal envelopes were computed using the Hilbert transform.

For the frequency-domain method, the following formula was used (Morse and Ingard, 1986; Duck, 1990):

$$c_g = \frac{c_{pc}}{1 - \frac{\omega_c}{c_{pc}}\left(\frac{\partial c_p}{\partial \omega}\right)_{\omega=\omega_c}}, \tag{3}$$

where c_{pc} is the phase velocity at the center frequency of the pulse (500 kHz), ω_c. This method was previously employed by Strelitzki and Evans, 1996.

Substitution techniques can exhibit appreciable error if the velocity differs substantially between the sample and the reference (Kaufman *et al.*, 1995). However, this diffraction-related error has been reported to be negligible in cancellous specimens from human calcaneus *in vitro* (Droin *et al.*, 1998).

III. RESULTS

Figure 2 shows measurements of phase velocity in the polycarbonate plate obtained using (1) the GE Achilles Insight, and (2) Panametrics transducers in a water tank (Wear, 2000). The two sets of measurements are in good agreement with each other. This agreement establishes confidence in the clinical data acquisition and analysis systems.

Figure 3 compares time-domain and frequency-domain measurements of group velocity in 73 women. The two sets of measurements are in excellent agreement and reinforce confidence in the group and phase velocity measurement methods. The fact that some group velocity estimates are quite low (near 1350 m/s) may be partially due to the fixed heel width assumption (2.5 cm). Equations (1) and (2) underestimate group velocity when d is underestimated and $\Delta t > 0$.

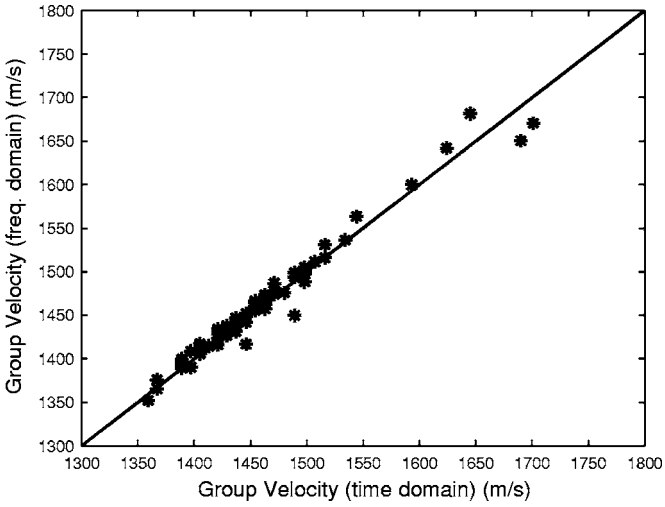


FIG. 3. Time domain vs frequency domain measurements of group velocity in 73 women.

Figure 4 compares group and phase velocity measurements in 73 women. The phase velocity measurements tend to be higher, as is expected in negatively dispersive media [see Eq. (3)].

Figure 5 shows average frequency-dependent phase velocity in 73 women. Average phase velocity at 500 kHz was 1489±55 m/s (mean ± standard deviation). Average dispersion rate (obtained from a linear least-squares regression fit) was −59±52 m/s MHz.

IV. DISCUSSION

The mean calcaneal dispersion rate of −59 m/s MHz measured here *in vivo* was more negative than those previously reported *in vitro*, which range from −15 to −40 m/s MHz (average of the four studies: −26.25 m/s MHz). See Table II. Two main differences between the present *in vivo* experiment and previous *in vitro* experiments may account for part or all of this disparity. First, pores in the cancellous bone frame *in vivo* are filled with marrow rather than water. Second, the present *in vivo* experiment was performed on living women while the pre-

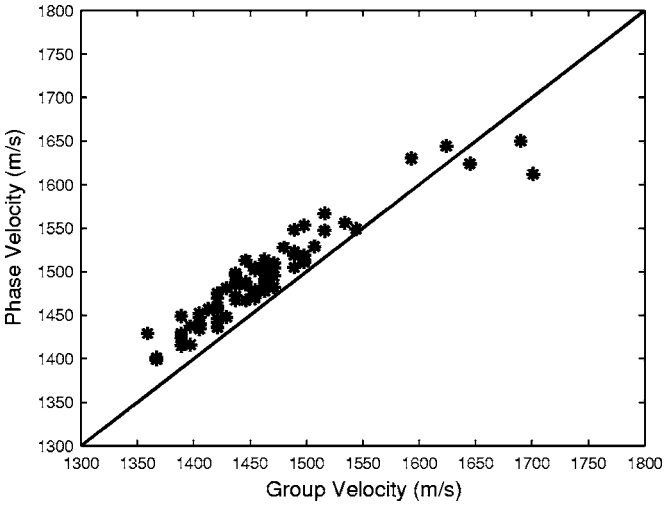


FIG. 4. Phase velocity vs group velocity in 73 women.

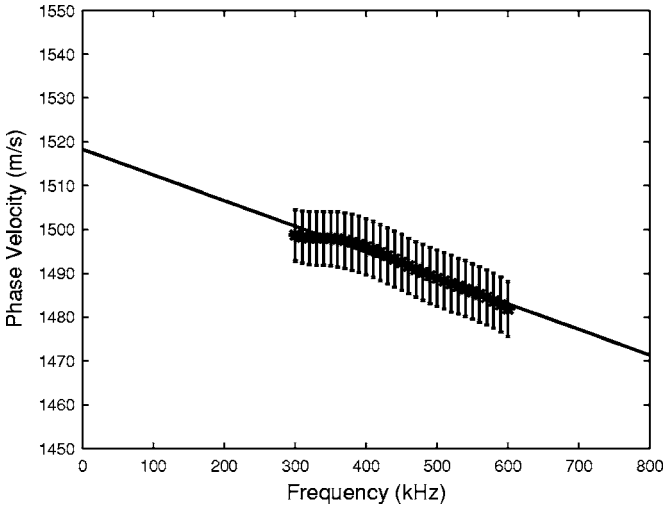


FIG. 5. Average frequency-dependent phase velocity in 73 women. Error bars denote standard errors.

vious *in vitro* experiments were performed on calcaneus samples from cadavers. The former are likely to have had lower-porosity bone.

The stratified model was employed to investigate differences in phase velocity and dispersion between the *in vivo* and *in vitro* measurements. The stratified model is illustrated in Fig. 1. Two homogeneous materials are arranged in alternating layers. Each material is characterized by its density, ρ, and its first and second Lamé constants, λ and μ. The widths of the alternating layers are denoted by h_1 and h_2. The width of the periodic unit is $h=h_1+h_2$ and is assumed to be small relative to the wavelength. The structural periodicity imposes periodic conditions on the solution to the wave equation. Thus velocities and pressures at a location z are the same as those at $z+nh$ where n is an integer. Continuity of velocities and pressures is also assumed at layer boundaries.

For plane wave propagation perpendicular to the planar interfaces between adjacent layers, the dispersion relation for the longitudinal wave is given by (Brekhovskikh, 1980)

$$\cos \xi h = \cos k_1 h_1 \cos k_2 h_2 - [(1+s)/2s]\sin k_1 h_1 \sin k_2 h_2, \tag{4}$$

where $k_1=\omega/c_1$, $k_2=\omega/c_2$,

$$s = \frac{(\lambda_2 + 2\mu_2)k_2}{(\lambda_1 + 2\mu_1)k_1} \tag{5}$$

and the phase velocity of the longitudinal wave for propagation perpendicular to the layers is given by $c_{zz}=\omega/\xi$ where $\omega=2\pi f$ and f is the frequency of the wave. The right-hand side of Eq. (4) may be computed from the structural and material properties of the two media. After taking an inverse cosine and dividing by h, ξ is obtained. Phase velocity is then computed from $c_{zz}=\omega/\xi$.

The stratified model requires assumptions for values of structural and material parameters of the fluid filler (marrow *in vivo* or water *in vitro*) and the bone trabeculae. The structural parameters (h_1 and h_2) were taken from microarchitectural analysis of 60 human calcanea (Ulrich *et al.*, 1999). The lattice spacing $h=h_1+h_2$ was taken to be 811 μm, which is the sum of the mean values for trabecular separation (Tb.Sp=684 μm) and trabecular thickness (Tb.Th =127 μm) in human calcaneus. Three values for the ratio of bone volume to total volume, BV/TV$=h_1/(h_1+h_2)$, were used: 0.083, 0.117, and 0.150. These values correspond to the mean plus or minus one standard deviation for BV/TV measured from microCT analysis of human calcaneus (Ulrich *et al.*, 1999). BV/TV=1−porosity. The layer thicknesses were obtained from $h_1=h^*$BV/TV and $h_2=h-h_1$.

The material parameters for marrow were obtained as follows. Generally speaking, marrow may contain both hematopoietic and adipose constituents. The relative proportion of adipose increases with age, however, and the conversion of marrow from hematopoietic to adipose in calcaneus is nearly complete by the age of 20 (Christy, 1981; Les *et al.*, 2002). Therefore, material properties for marrow were assigned to values for fat, which were obtained from the material properties library from WAVE 2000 PRO® software (Cyberlogic, New York).

Figure 6 shows stratified model predictions for phase velocity versus frequency in cancellous bone assuming four different sets of material parameters for the trabecular material—obtained from Grenoble *et al.* (1972), Cowin (1989), and Luo *et al.* (1999). The four different parameter sets yield similar phase velocity curves, all with negative dispersion. The stratified models underestimate phase velocity at 500 kHz, however, by about 35 m/s. The parameters from Cowin (1999) were used to generate phase velocity predictions given in the following. Table III shows values assumed for all structural and material parameters.

TABLE II. Estimates of dispersion rate in human calcaneus from Wear (2000), Nicholson *et al.* (1996, Table 1), Strelitzki and Evans (1996, Table 2), Droin *et al.* (1998, Table 1), and the present paper.

Author(s)		No.	Frequency range (kHz)	Age range (years)	Dispersion rate (mean ± standard deviation) (m/s MHz)
Nicholson *et al.*	*in vitro*	70	200–800	22–76	−40
Strelitzki and Evans	*in vitro*	10	600–800	Unknown	−32±27
Droin *et al.*	*in vitro*	15	200–600	69–89	−15±13
Wear (2000)	*in vitro*	24	200–600	Unknown	−18±15
Wear (present paper)	*in vivo*	73	300–600	21–78	−59±52

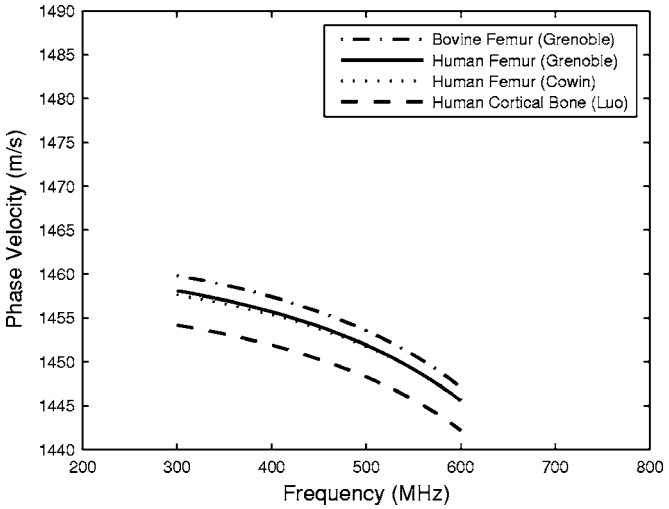


FIG. 6. Stratified model predictions for phase velocity assuming four different sets of material parameters for bone trabeculae.

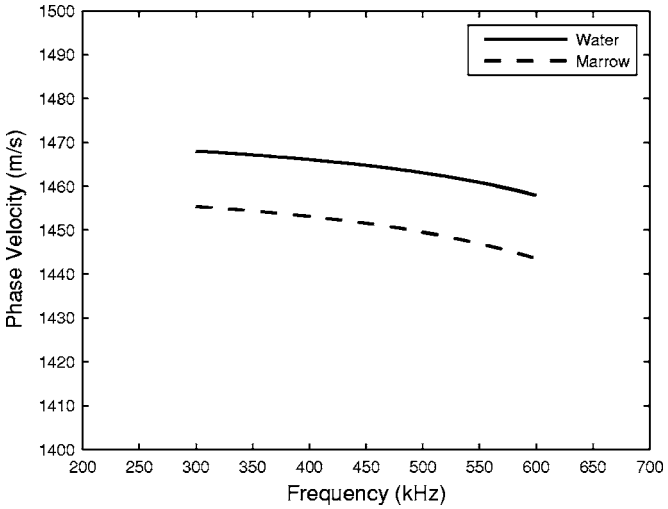


FIG. 7. Stratified-model predictions for phase velocity in cancellous bone based on two options for the fluid filler: water (*in vitro* experiments) and marrow (*in vivo* experiments).

Figure 7 shows the stratified model prediction of the effect of changing the fluid filler from water (*in vitro*) to marrow (*in vivo*). Phase velocity at 500 kHz drops by 14 m/s. This drop can explain about two-thirds of the difference in phase velocity at 500 kHz previously reported *in vitro*, 1511 m/s (Wear, 2000), with the value reported here *in vivo*, 1489 m/s. Other investigators have also measured velocity to be lower when marrow is present. Nicholson and Bouxsein (2002) observed about twice as much reduction in phase velocity at 600 kHz, 43 m/s, in human calcaneus samples *in vitro*. Alves *et al.* (1996) observed a reduction of SOS at 500 kHz of 35 m/s, while Hoffmeister *et al.* (2002) found no significant difference at 2.25 MHz, in bovine cancellous bone *in vitro*.

Figure 7 shows that dispersion rate (obtained from least-squares linear regression fits to the curves in Fig. 7) drops from −33 to −38 m/s MHz. Therefore, replacing water with marrow can be expected to reduce the dispersion rate by an amount on the order of 5 m/s MHz.

Figure 8 shows stratified model predictions for BV/TV values equal to the mean plus or minus one standard deviation for values reported by Ulrich *et al.* (1999) for human calcaneus. This analysis is relevant because it is plausible that the living human subjects in the current *in vivo* study tended to have higher BV/TV (i.e., lower porosity) than the cadaveric calcaneus specimens in the *in vitro* studies. Figure 8 shows that each reduction of BV/TV by one standard deviation is accompanied by a reduction in dispersion rate of about 25 m/s MHz. Therefore, a presumed lower porosity (i.e., higher BV/TV) of younger, living bone would help explain measurements of steeper negative dispersion. (This is consistent with the fact that Nicholson *et al.*, whose test population had an unusually high representation of young women, measured the steepest negative dispersion among the four *in vitro* studies. See Table II.)

V. CONCLUSION

This study represents the first large-scale investigation of calcaneal dispersion *in vivo*. Negative dispersion, previously reported in human calcaneal cancellous bone samples *in vitro*, is also exhibited *in vivo*. Phase velocity is lower, and dispersion rate is more negative, in calcanea from women *in vivo* than in cancellous bone samples *in vitro*. Although the stratified model tends to underestimate phase velocity at 500 kHz by about 35 m/s, it is useful for predicting dispersion, and for quantitatively explaining the differences in phase velocity and dispersion between *in vivo* and *in vitro* measurements.

TABLE III. Material and structural properties used as inputs for the stratified model. Material parameters for bone were taken from Cowin (1989). Structural parameters for bone were taken from Ulrich *et al.* (1999). Material parameters for water and marrow (i.e., fat) were taken from the material properties library WAVE 2000 PRO software (Cyberlogic, New York) and Kaye and Laby, 1973.

	Bone trabeculae	Water	Marrow
Density (kg/m^3)	1850	1000	1055
Longitudinal velocity (m/s)	3260	1482 (at 20 °C)	1479
Transverse velocity (m/s)	1644	3.5	34.5
First Lamé constant (λ) (MPa)	9700	2241	2050
Second Lamé constant (μ) (MPa)	5000	0	0
Bone volume fraction (BV/TV) (%)	11.65±3.33	⋯	⋯
Lattice spacing: $h=h_1+h_2$ (μm)	811	⋯	⋯

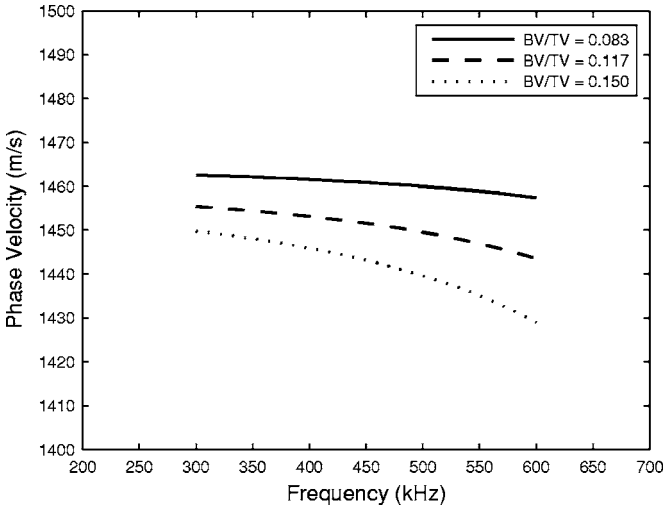


FIG. 8. Stratified-model predictions for phase velocity in cancellous bone as a function of ratio of bone volume to total volume (BV/TV). The values for BV/TV shown correspond to the mean ± one standard deviation for BV/TV reported in human calcaneus (Ulrich *et al.*, 1999).

ACKNOWLEDGMENTS

The author is grateful for funding provided by the US Food and Drug Administration Office of Women's Health. The author is grateful to Rich Morris, General Electric Corporation, Madison, WI, for providing the custom radio-frequency interface for the GE Achilles. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Department of Health and Human Services.

Alves, J. M., Ryaby, J. T., Kaufman, J. J., Magee, F. P., and Siffert, R. S. (**1996**). "Influence of marrow on ultrasonic velocity and attenuation in bovine trabecular bone," Calcif. Tissue Int. **58**, 362–367.

Bauer, D. C., Gluer, C. C., Cauley, J. A., Vogt, T. M., Ensrud, K. E., Genant, H. K., and Black, D. M. (**1997**). "Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women," Arch. Intern Med. **157**, 629–634.

Bouxsein, M. L., and Radloff, S. E. (**1997**). "Quantitative ultrasound of the calcaneus reflects of the mechanical properties of calcaneal trabecular bone," J. Bone Miner. Res. **12**, 839–846.

Brekhovskikh, L. M. (**1980**). *Waves in Layered Media* (Academic, New York), p. 81.

Bruggeman, D. A. G. (**1935**). "Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen ("Calculation of physical constants from heterogeneous substances")," Ann. Phys. **24**, 636-664.

Chaffai, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**). "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone (N.Y.), **30**, 229–237.

Chen, P., and Chen, T. (**2006**). "Measurements of acoustic dispersion on calcaneus using split spectrum processing technique," Med. Eng. Phys. **28**, 187–193.

Christy, M. (**1981**). "Active bone marrow distribution as a function of age in humans," Phys. Med. Biol. **26**, 389–400.

Cowin, S. C. (**1989**). "The mechanical properties of cortical bone," in *Bone Mechanics*, edited by S. C. Cowin (CRC Press, Boca Raton, FL), pp. 97–127.

Droin, P., Berger, G., and Laugier, P. (**1998**). "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592.

Duck, F. A. (**1990**). *Physical properties of tissue* (University Press, Cambridge, UK, **1990**).

Glüer, C. C., Cummings, S. R., Bauer, D. C., Stone, K., Pressman, A., Mathur, A., and Genant, H. K. (**1996**). "Osteoporosis: Association of recent fractures with quantitative US findings," Radiology, **199**, 725–732.

Glüer, C. C., *et al.* (**2004**). "Association of five quantitative ultrasound devices and bone densitometry with osteoporotic vertebral fractures in a population-based sample: The OPUS study," J. Bone Miner. Res. **19**, 782–793.

Grenoble, D. E., Katz, J. L., Dunn, K. L., Gilmore, R. S., and Lingamurty, K. (**1972**). "The elastic properties of hard tissues and apatites," J. Biomed. Mater. Res. **6**, 221–233.

Hans, D., Dargent-Molina, P., Schott, A. M., Sebert, J. L., Cormier, C., Kotzki, P. O., Delmas, P. D., Pouilles, J. M., Breart, G., and Meunier, P. J. (**1996**). "Ultrasonographic heel measurements to predict hip fracture in elderly women: The EPIDOS prospective study," Lancet **348**, 511–514.

Hans, D., Wu, C., Njeh, C. F., Zhao, S., Augat, P., Newitt, D., Link, T., Lu, Y., Majumdar, S., and Genant, H. K. (**1999**). "Ultrasound velocity of cancellous cubes reflects mainly bone density and elasticity," Calcif. Tissue Int. **64**, 18–23.

Hoffmeister, B. K., Whitten, S. A., and Rho, J. Y. (**2000**). "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone (N.Y.) **26**, 635–642.

Hoffmeister, B. K., Whitten, S. A., Kaste, S. C., and Rho, J. Y. (**2002**). "Effect of collagen and mineral content on the high-frequency ultrasonic properties of human cancellous bone," **13**, 26–32.

Hoffmeister, B. K., Auwarter, J. A., and Rho, J. Y. (**2002**). "Effect of marrow on the high frequency ultrasonic properties of cancellous bone," Phys. Med. Biol. **47**, 3419–3427.

Hughes, E. R., Leighton, T. G., Petley, G. W., and White, P. R. (**1999**). "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 881–821.

Kaye, G. W. C., and Laby, T. H. (**1973**). *Table of Physical and Chemical Constants* (Longman, London).

Kaufman, J. J., Xu, W., Chiabrera, A. E., and Siffert, R. S. (**1995**). "Diffraction effects in insertion mode estimation of ultrasonic group velocity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 232–242.

Laugier, P., Droin, P., Laval-Jeantet, A. M., and Berger, G. (**1997**). "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165.

Laugier, P. (**2004**). "An overview of bone sonometry," International Congress Series **1274**, 23–32.

Lee, K. I., Roh, H., and Yoon, S. W. (**2003**). "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293.

Lee, K. I., and Yoon, S. W. (**2006**). "Predictions of the stratified model for the dependence of phase velocity on microarchitectural parameters and porosity in trabecular bone," Proceedings WESPAC IX, Seoul, South Korea.

Les, C. M., Whalen, R. T., Beaupre, G. S., Yan, C. H., Cleek, T. M., and Wills, J. S. (**2002**). "The x-ray attenuation characteristics and density of human calcaneal marrow do not change significantly during adulthood," J. Orthop. Res. **20**, 633–641.

Luo, G., Kaufman, J. J., Chiabrera, A. E., Bianco, B., Kinney, J. H., Haupt, D., Ryaby, J. T., and Siffert, R. S. (**1999**). "Computational methods for ultrasonic bone assessment," Ultrasound Med. Biol. **25**, 823–830.

Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2006b**). "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–EL61.

Marutyan, K. R., Bretthorst, G. L., and Miller, J. G. (**2006a**). "Bayesian estimation of the underlying bone properties from mixed fast and slow mode ultrasonic signals," J. Acoust. Soc. Am. **121**, EL8–EL15.

Mobley, J., Waters, K. R., and Miller, J. G. (**2003**). "Finite-bandwidth effects on the causal prediction of ultrasonic attenuation of the power-law form," J. Acoust. Soc. Am. **114**, 2782–2790.

Morse, P. M., and Ingard, K. U. (**1986**). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ), Chap. 9.

Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and Van der Perre, G. (**1996**). "Comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurements in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.

Nicholson, P. H. F., Lowet, G., Cheng, X. G., Boonen, S., Van der Perre, G., and Dequeker, J. (**1997**). "Assessment of the strength of the proximal femur in vitro: Relationship with ultrasonic measurements of the calcaneus," Bone (N.Y.) **20**, 219–224.

Nicholson, P. H. F., Muller, R., Lowet, G., Cheng, X. G., Hildebrand, T.,

Ruegsegger, P., Van Der Perre, G., Dequeker, J., and Boonen, S. (**1998**). "Do quantitative ultrasound measurements reflect structure independently of density in human vertebral cancellous bone?," Bone (N.Y.) **23**, 425–431.

Nicholson, P. H. F., and Bouxsein, M. L. (**2002**). "Bone marrow influences quantitative ultrasound measurements in human cancellous bone," Ultrasound Med. Biol. **28**, 369–375.

Njeh, C. F., Hodgskinson, R., Currey, J. D., and Langton, C. M. (**1996**). "Orthogonal relationships between ultrasonic velocity and material properties of bovine cancellous bone," Med. Eng. Phys. **18**, 373–381.

O'Donnell, M., Jaynes, E. T., and Miller, J. G. (**1981**). "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–700.

Padilla, F., and Laugier, P. (**2000**). "Phase and group velocities of fast and slow compressional waves in trabecular bone," J. Acoust. Soc. Am. **108**, 1949–1952.

Plona, T. J., Winkler, K. W., and Schoenberg, M. (**1987**). "Acoustic waves in alternating fluid/solid layers," J. Acoust. Soc. Am. **81**, 1227–1234.

Postma, G. W. (**1955**). "Wave propagation in a stratified medium," Geophysics **20**, 80.

Rossman, P., Zagzebski, J., Mesina, C., Sorenson, J., and Mazess, R. (**1989**). "Comparison of speed of sound and ultrasound attenuation in the os calcis to bone density of the radius, femur and lumbar spine," Clin. Phys. Physiol. Meas. **10**, 353–360.

Riznichenko, Y. V. (**1949**). "Propagation of seismic waves in discrete and heterogeneous media," Izv. Akad. Nauk SSSR, Ser. Geogr. Geofiz. **13**, 115.

Rytov, S. M. (**1956**). "Acoustical properties of a finely layered medium," Akust. Zh. **2**, 71; Sov. Phys. Acoust. **2**, 67-80.

Schott, M., Weill-Engerer, S., Hans, D., Duboeuf, F., Delmas, P. D., and Meunier, P. J. (**1995**). "Ultrasound discriminates patients with hip fracture equally well as dual energy X-ray absorptiometry and independently of bone mineral density," J. Bone Miner. Res. **10**, 243–249.

Strelitzki, R., and Evans, J. A. (**1996**). "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.

Strelitzki, R., Evans, J. A., and Clarke, A. J. (**1997**). "The influence of porosity and pore size on the ultrasonic properties of bone investigated using a phantom material," Osteoporosis Int. **7**, 370–375.

Tarkov, A. G. (**1940**). "The problem of the anisotropy of elastic properties in rocks," Mater. Vses. N.-I. Geol. In-ta Obshch. Seriya. Sb. **5**, 209.

Tavakoli, M. B., and Evans, J. A. (**1991**). "Dependence of the velocity and attenuation of ultrasound in bone on the mineral content," Phys. Med. Biol. **36**, 1529–1537.

Thompson, P., Taylor, J., Fisher, A., and Oliver, R. (**1998**). "Quantitative heel ultrasound in 3180 women between 45 and 75 years of age: Compliance, normal ranges and relationship to fracture history," Osteoporosis Int. **8**, 211–214.

Trebacz, H., and Natali, A. (**1999**). "Ultrasound velocity and attenuation in cancellous bone samples from lumbar vertebra and calcaneus," Osteoporosis Int. **9**, 99–105.

Turner, H., Peacock, M., Timmerman, L., Neal, J. M., and Johnston, C. C. Jr. (**1995**). "Calcaneal ultrasonic measurements discriminate hip fracture independently of bone mass," Osteoporosis Int. **5**, 130–135.

Ulrich, D., van Rietbergen, B., Laib, A., and Rüegsegger, P. (**1999**). "The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone," Bone (N.Y.) **25**, 55–60.

Waters, K. R., Hughes, M. S., Mobley, J., Brandenburger, G. H., and Miller, J. G. (**2000**). "On the applicability of Kramers-Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563.

Waters, K. R., and Hoffmeister, B. K. (**2005**). "Kramers-Kronig analysis of attenuation and dispersion in trabecular bone," J. Acoust. Soc. Am. **118**, 3912–3920.

Wear, K. A. (**2000**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.

Wear, K. A. (**2001**). "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083.

Yamoto, Y., Matsukawa, M., Otani, T., Yamazaki, K., and Nagano, A. (**2006**). "Distribution of longitudinal wave properties in bovine cortical bone in vitro," Ultrasonics **44**, e233–e237.

Zagzebski, J. A., Rossman, P. J., Mesina, C., Mazess, R. B., and Madsen, E. L. (**1991**). "Ultrasound transmission measurements through the os calcis," Calcif. Tissue Int. **49**, 107–111.

Shear wave speed recovery using moving interference patterns obtained in sonoelastography experiments

Joyce McLaughlin and Daniel Renzi[a)]
Mathematics Department, Rensselaer Polytechnic Institute, Troy, New York 12180

Kevin Parker
ECE Department, University of Rochester, Rochester, New York 14627

Zhe Wu
GE Healthcare, Milwaukee, Wisconsin 53201

(Received 17 November 2005; revised 4 December 2006; accepted 6 January 2007)

Two new experiments were created to characterize the elasticity of soft tissue using sonoelastography. In both experiments the spectral variance image displayed on a GE LOGIC 700 ultrasound machine shows a moving interference pattern that travels at a very small fraction of the shear wave speed. The goal of this paper is to devise and test algorithms to calculate the speed of the moving interference pattern using the arrival times of these same patterns. A geometric optics expansion is used to obtain Eikonal equations relating the moving interference pattern arrival times to the moving interference pattern speed and then to the shear wave speed. A cross-correlation procedure is employed to find the arrival times; and an inverse Eikonal solver called the *level curve method* computes the speed of the interference pattern. The algorithm is tested on data from a phantom experiment performed at the University of Rochester Center for Biomedical Ultrasound.
 [DOI: 10.1121/1.2534717]

PACS number(s): 43.80.Qf, 43.20.Jr [FD] Pages: 2438–2446

I. INTRODUCTION

The target in this paper is to produce an image of tissue where the imaging functional is a measure of shear stiffness. This problem has been addressed for over 10 years and is motivated by the fact that shear stiffness is the tissue elastic property that is felt in a palpation exam. Three types of experiments have emerged.

(a) Static experiment: The tissue is compressed (Konofagou *et al.*, 1998, 2000a, b; Konofagou, 2000; Ophir *et al.* 1991).
(b) Transient experiments: (1) A wave is initiated with a line source on the boundary, (Bercoff *et al.*, 2001; Catheline *et al.*, 1999; Gennisson *et al.* 2003; Sandrin *et al.*, 2001, 2002a, b; Tanter *et al.* 2003), or in the interior (Bercoff *et al.*, 2002, 2004), and a wave with a front propagates away from the source; (2) a wave is initiated at a point (Nightingale *et al.* 2002, 2003), and propagates away from the source; and (3) a traveling wave is produced by harmonic excitation at two different points, each excited at two different but nearby frequencies (Wu *et al.*, 2004, 2006).
(c) Dynamic excitation: (1) A time harmonic excitation made on the boundary creates a time harmonic wave in the tissue (Lerner *et al.*, 1988; Gao *et al.*, 1995; Levinson and Sata, 1995; Taylor *et al.*, 2000; Manduca *et al.*, 2001; Wu *et al.*, 2002); and (2) a time harmonic excitation in the interior (Greenleaf and Fatemi, 1998) creates a time harmonic radiating wave.

For most of these experiments interior displacement on a fine grid of points in an imaging plane is measured with ultrasound or magnetic resonance and the excitation is low frequency (50–200 Hz); in Greenleaf and Fatemi (1998). an interior point source excitation at a few kilohertz yields a radiating wave which is measured on the surface of the body.

In McLaughlin and Renzi (2006a, b) the authors developed the Arrival Time algorithm for the transient elastography experiment developed in the laboratory of Fink (Catheline *et al.*, 1999; Sandrin *et al.*, 2002a, b). Important features in this work are: (1) A line source, with central frequency (50–200 Hz), initiates a shear wave with a front propagating in the interior; (2) the ultrafast imaging system developed by Fink *et al.* has a frame rate of up to 10 000 frames/s enabling identification of the wave front and its arrival time on a sufficiently fine grid in the image plane; and (3) the Arrival Time algorithm recovers the shear stiffness in the imaging plane from the space/time position of the wave front.

In this paper we focus on the application of the Arrival Time algorithm to image shear stiffness using data from two new sonoelasticity experiments developed by Wu and Parker at the University of Rochester. The key feature of these experiments is that the display of the Doppler spectral variance on a GE LOGIC 700 Doppler ultrasound machine shows a very slow moving traveling wave.

In the crawling wave experiment (Wu *et al.* 2004, 2006), two time harmonic excitations at nearby but not equal frequencies, are created on opposite sides of the tissue. In the holographic wave experiment (Wu *et al.* 2006), one time harmonic excitation is made in the tissue, and a second os-

[a)]Author to whom correspondence should be addressed; Electronic mail: renzid@rpi.edu

cillation at a nearby but not equal frequency is made in the ultrasound transducer where a gel is applied so that no resultant wave propagates into the tissue.

The common features of the two experiments are: (1) The GE LOGIC 700 display shows the radial component of a slowly moving interference pattern which would be stationary if the frequencies were the same; (2) the speed of the interference pattern is a small fraction of the shear wave speed; and (3) a Doppler ultrasound scanner samples the very slowly moving interference pattern effectively at a frame rate similar to the ultrafast imaging system frame rate for a wave moving at the shear wave speed. The main differences of the two experiments are: (1) The crawling wave interference display is governed by the sum of the two waves generated by the sources; and the holographic wave display is governed by the relative motion of the tissue oscillation and the transducer oscillation; (2) the interference pattern for the crawling wave has nearly parallel interference maxima; and for the holographic wave the interference patterns are more circular; (3) the amplitude of the holographic wave decreases with distance from the source; and the amplitude of the crawling wave is more uniform; (4) for the holographic wave the speed of the moving interference pattern is directly proportional to the shear wave speed; and for the crawling wave the relationship between the speed of the interference pattern and the shear wave speed is quite complicated in inhomogeneous regions; see Sec. IV; and (5) an advantage of the holographic wave experiment is that the tissue need only be accessible for excitation in one location.

To create our images we treat a stripe in a moving interference pattern as a wave packet and then: (a) Identify the arrival time of the wave packet at each point in the image plane; (b) find the Eikonal equation satisfied by this arrival time; and (c) apply the *level curve method* for the Arrival Time algorithm (McLaughlin and Renzi, 2006b) to the Eikonal equation to solve an inverse problem and image stiffness. Prior to this work, Wu *et al.* (2004) used the distance between interference maxima in the crawling wave experiment in a homogeneous phantom to determine the constant speed. This method, called the low frequency estimation method, is also applied to both experiments performed on a phantom with inclusion in Wu *et al.* (2006). In the current paper, the image of the inclusion is significantly improved when the image is created with the Arrival Time algorithm. Furthermore, the Arrival Time algorithm is fully two dimensional, taking into consideration the depth and transverse spreading of the lines of constant phase.

The rest of this paper is composed as follows: The experimental setups for each experiment are illustrated in Sec. II. Next, the mathematical model for each experiment is given in Sec. III. The equations relating the moving interference pattern speed and the shear wave speed are derived in Sec. IV. Two subalgorithms needed to create the shear stiffness images are explained in Secs. V and VI. Section VII contains reconstructions of a heterogeneous phantom with data from both experiments. Concluding remarks are given in Sec. VIII.

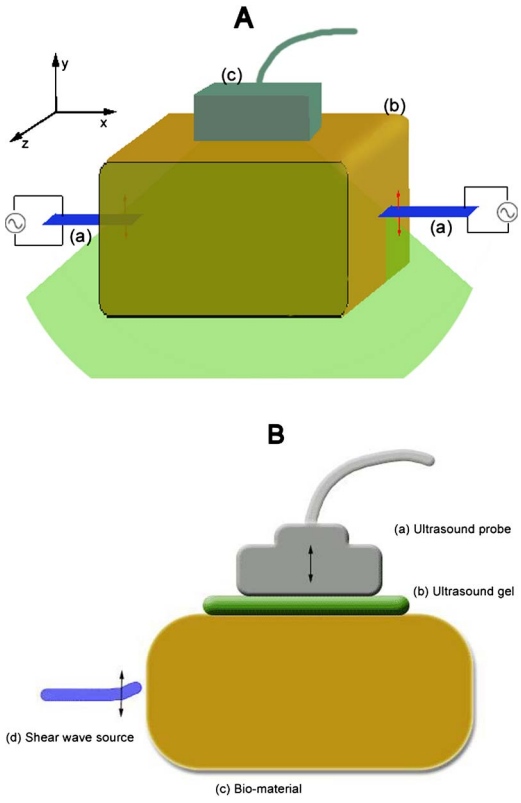


FIG. 1. (Color online) (A) Setup of the crawling wave experiment. Here the shear wave vibration sources are at (a), and the biomaterial and ultrasound probe are at (b) and (c) respectively. (B) Setup of the holographic wave experiment. Figure taken from Wu *et al.* (2006).

II. EXPERIMENTAL SETUPS AND THE DATA

The experimental apparatus for the crawling and holographic wave experiments are shown in Figs. 1(A) and 1(B). The GE LOGIC 700 Doppler ultrasound machine presents a 4 bit display of the Doppler spectral variance which is a good estimation of vibration amplitude (Wu *et al.*, 2004) and therefore displays a wave slowly moving across the screen. The Doppler spectral variance is the square of the sum of the excitations in the crawling wave experiment or the square of the difference in the excitations in the holographic wave experiment under some assumptions explained in the following.

For the crawling wave experiment we represent the radial components of the wave from each source as $A_r \sin(\omega_1 t - \omega_1 \phi_1)$, $B_r \sin(\omega_2 t + \omega_2 \phi_2)$. The backscattered ultrasound signal (see Huang *et al.*, 1990) can be represented as

$$s(t) = \cos\left(\omega_0 t + \frac{4\pi}{\lambda_0}(A_r \sin(\omega_1(t-\phi_1)) + B_r \sin(\omega_2(t+\phi_2)))\right)$$

where ω_0 is the ultrasound frequency and λ_0 is the wavelength of the ultrasound wave. Using basic trigonometry we can rewrite $s(t)$ as

$$s(t) = \cos\Big(\omega_0 t + \frac{4\pi}{\lambda_0} G \sin(t(\omega_1+\omega_2)/2 - (\omega_1\phi_1 - \omega_2\phi_2)/2 + \alpha)\Big),$$

where

$$G = [(A_r + B_r)^2 \cos^2(t(\omega_1-\omega_2)/2 - (\omega_1\phi_1 + \omega_2\phi_2)/2) + (A_r - B_r)^2 \sin^2(t(\omega_1-\omega_2)/2 - (\omega_1\phi_1+\omega_2\phi_2)/2)]^{1/2}$$

and

$$\alpha = \tan^{-1}\left[\left(\frac{A_r - B_r}{A_r + B_r}\right)\tan(t(\omega_1-\omega_2)/2 - (\omega_1\phi_1+\omega_2\phi_2)/2)\right].$$

In our examples $\omega_1-\omega_2 \sim 0.1$ Hz and $\omega_1+\omega_2 \sim 400$ Hz, so G and α are slowly varying and can be considered as stationary during the time interval for calculating the spectral variance. The Doppler spectral variance is then proportional to G^2 which is, again using basic trigonometry, with $\psi(\boldsymbol{x},t) = t(\omega_1-\omega_2) - (\omega_1\phi_1+\omega_2\phi_2)$,

$$G^2 = A_r^2 + B_r^2 + 2A_rB_r \cos\psi(\boldsymbol{x},t). \tag{1}$$

For the holographic wave experiment let $B\sin(\omega_2 t)$ be the vibration of the ultrasound transducers; the backscattered signal is the vibration of the tissue relative to the vibration in the transducer, so

$$s(t) = \cos\Big(\omega_0 t + \frac{4\pi}{\lambda_0}(A_r \sin(\omega_1 t - \phi_1) - B\sin(\omega_2 t))\Big) = \cos\Big(\omega_0 t + \frac{4\pi}{\lambda_0}\tilde{G}\sin(t(\omega_1+\omega_2)/2 - \omega_1\phi_1/2 + \tilde{\alpha})\Big)$$

where

$$\tilde{\alpha} = \tan^{-1}\left[\frac{A_r + B}{A_r - B}\tan^{-1}(t(\omega_1-\omega_2)/2 - \omega_1\phi_1/2)\right].$$

Here the displayed quantity is

$$\tilde{G}^2 = A_r^2 + B^2 - 2A_rB\cos\psi(\boldsymbol{x},t), \tag{2}$$

where now $\phi_2 = 0$ and so $\psi(x,t) = t(\omega_1-\omega_2) - \omega_1\phi_1$.

Both G^2 and $\tilde{G}^2$ can be rewritten as the amplitude of the sum (or difference) of the complexification of the induced vibrations

$$G^2 = |A_r e^{i\omega_1(t-\phi_1)} + B_r e^{i\omega_2(t-\phi_2)}|^2,$$

$$\tilde{G}^2 = |A_r e^{i\omega_1(t-\phi_1)} - Be^{i\omega_2 t}|^2.$$

These are the identities presented in Wu *et al.* (2006). Here we explain the assumptions under which they are obtained from the spectral variance.

III. MATHEMATICAL MODEL

Because stiffness is an elastic property and the displacements generated from the vibrators are small (on the order of microns), we use the linear elastic system of differential equations as our mathematical model. Assuming also that the medium is isotropic, the vector elastic displacement, $\boldsymbol{u}$, is then governed by the following system of equations:

$$(\lambda u_{j,j})_{,i} + (\mu(u_{i,j} + u_{j,i}))_{,j} - \rho u_{i,tt} = 0, \tag{3}$$

where λ, μ are the Lamé parameters and ρ is the density.

IV. EQUATIONS FOR THE IMAGING FUNCTIONALS

In the crawling wave experiments, the displacements $\boldsymbol{u}^1, \boldsymbol{u}^2$, from the first and second vibration sources, respectively, and $\boldsymbol{u} = \boldsymbol{u}^1 + \boldsymbol{u}^2$ all satisfy Eq. (3). The goal here is to find an equation for the phases for each of these quantities and then to derive a relationship between the phase $\psi(\boldsymbol{x},t) = t(\omega_1-\omega_2) - (\omega_1\phi_1+\omega_2\phi_2)$, or $\psi(\boldsymbol{x},t) = t(\omega_1-\omega_2) - \omega_1\phi_1$, seen in expressions (1) and (2), respectively, and the shear wave speed.

To accomplish this goal we first assume that $\boldsymbol{u}, \boldsymbol{u}_1, \boldsymbol{u}_2$ represent the complexification of the corresponding displacements. We use the geometric optics approximation (Ji *et al.* 2003; Ji and McLaughlin, 2004) for $\boldsymbol{u}^1$,

$$\boldsymbol{u}^1(\boldsymbol{x},t) = \boldsymbol{A}e^{i\omega_1(t-\phi_1)}, \tag{4}$$

where $\boldsymbol{A}$ is represented by the asymptotic expansion, $\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{A}_1/(i\omega_1) + \boldsymbol{A}_2/(i\omega_1)^2 + \cdots$. Substituting this expansion into Eq. (3), writing the left-hand side of Eq. (3) in powers of ω_1, and setting the coefficient of the highest order terms of ω_1 equal to zero results in (see Ji and McLaughlin, 2004)

$$0 = M\boldsymbol{A}_0, \tag{5}$$

where M is the following matrix:

$$M = [(\lambda+\mu)\nabla\phi_1(\nabla\phi_1)^T + (\mu|\nabla\phi_1|^2 - \rho)I]. \tag{6}$$

The assumption here is that there is enough separation of scales so that the coefficient of each power of ω_1 is separately equal to zero. For Eq. (5) to have a solution, the matrix M must be singular. Setting the determinant of M equal to zero yields that either

$$|\nabla\phi_1(\boldsymbol{x})| = \sqrt{\rho/\mu} = 1/C_s, \tag{7}$$

or

$$|\nabla\phi_1(\boldsymbol{x})| = \sqrt{\rho/(\lambda+2\mu)} = 1/C_p, \tag{8}$$

where C_s and C_p are the shear and compression wave speeds, respectively. Equations (7) and (8) are called Eikonal equations. In soft tissue, λ is several orders of magnitude greater than μ (Sarvazyan *et al.*, 1995). Furthermore, for the constant coefficient case in an elastic half space, the exact solution of Eq. (3) has been found in Miller and Pursey (1954) and from this solution it is clear that the amplitude of the compression wave is very small, $O((\mu/\lambda)^2)$, when the ratio λ/μ is large. For this reason we will assume that Eq. (7) is satisfied. Likewise, we write the displacement, $\boldsymbol{u}^2$, from the second source as

$$\boldsymbol{u}^2(\boldsymbol{x},t) = \boldsymbol{B}e^{i\omega_2(t-\phi_2(t))} \tag{9}$$

and as above, the phase, ϕ_2, satisfies

$$|\nabla\phi_2(\boldsymbol{x})| = \sqrt{\rho/\mu} = 1/C_s. \tag{10}$$

To obtain the equation for ψ we use the idea that ψ and the speed, F, of the moving interference pattern in the direction $-\nabla\psi$ satisfy the Eikonal equation $|\nabla\psi(\boldsymbol{x},t)|F(\boldsymbol{x})=\psi_t$. (See Osher and Sethian (1988); Sethian (1999); Osher and Fedkiw (2002), and Appendix A.)

In addition, because ψ_t is the constant $\Delta\omega=\omega_1-\omega_2$, only the spatially varying component of the phase, which is $\hat{\psi}(\boldsymbol{x})=\omega_1\phi_1+\omega_2\phi_2$ for the crawling wave and $\hat{\psi}(\boldsymbol{x})=\omega_1\phi_1$ for the holographic wave is present in the formula for the speed, F. So

$$F=\frac{\omega_1-\omega_2}{|\nabla\hat{\psi}(\boldsymbol{x})|}=\frac{\Delta\omega}{|\nabla\hat{\psi}(\boldsymbol{x})|}. \tag{11}$$

We use this equation and the Arrival Time algorithm to find F and use $\omega_1F/\Delta\omega$ and $2\omega_1F/\Delta\omega$ as our imaging functionals for the holographic wave and crawling wave experiments, respectively, as we explain below.

The speed F is a simple multiple of the shear wave speed in the holographic wave experiment but not for the crawling wave experiment. To show this we first calculate $|\nabla\hat{\psi}(\boldsymbol{x})|^2$ as

$$\begin{aligned}|\nabla\hat{\psi}(\boldsymbol{x})|^2&=|\omega_1\nabla\phi_1(\boldsymbol{x})+\omega_2\nabla\phi_2(\boldsymbol{x})|^2\\&=\omega_1^2|\nabla\phi_1(\boldsymbol{x})|^2+\omega_2^2|\nabla\phi_2(\boldsymbol{x})|^2\\&\quad+2\omega_1\omega_2\nabla\phi_1(\boldsymbol{x})\cdot\nabla\phi_2(\boldsymbol{x})\\&=\omega_1^2|\nabla\phi_1(\boldsymbol{x})|^2+\omega_2^2|\nabla\phi_2(\boldsymbol{x})|^2\\&\quad+2\omega_1\omega_2|\nabla\phi_1(\boldsymbol{x})||\nabla\phi_2(\boldsymbol{x})|\cos(\theta),\end{aligned} \tag{12}$$

where θ is the angle between $\nabla\phi_1(\boldsymbol{x})$ and $\nabla\phi_2(\boldsymbol{x})$. Now, for the holographic wave experiment $\phi_2=0$ and

$$\begin{aligned}F^2&=\frac{\Delta\omega^2}{|\nabla\hat{\psi}|^2}=\frac{\psi_t^2(\boldsymbol{x},t)}{|\nabla\psi(\boldsymbol{x},t)|^2}=\frac{\Delta\omega^2C_s^2}{(\omega_1^2)}\\&\Rightarrow F=\frac{\Delta\omega C_s}{\omega_1}\text{ or }C_s=\frac{\omega_1F}{\Delta\omega}.\end{aligned} \tag{13}$$

For the crawling wave experiment the relationship between the crawling wave speed, F, and the shear wave speed, C_s, is more complicated. Therefore, using Eqs. (7), (10), and (12) we have

$$|\nabla\hat{\psi}(\boldsymbol{x})|^2=(1/C_s^2)(\omega_1^2+\omega_2^2+2\omega_1\omega_2\cos(\theta)). \tag{14}$$

Calculating the ratio $\psi_t^2/|\nabla\psi(\boldsymbol{x},t)|^2$ we obtain

$$\begin{aligned}F^2&=\frac{\Delta\omega^2}{|\nabla\hat{\psi}(\boldsymbol{x})|^2}=\frac{\psi_t^2}{|\nabla\psi|^2}\\&=\frac{\Delta\omega^2C_s^2}{2\omega_1^2(1+\cos(\theta))+O(\omega_1\Delta\omega_1)+O(\Delta\omega^2)}\end{aligned} \tag{15}$$

$$=\frac{\Delta\omega^2C_s^2}{4\omega_1^2\cos^2(\theta/2)+O(\omega_1\Delta\omega_1)+O(\Delta\omega^2)}. \tag{16}$$

This equation cannot be used to directly find the wave speed, $C_s(\boldsymbol{x})$, from the phase, $\psi(\boldsymbol{x},t)$, because in the inhomogeneous medium case the $\cos(\theta/2)$ term depends on the unknowns $\phi_1(\boldsymbol{x})$ and $\phi_2(\boldsymbol{x})$. However, Eqs. (7), (10), and (15) are a coupled system of three equations that can be solved for C_s, $\phi_1(\boldsymbol{x})$, and $\phi_2(\boldsymbol{x})$. This will be the subject of a future paper. So here for the crawling wave experiment we do not image C_s but instead use the quantity $2\omega_1F/\Delta\omega$ as an imaging functional. Note that similar expressions relating phase wave speed to C_s are found in Wu *et al.* (2006), under a locally constant assumption and without showing the relationship to the underlying elastic equations.

Remark 1: We measure only one component of the vibration amplitude. This is not a restriction since all components of the vibrational amplitude have the same phase under the geometric optics assumption.

Remark 2: The vibration amplitude is measured in a plane. So, out-of-plane derivatives cannot be calculated and are assumed to be zero. If there is significant out-of-plane motion of the moving interference pattern, this assumption causes overestimation of the imaging functional.

Remark 3: For the holographic experiment the direction of propagation of the shear wave is $\nabla\hat{\psi}=\omega_1\nabla\phi_1/\Delta\omega$; so the moving interference pattern moves either in the same (if $\omega_1>\omega_2$) or directly opposite direction (if $\omega_1<\omega_2$) as the shear wave induced by the source vibrating in the phantom (or tissue). For the crawling wave experiment $\nabla\hat{\psi}=(\omega_1\nabla\phi_1+\omega_2\nabla\phi_2)/\Delta\omega$; so the direction of the moving interference pattern is not, in general, in the same direction as either of the shear waves induced individually by the two sources.

V. CALCULATING PHASE AND ARRIVAL TIME

To utilize $|\nabla\hat{\psi}|F=\Delta\omega$ we must first construct a continuously varying phase, $\hat{\psi}$, from the data, $|u_r|^2=A_r^2+B_r^2+2A_rB_r\cos(\Delta\omega t-\hat{\psi})$. This is related to the classic phase unwrapping problem.

Furthermore, we can interpret a multiple of $\hat{\psi}$ as an arrival time. This is based on the observation that at an arbitrary fixed point, $\boldsymbol{x}_0$, the time trace of the data can be represented by a constant plus $2(A_rB_r)(\boldsymbol{x})\cos((\omega_2-\omega_1)t-\hat{\psi}(\boldsymbol{x}))$. After filtering out the constant, consider the time trace $V(\boldsymbol{x}_0,t)=2(A_rB_r(\boldsymbol{x}_0))\cos((\omega_1-\omega_2)t-\hat{\psi}(\boldsymbol{x}_0))$; an example of this is shown in Fig. 2(A) with a solid line. Now consider the time trace at a second fixed point, $\boldsymbol{x}_1$, with the additive constant also removed, $V(\boldsymbol{x}_1,t)=2(A_rB_r)(\boldsymbol{x}_1)(\cos((\omega_1-\omega_2)t-\hat{\psi}(\boldsymbol{x}_1)))$; see the dotted line in Fig. 2(A). Notice that, except for magnitude, $V(\boldsymbol{x}_1,t)$ is very nearly $V(\boldsymbol{x}_0,t)$ except that it is time delayed by $(\hat{\psi}(\boldsymbol{x}_1)-\hat{\psi}(\boldsymbol{x}_0))/\Delta\omega$. That is $(\omega_1-\omega_2)t-\hat{\psi}(\boldsymbol{x}_0)=(\omega_1-\omega_2)(t+\delta)t-\hat{\psi}(\boldsymbol{x}_1)$, and we can interpret $\hat{\psi}(\boldsymbol{x}_1)-\hat{\psi}(\boldsymbol{x}_0)$ as a scaled time delay. So it is appropriate to define the quantity $\hat{\psi}(\boldsymbol{x}_1)/\Delta\omega=T(\boldsymbol{x}_1)$ as the arrival time, of the signal $V(\boldsymbol{x}_0,t)$ at the point $\boldsymbol{x}_1$. With this in mind, for the rest of this paper, we will refer to the scaled phase $\hat{\psi}(\boldsymbol{x})/\Delta\omega$ as the arrival time, $T(\boldsymbol{x})$.

We compute the arrival time using

$$C(\boldsymbol{x},\delta t):=\frac{1}{T}\int_0^T\tilde{v}(\boldsymbol{x}_0,t)\tilde{v}(\boldsymbol{x},t+\delta t)dt,$$

where

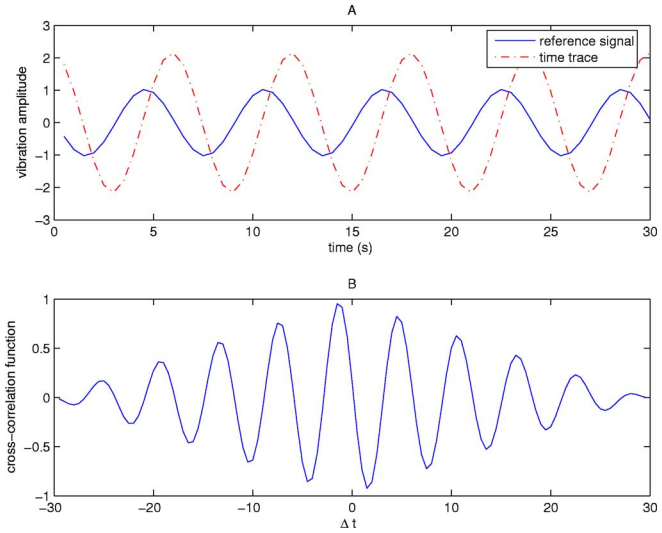


FIG. 2. (Color online) (A) Time traces of vibrational amplitude $V(\boldsymbol{x}_0,t)$ (solid line) and $V(\boldsymbol{x}_1,t)$ (dashed line). (B) Cross-correlation function of the two signals, $\tilde{v}(\boldsymbol{x}_1,t)$ and $\tilde{v}(\boldsymbol{x}_0,t)$.

$$\tilde{v}(\boldsymbol{x},t)=\begin{cases} V(\boldsymbol{x},t) & \text{if } 0\le t\le \hat{T} \\ V(\boldsymbol{x},t-\hat{T}) & \text{if } t>\hat{T} \\ V(\boldsymbol{x},t+\hat{T}) & \text{if } t>0. \end{cases}$$

See Fig. 2(B). Now we estimate the arrival times by $T(\boldsymbol{x}) \approx \delta t_{\max}$, where

$$\delta t_{\max} := \operatorname{argmax}_{\delta t\in[0,\hat{T}]} C(\boldsymbol{x},\delta t).$$

That is, δt, maximizes the correlation between the signals $\tilde{v}(\boldsymbol{x}_0,t)$ and $\tilde{v}(\boldsymbol{x},t+\delta t)$. To eliminate nonuniqueness that occurs because the two signals are cyclical, we: (1) Choose one of the maximums arbitrarily for the first point, $\boldsymbol{x}_0$; then, (2) for points neighboring $\boldsymbol{x}_0$, we choose local maxima in the cross-correlation function near the value $T(\boldsymbol{x}_0)$; and (3) to add stability to our procedure when finding the arrival time, $T(\boldsymbol{x})$, at a new point, $\boldsymbol{x}$, we use the median value of T at nearby points, that already have a computed arrival time, as a starting point.

We will use these computed arrival times as input to an inverse Eikonal solver described in the following; see also Ji *et al.* (2003); McLaughlin and Renzi (2006a, b). The output of this solver will be the speed of the moving interference pattern.

VI. SOLVING THE INVERSE EIKONAL EQUATION

The quantities $2\omega_1 F/\Delta\omega$ and $\omega_1 F/\Delta\omega$ are our imaging functionals for the crawling and holographic wave experiments, respectively. The goal now is to calculate $F=|\nabla T|^{-1}$ in a smart way, avoiding the essentially unstable calculation of dividing by derivatives of noisy data.

A slow, but robust, second-order method approximates the speed of the moving interference pattern using the elementary idea that speed is distance divided by time. So,

$$F\approx\left\{\frac{1}{2\Delta t}\left(\min_{\hat{x}^+}|\boldsymbol{x}-\hat{\boldsymbol{x}}^+|+\min_{\hat{x}^-}|\boldsymbol{x}-\hat{\boldsymbol{x}}^-|\right):\hat{\boldsymbol{x}}^{\pm}\text{ satisfies } T(\hat{\boldsymbol{x}}^{\pm})=T(\boldsymbol{x})\pm\Delta t\right\}. \tag{17}$$

We call this method for finding F the *distance method*. This is justified in McLaughlin and Renzi (2006a). A faster $0(m\log m)$ algorithm is described below.

Starting with the surface $S_T=\{(\boldsymbol{x},t)\,|\,T(\boldsymbol{x})=t, 0<t<T, \boldsymbol{x}\in\Omega\}$ where Ω is the image plane, define the higher dimensional function

$$\gamma(\boldsymbol{x},t)=\pm\min_{\hat{x}^+}\{|\boldsymbol{x}-\hat{\boldsymbol{x}}|:\hat{\boldsymbol{x}}\text{ satisfies } T(\boldsymbol{x})=t\}$$

where plus (minus) is chosen if $t>T(\boldsymbol{x})(t<T(\boldsymbol{x}))$, respectively. Then

$$\gamma(\boldsymbol{x},T(\boldsymbol{x}))=0 \text{ for } \boldsymbol{x}\in\Omega,\quad |\nabla\gamma|=1$$

so that

$$\gamma_t=|\nabla T|^{-1}=F \text{ on } \{(\boldsymbol{x},t)\,|\,\gamma(\boldsymbol{x},t)=0\}=S_T.$$

The potentially unstable term $|\nabla T|^{-1}$ is now replaced by γ_t which is in the numerator; and furthermore, no additional approximations are made to achieve this equation [see Osher and Fedkiw (2002); Sethian (1999), and Appendix A). For our inverse problem to obtain the $0(m\log m)$ algorithm speed, the extension from S_T to γ is made quickly and simultaneously for all times in our discretization. For those details, refer to McLaughlin and Renzi (2006b). The full algorithm for calculating the speed, F, in this way is called the *level curve method*. Note that as a final step we apply total variation minimization, (Rudin *et al.*, 1992).

VII. PHANTOM EXPERIMENTS

Combining the ideas from Sec. V (arrival time calculation) and Sec. VI (speed calculation from arrival times) gives a complete algorithm to recover interference pattern speed. The data are obtained using a Zerdine tissue mimicking phantom (CIRS Norfolk, VA), which is bowl-shaped and measures approximately $15\times15\times15$ cm in size. The phantom contains an isotropic background and a 1.3-cm-diam isotropic spherical stiff inclusion. The shear wave speed in the stiff inclusion is approximately $\sqrt{7}\approx2.65$ times faster than the background shear wave speed.

For the crawling wave experiment, two vibration sources are on opposite ends of the Phantom at frequencies, 250 and 250.15 Hz. Figure 3(A) shows a snapshot of the interference pattern in a middle region in the plane containing the two sources and the ultrasound transducer.

The first step to generate a shear wave speed reconstruction is to find the arrival times, T, from the spectral variance data. Before we do this, we preprocess the data. We use the one-dimensional fast Fourier transform on each time trace, and filter out all the frequencies except for a narrow band

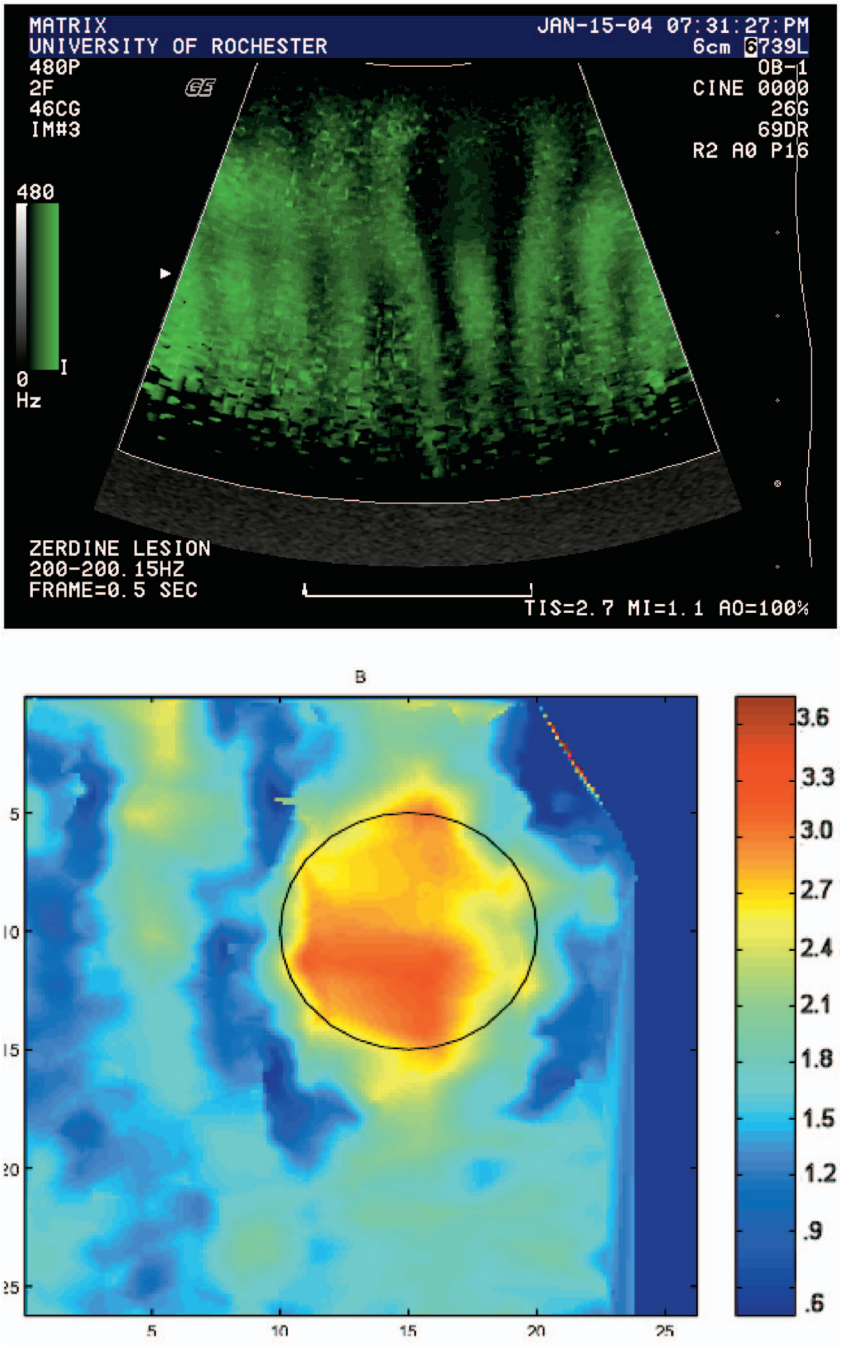
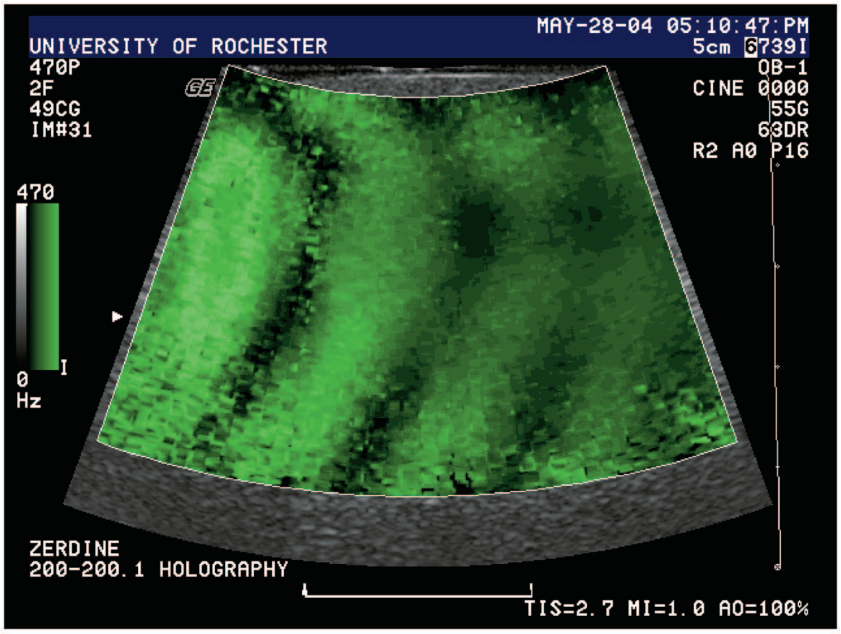


FIG. 3. (A) Snapshot of the moving interference pattern in the crawling wave experiment. (B) Imaging functional, $2\omega_1 F/\Delta\omega$, related to the shear wave speed in the crawling wave experiment. The units for the axis are millimeters and the color bar units are m/s.

around the driving frequency, $\Delta\omega=0.15$ Hz. Then we find the arrival time as discussed in Sec. VI. The interference pattern speed, $F=|\nabla T|^{-1}=\gamma_t$, is calculated with the inverse *level curve method* for the Arrival Time algorithm. The imaging functional, $2\omega_1 F/\Delta\omega$, is shown in Fig. 3(B). The wave speed contrast of the reconstruction is about 2.33, which is close to the actual wave speed contrast of 2.65. Note also the ring-like artifact around the recovered inclusion. This is likely due to the omission of the $\cos(\theta/2)$ term in our equation for the speed.

The interference patterns look very similar to a plane wave when two point sources are used. For an explanation, see Appendix B.

For the holographic wave experiment the frequency of the vibration source is 200.1 Hz. The ultrasound transducer is vibrated at 200 Hz. Figure 4(A) shows a snapshot of the moving interference pattern; it looks like an expanding half circle as one would expect from a point source. We find the arrival times, the interference pattern speed, F, as outlined in Secs. V and VI, and image the shear wave speed $C_s=\omega_1 F/\Delta\omega$; see Fig. 4(B). The imaging planes in the two experiments are at slightly different locations in the phantom. The black circle indicates the size of the stiff inclusion. In this reconstruction the wave speed contrast is almost 2, compared to the actual wave speed contrast of 2.65. There are fewer artifacts in this reconstruction. This is due to the more accurate relationship between the interference speed and the shear wave speed for the holographic wave experiment.



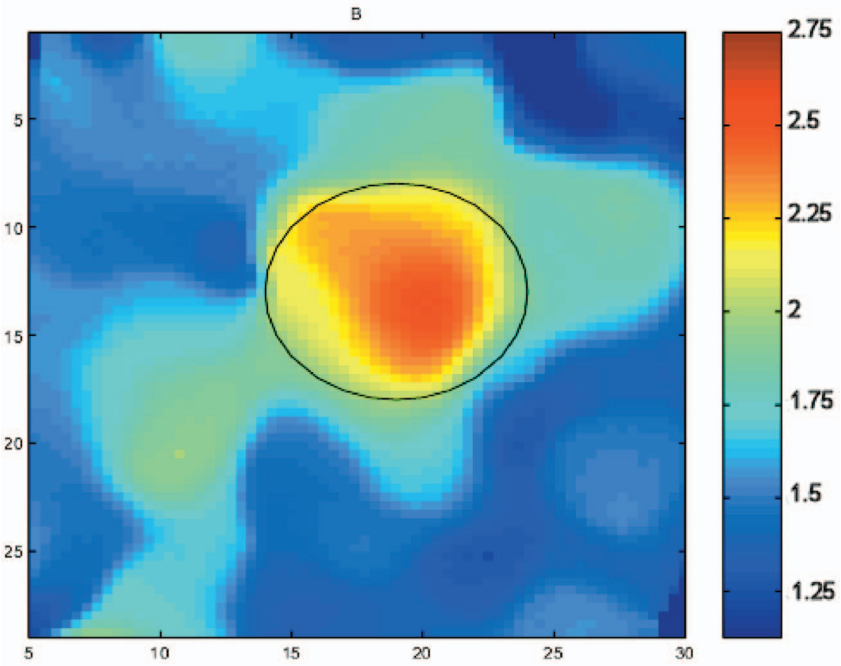


FIG. 4. (A) Snapshot of the moving interference pattern in the holographic wave experiment. (B) Recovery of the shear wave speed C_s in the holographic wave experiment. The units for the axis are millimeters and the color bar units are m/s.

VIII. SUMMARY AND CONCLUSION

We have developed a new algorithm, composed of two subalgorithms, to image the speed of moving interference patterns. The first subalgorithm finds the *arrival times* of one of the curves of interference maxima. The second subalgorithm takes as input the arrival times found by the first subalgorithm, and finds the moving interference pattern speed by solving the inverse Eikonal equation using the inverse *level curve method* for the Arrival Time algorithm. Our method is fully two dimensional taking into account both vertical and horizontal spread in the phase, and would easily generalize to three dimensions. The imaging functional is a multiple of the speed of the moving interference pattern.

There are two sources of artifacts in the images. One is the low bit rate achieved with only 16 levels of quantization in the display. We expect significantly less artifacts when the data gives a 256 color quantization. The second applies to the crawling wave experiment. Some artifacts occur because the imaging functional is a nonlinear function of the shear wave speed. These artifacts may be removed by solving the equations for ϕ_1, ϕ_2, and F simultaneously.

Note also, here we only consider interference pattern speed which is determined from the phase. When 256 color quantization data are available, one might consider also using a Helmholtz equation model, which has been considered in McLaughlin *et al.* (2006c); Dutt *et al.* (1997); Oliphant *et al.* (2000); Bishop *et al.* (2000); and Brown *et al.* (2001). Helmholtz inversion may be possible for the product $A\cos(\Delta\omega t - \omega_1\phi_1)$ obtained from the spectral variance calculation for the holographic wave experiment for experiments

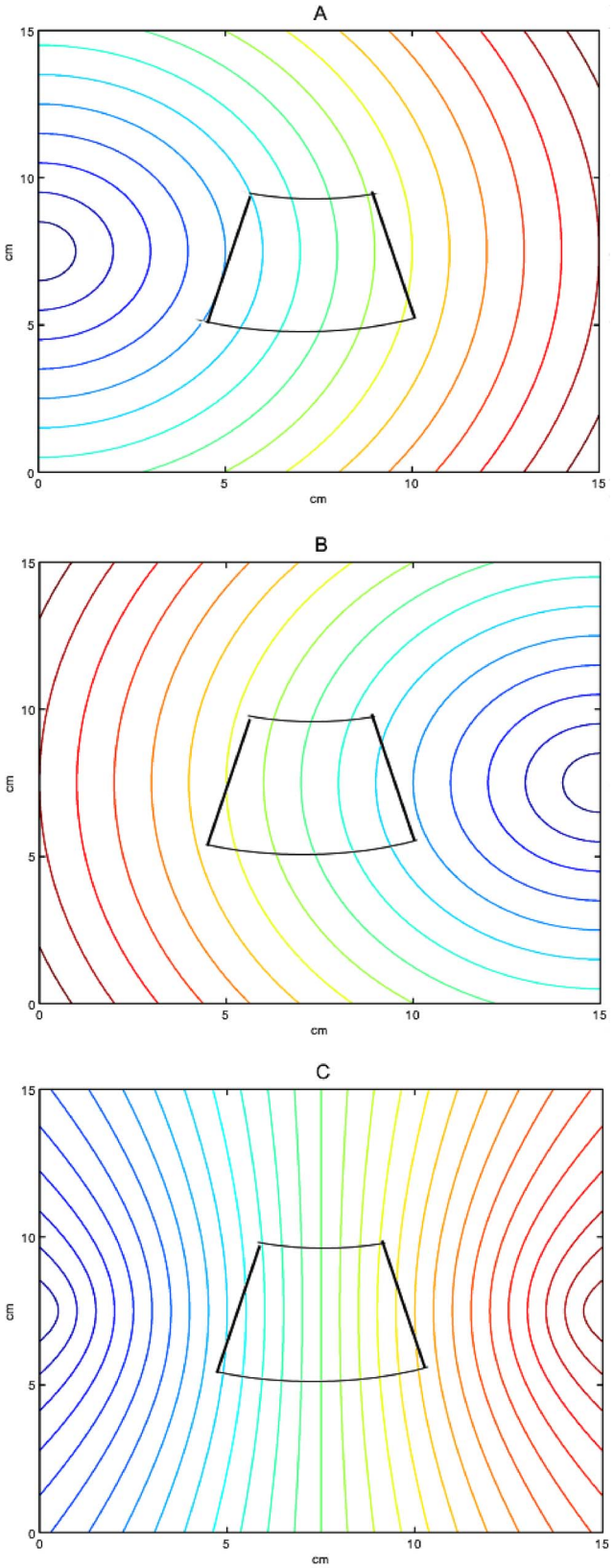


FIG. 5. (Color online) (A) Lines of constant phase for ϕ_1. (B) Lines of constant phase for ϕ_2. (C) Lines of constant phase for $\phi_1+\phi_2$. The solid black lines enclose the imaging region.

where: (1) There is very little out-of-of-plane motion; and (2) almost all excitation occurs in the radial direction (or downward direction if that is the direction of measurement). For the crawling wave experiment, this approach is more complicated since a differential equation that the expression $A_r B_r \cos(\Delta\omega t-(\omega_1\phi_1+\omega_2\phi_2))$ satisfies, even under the simplifying assumptions given earlier, is substantially more complicated.

It may, however, be possible to obtain a more accurate phase calculation by using the rf data more directly.

APPENDIX A

In this appendix we show that, on a level surface of the phase, $\psi(\boldsymbol{x},t)=k$,

$$\psi_t=|\nabla\psi|F \tag{A1}$$

is satisfied where F is the component of the velocity in the direction, $-\nabla\psi$, which is normal to the level curve $\psi(\boldsymbol{x},t)=k$, fixed k. Let $\boldsymbol{X}(t)$ be a parametric representation of a point lying on $\psi(\boldsymbol{x},t)=k$ with $x(t_0)=x_0$, some $(\boldsymbol{x}_0,t_0)$ satisfying $\psi(\boldsymbol{x}_0,t_0)=k$. Since $-\nabla\psi/|\nabla\psi|$ is normal to the curve $\psi(\boldsymbol{x},t_0)=k$, at $\boldsymbol{x}=\boldsymbol{x}_0$,

$$F(\boldsymbol{x}_0)=\boldsymbol{X}_t\cdot(-\nabla\psi/)|\nabla\psi|. \tag{A2}$$

Taking a time derivative of $\psi(\boldsymbol{x}(t),t)=k$ yields

$$\psi_t+\nabla\psi\cdot\boldsymbol{X}_t=0. \tag{A3}$$

Multiplying through Eq. (A3) by $1/|\nabla\psi|$, and using Eq. (A2), leads to

$$\psi_t=|\nabla\psi|F. \tag{A4}$$

Since $(\boldsymbol{x}_0,t_0)$ is arbitrarily chosen on $\psi(\boldsymbol{x},t)=k$, Eq. (A1) is established.

APPENDIX B

The interference pattern lines in Fig. 3 are similar to a plane wave for the following reasons: $\nabla\hat{\psi}/\Delta\omega\approx\omega_1(\nabla\phi_1+\nabla\phi_2)/\Delta\omega$ determines the direction of motion of the interference pattern; and (2) the image plane window is some distance from each source and the individual waves from each source move in opposite directions; this implies that in the background the vertical components of $\nabla\phi_1$ and $\nabla\phi_2$ will have opposite sign. To demonstrate we solve the equations $|\nabla\phi_1|=1$, and $-|\nabla\phi_2|=-1$, on a 15 cm$\times$15 cm square with the initial conditions $\phi_1(0,7.5)=0$, and $\phi_2(15,7.5)=0$. The lines of constant phase for ϕ_1 and ϕ_2 are simply expanding circles and are shown in Figs. 5(A) and 5(B). For $\phi_1+\phi_2$ the lines of constant phase, shown in Fig. 5(C), are given by a family of parabolas (Wu 2005). However, near the line equidistant to the two point sources the lines of constant phase for $\phi_1+\phi_2$ are nearly vertical; see Fig. 5(C). Note also that in Fig. 5(C) there are twice as many lines of constant phase as in Figs. 5(A) and 5(B). This is consistent with the additional factor of 2 in the crawling wave imaging functional.

Bercoff, J., Tanter, M., Chaffai, S., Sandrin, L., and Fink, M. (**2002**). "Ultrafast imaging of beam formed shear waves induced by the acoustic radiation force. Application to transient elastography," Proc.-IEEE Ultrason. Symp. **2**, 1899–1902.

Bercoff, J., Tanter, M., and Fink, M. (**2004**). "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **19**, 396–409.

Bercoff, J., Tanter, M., Sandrin, L., Catheline, S., and Fink, M. (**2001**). "Ultrafast compound imaging for 2D displacement vector measurements:

Application to transient elastography and color flow mapping," Proc.-IEEE Ultrason. Symp. **2**, 1619–1622.
Bishop, J., Samani, A., Sciaretta, J., and Plewes, D. B. (**2000**). "Two dimensional MR elastography with linear inversion reconstruction: Methodology and noise analysis," Phys. Med. Biol. **45**, 2081–2091.
Catheline, S., Thomas, J.-L., Wu, F., and Fink, M. (**1999**). "Diffraction field of a low frequency vibrator in soft tissues using transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1013–1019.
Dutt, V., Manduca, A., Muthupillai, R., Ehman, R., and Greenleaf, J. (**1997**). "Inverse approach to elasticity reconstruction in shear wave imaging," Proc.-IEEE Ultrason. Symp. **2**, 1415–1418.
Gao, L., Parker, K. J., and Alam, S. K. (**1995**). "Sonoelasticity imaging: Theory and experimental verification," J. Acoust. Soc. Am. **97**, 3875–3880.
Gennisson, J. L., Catheline, S., Chaffai, S., and Fink, M. (**2003**). "Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear wave speeds in muscles," J. Acoust. Soc. Am. **114**, 536–541.
Greenleaf, J. F., and Fatemi, M. (**1998**). "Ultrasound-stimulated vibroacoustic spectrography," Science **280**, 82–85.
Huang, S.-R., Lerner, R. M., and Parker, K. J. (**1990**). "On estimating the amplitude of harmonic vibration from the Doppler spectrum of reflected signals," J. Acoust. Soc. Am. **88**, 2702–2712.
Ji, L., and McLaughlin, J. R. (**2003**). "Shear stiffness identification in biological tissues: The full elastic model" (unpublished).
Ji, L., and McLaughlin, J. R. (**2004**). "Recovery of the Lamé parameter μ in biological tissues," Inverse Probl. **20**, 1–24.
Ji, L., McLaughlin, J. R., Renzi, D., and Yoon, J.-R. (**2003**). "Interior elastodynamics inverse problems: Shear wave speed reconstruction in transient elastography," Inverse Probl. **19**, S1–29.
Konafagou, E. E. (**2000**). "Precision estimation and imaging of normal and shear components of the 3D strain tensor in elastography," Phys. Med. Biol. **45**, 1553–1563.
Konafagou, E. E., Harrigan, T., and Ophir, J. (**2000a**). "Shear strain estimation and lesion mobility assessment in elastography," Ultrasonics **38**, 400–404.
Konafagou, E. E., and Ophir, J. (**1998**). "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.
Konafagou, E. E., Varghesse, T., and Ophir, J. (**2000b**). "Theoretical bounds on the estimation of transverse displacement, transverse strain and Poisson's ratio in elastography," Ultrason. Imaging **22**, 153–177.
Lerner, R. M., Parker, K. J., Holen, J., Gramiak, R., and Waag, R. C. (**1988**). "Sono-elasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets," Acoust. Imaging **16**, 317–327.
Levinson, S. F. M., and Sata, T. (**1995**). "Sonoelastic determination of human skeletal-muscle elasticity," J. Biomech. **28**, 1145–1154.
Manduca, A., Oliphant, T. E., Dresner, M. A., Mahowald, J. L., Kruse, S. A., Amromin, E., Felmlee, J. P., Greenleaf, J. F., and Ehman, R. L. (**2001**). "Magnetic resonance elastography: Non-invasive mapping of tissue elasticity," Med. Image Anal **5**, 237–254.
McLaughlin, J. R., and Renzi, D. (**2006a**). "Shear wave speed recovery in transient elastography and supersonic imaging using propagating fronts," Inverse Probl. **22**, 681–706.
McLaughlin, J. R., and Renzi, D. (**2006b**). "Using level set based inversion of arrival times to recover shear wave speed in transient elastography and supersonic imaging," Inverse Probl. **22**, 706–725.
McLaughlin, J. R., Renzi, D., Yoon, J.-R., Ehman, R. L., and Manducca, A. (**2006c**). "Variance controlled shear stiffness images for MRE data," *IEEE International Symposium on Biomedical Imaging, Macro to Nano* p. 960–963.
Miller, G., and Pursey, H. (**1954**). "The field and radiation impedance of mechanical radiators on the free surface of a semi-infinite isotropic solid," Proc. R. Soc. London, Ser. A **223**, 521–544.
Nightingale, K., Mcaleavey, S., and Trahey, V. (**2003**). "Shear-wave generation using acoustic radiation force: In vivo and ex vivo results," Ultrasound Med. Biol. **29**, 1715–1723.
Nightingale, K., Stutz, D., Bentley, R., and Trahey, G. (**2002**). "Acoustic radiation force impulse imaging: Ex vivo and in vivo demonstration of transient shear wave propagation," IEEE International Symposium on Biomedical Imaging, Cat. No. 02EX608 528-8
Oliphant, T. E., Kinnick, R. R., Manduca, A., Ehman, R. L., and Greenleaf, J. F. (**2000**). "An error analysis of Helmholtz inversion for incompressible shear vibration elastography with application to filter-design for tissue characterization," *Ultrasonics Symposium, 2000 IEEE*, Vol. **2**, pp. 1795–1798.
Ophir, J., Cespedes, I., Ponnekanti, H., Yazdi, Y., and Li, X. (**1991**). "Elastography: A quantitative method for imaging the elasticity of biological tissues," Ultrason. Imaging **13**, 111–134.
Osher, S. J., and Fedkiw, R. (**2002**). *Level Set Methods and Dynamic Implicit Surfaces* (Springer, Berlin).
Osher, S. J., and Sethian, J. A. (**1988**). "Front propagation with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," J. Comput. Phys. **79**, 12–49.
Rudin, L. I., Osher, S., and Fatemi, E. (**1992**). "Nonlinear total variation based noise removal algorithms," Physica D **60**, 259–268.
Sandrin, L., Tanter, M., Catheline, S., and Fink, M. (**2002a**). "Shear modulus imaging with 2-D transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 426–435.
Sandrin, L., Tanter, M., Catheline, S., and Fink, M. (**2002b**). "Time-resolved 2D pulsed elastography. Experiments on tissue-equivalent phantoms and breast in-vivo," Proc. SPIE **4325**, 120–126.
Sandrin, L., Tanter, M., Gennisson, J. L., Catheline, S., and Fink, M. (**2001**). "Shear elasticity probe for soft tissues with 1-D transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 436–446.
Sarvazyan, A. P., Skovoroda, A. R., Emalianov, S. Y., Fowlkes, L. B., Pipe, J. G., Adler, R. S., and Carson, P. L. (**1995**). "Biophysical bases of elasticity imaging," Acoust. Imaging **21**, 223–240.
Sethian, J. A. (**1999**). *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science* (Cambridge University Press, Cambridge).
Tanter, M., Bercoff, J., Sandrin, L., and Fink, M. (**2002**). "Ultrafast compound imaging for 2-D motion vector estimation: Application to transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1363–1374.
Taylor, L. S., Porter, B. C., Rubens, D. J., and Parker, K. J. (**2000**). "Three-dimensional sonoelastography: Principles and practices," Phys. Med. Biol. **45**, 1477–1494.
Wu, Z. (**2005**). "Shear wave interferometry and holography, an application of sonoelasticity," Ph.D. thesis, University of Rochester, Rochester, NY.
Wu, Z., Rubens, D. J., and Parker, K. J. (**2006**). "Sonoelastographic imaging of interference patterns for estimation of the shear velocity distribution in biomaterials," J. Acoust. Soc. Am. **120**, 535-545.
Wu, Z., Taylor, L. S., Rubens, D. J., and Parker, K. J. (**2002**). "Shear wave focusing for three-dimensional sonoelastography," J. Acoust. Soc. Am. **111**, 439–446.
Wu, Z., Taylor, L. S., Rubens, D. J., and Parker, K. J. (**2004**). "Sonoelastographic imaging of interference patterns for estimation of the shear velocity of homogeneous biomaterials," Phys. Med. Biol., **49**, 911–922.

Erratum: "Depth-pressure relationships in the oceans and seas" [J. Acoust. Soc. Am. 103(3), 1346–1352 (1998)]

Claude C. Leroy
IMM Vivaldi, 30 Chemin de la Baou, Sanary Sur Mer, 83110, France

(Received 5 January 2007; accepted 5 January 2007)

[DOI: 10.1121/1.2534201]

PACS number(s): 43.30.Pc, 43.30.Es, 43.28.Fp [RAS]

A number of errors in various coefficients of equations originally presented in this paper were recently detected following the choice of the algorithm by the UK National Physical Laboratory in its on-line Technical Guides. They are as follows:

— Equations 3 and 5: The term in P^2 should read: $-2.2512\times10^{-1}\ P^2$ (instead of -2.512 etc)

— Equation 11: The first coefficient should read: 9.7803 (instead of 0.7803)

— Equation 12: The first term should read, as in Table II: $1.0\times10^{-2}\ Z/(Z+100)$ (instead of $0.8\ Z/(Z+100)$)

 0001-4966/2007/121(4)/2447/1/$23.00

Erratum: "2aPP29. An upper bound on the temporal resolution of human hearing" [J. Acoust. Soc. Am 120(5), 2085 (2006)]

Milind N. Kunchur
Univ. of South Carolina, Dept. of Physics & Astronomy, Columbia, South Carolina 29208

(Received 2 February 2007; accepted 2 February 2007)
[DOI: 10.1121/1.2711424]

PACS number(s): 43.66.Mk, 43.10.Vx, 43.28.Fp

There is a publisher's error in the first sentence of this abstract. The word "microseconds" was abbreviated and was published as "ms." The correct first sentence of this abstract should read:

"This work obtained an upperbound of about 5 microseconds for the temporal resolution of hearing by studying human discriminability for bandwidth restriction by low-pass filtering."

 0001-4966/2007/121(4)/2448/1/$23.00